Personalized Motivation-supportive Messages for Increasing Participation in Crowd-civic Systems

PAUL GRAU, Korea Advanced Institute of Science and Technology, Republic of Korea
BABAK NADERI, Technische Universität Berlin, Germany
JUHO KIM, Korea Advanced Institute of Science and Technology, Republic of Korea

In crowd-civic systems, citizens form groups and work towards shared goals, such as discovering social issues or reforming official policies. Unfortunately, many real-world systems have been unsuccessful in continually motivating large numbers of citizens to participate voluntarily, despite various approaches such as gamification and persuasion techniques. In this paper, we examine the influence of personalized messages designed to support motivation as asserted by the Self-Determination Theory (SDT). We designed a crowd-civic platform for collecting community issues with personalized motivation-supportive messages and conducted two studies: a pair-comparison experiment with 150 participants on Amazon's Mechanical Turk and a live deployment study with 120 university members. Results of the pair-comparison study indicate applicability of SDT's perspective in crowd-civic systems. While applying it in the live system surfaced several challenges, including recruiting participants without interfering with general motivations, the collected data exhibited similar promising trends.

Additional Key Words and Phrases: Crowd-civic systems; personalization; motivation; self-determination theory

ACM Reference Format:
Paul Grau, Babak Naderi, and Juho Kim. 2018. Personalized Motivation-supportive Messages for Increasing Participation in Crowd-civic Systems. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 60 (November 2018), 22 pages. https://doi.org/10.1145/3274329

1 INTRODUCTION

Interest in civic technology has been growing quickly in the past decade. Researchers have been seeking novel technical solutions to fundamental problems of communities: how to "govern, organize, serve, and identify matters of concern" [3]. In crowd-civic systems, large numbers of citizens form groups in processes that are related to crowdsourcing [22]: they join online communities and contribute small pieces towards common goals, such as the collection of local knowledge, discovery of social issues, or even reforming official policies. Users of crowd-civic systems work together to achieve tangible outcomes, and they often do so voluntarily.

Encouraging people to join an activity and keeping them motivated has always been a major challenge for crowdsourcing system designers [31]. Even in paid micro-task crowdsourcing, in which direct influence of external rewards is expected, research has shown that a combination of

Authors' addresses: Paul Grau, paul@graycoding.com, Korea Advanced Institute of Science and Technology, School of Computing, Daejeon, Republic of Korea; Babak Naderi, babak.naderi@tu-berlin.de, Technische Universität Berlin, Software Engineering and Theoretical Computer Science, Berlin, Germany; Juho Kim, juhokim@kaist.ac.kr, Korea Advanced Institute of Science and Technology, School of Computing, Daejeon, Republic of Korea.

2573-0142/2018/11-ART60 $15.00
https://doi.org/10.1145/3274329

factors, including the task's interestingness, expected workload, and profitability determines the workers' decision to opt in [29]. In crowd-civic systems, citizens are typically asked to contribute voluntarily for a variety of reasons, including the greater good (e.g. as a civic duty) and other indirect benefits (e.g. social or epistemic aspects) [2], in absence of monetary rewards. Thus, due to diverse motivations of citizens [2], the issues of self-selection [1] and the long-tail of participation [36], often a particular, small, and homogeneous group of citizens ends up contributing most of the work. This is less than desirable for fulfilling the democratic ideal of representativeness [1]. For technology to advance social progress, all voices need to be heard. However, there is a large and underrepresented group of people that participate rarely or never in these kinds of systems, which has lead researchers to call for more inclusive, differentiated design — for moving away from the currently prevalent "one size fits all designs of civic technology" [2].

In the field of work motivation, theories often differentiate between *intrinsic* and *extrinsic* motivation based on the origin of the motivation. Despite the positive effect of intrinsic motivation on performance [8], in reality, many tasks are not composed to produce intrinsic motivation. The Self-Determination Theory (SDT) [32] is a well-established theory of motivation which is empirically validated in different domains like education, health care, and work. It distinguishes between intrinsic motivation and spectra of extrinsic motivation in terms of involved behavior regulation processes. Besides the task itself, SDT also considers the individuals who perform an activity and how they internalize its reasons and goals. It provides a unique framework for studying the influence of user and environmental factors, as well as for investigating their interaction on motivation and, consequently, on desired outcomes.

Previous research has shown how users with varying personalities, as described by standard tests like the Big Five personality traits, have a preference for different motivational affordances [15]. This falls into the category of "personality-targeted design" [30], a framework for theory-driven personalized user interfaces. In this paper, we operationalize personality in the context of SDT, which regards motivational orientations as lasting aspects of one's personality. Addressing individual differences can effectively increase participation, not only in total but also from more diverse people, and as a result, boost representativeness [16]. To our knowledge, this idea has not been investigated empirically in the domain of crowd-civic systems. Thus, in this work, we address the research question: in what way can motivation theory inform the design of personalized user interfaces of crowd-civic systems to support different motivations?

This work investigates the influence of *personalized motivation-supportive design* on participation of participants in a crowd-civic platform with two studies. Results of a controlled pair-comparison study indicate the applicability of SDT in crowd-civic systems and the possibility of personalization. Data collected in a subsequent live deployment using an application based on the initial findings exhibited similar promising trends, but several challenges arose regarding personalization. By studying this concept in two settings, one focusing on control and one on ecological validity, we hope to provide a more balanced view on the possible advantages and disadvantages.

The contributions of this paper are: (1) evidence for the possibility of personalized motivation-supportive messages in a controlled setting, (2) evidence for the positive effect of personalized motivation-supportive messages on participation in a real-world crowd-civic system, and (3) artifacts of designing a crowd-civic system and translating theoretical constructs into concrete designs.

2 RELATED WORK

Our work is based on previous research in three fields: the application domain of crowd-civic systems, previous studies of message diversification, and personality-targeted design.

2.1 Motivation in Crowd-civic Systems

Crowd-civic systems are an emerging category of civic technology applications where citizens form groups working towards shared goals, such as the collection of local knowledge, discovery of social issues, or reforming official policies. It is the application of crowdsourcing to civic contexts that comes with a number of unique challenges [22]. A major issue of crowd-civic systems is attracting and retaining the "right" participants: depending on the application, different characteristics of participants might be desirable, such as expertise, diversity of opinions, and representativeness. The plain number of volunteers is also a critical issue. After an initial "hype", maintaining a critical mass for long-term engagement has been difficult for many platforms [36].

Aitamurto and Landemore [1] conducted a study in Finland employing elements of crowdsourcing for deliberation on lawmaking. In an open call, citizens were invited to join an online discussion forum to debate details of a proposed law change regarding off-road traffic. The researchers saw high user engagement that exhibited qualities of democratic deliberation as defined by previous research. However, the authors also note the unsolved problem of limited representativeness, owing mostly to the nature of self-selection and required level of technology skills.

Participation in crowd-civic systems is often voluntary. Designers of these systems have to clearly define and present the benefits to the users. Some people might be motivated by acting "for the greater good", for altruistically advancing societies. Usually, personal motivations are more complex than altruism and vary significantly. In another study by Aitamurto and Saldivar [2], they analyzed the dynamic nature of motivation factors. The authors argue that design for crowdsourced policymaking should support epistemic and interactive aspects to keep participants engaged, and more generally, that system designers should take different motivations into consideration.

In their work about voluntary participation in online experiments, Jun et al. [16] showed how motivation affects selection of participants. They suggest to diversify promotions to engage users with different motivations or from different demographics, which could mitigate self-selection biases and reduce dropout. In our work, we apply this idea in a civic context, and additionally attempt to employ personalization.

In summary, crowd-civic systems suffer from self-selection and non-diverse participants who aren't easily motivated to participate equally and for an extended period of time. This has lead numerous researchers to call for more inclusive, differentiated design. Combining these ideas and using the Self-Determination Theory (SDT), our work tries to contribute a step towards moving away from the currently prevalent "one size fits all designs" [2].

2.2 Message and Incentive Diversification

Recently, several works have investigated different methods of diversifying messaging strategies in various domains of voluntary action, such as social activism, behavior change, and online discussion.

Botivist [33] used bots on Twitter to individually address social media users that might contribute action to specific activism causes. The authors evaluated different strategies of messaging people with different message framings, such as referring to solidarity, loss, or gain. Unexpectedly, the most direct call-to-action was on average more effective than other strategies of persuasion. However, the authors didn't account for possible individual differences between participants.

Kocielnik and Hsieh [17] explored diversifying message-based triggers in a behavior change application. To counter users' annoyance and boredom when seeing a message repeatedly, they generated target-diverse and self-diverse message variants, referring to concepts that are cognitively close to either the action or recipient of a message. They found that self-diverse messages worked well as reminders and even motivators, but they note that in terms of motivation, all their messages

followed a single, generally positive motivational strategy. In our work, we explore messages which support diverse kinds of motivations that exist in individual users.

Investigating designs that encourage "one-time" or novice contributors of online communities, McInnis et al. [23] found that the phrasing of prompts matters. Small differences in call-to-action messages, such as referring to more or less specific goals, can affect the quantity and quality of contributions. The authors also used personal factors such as self-efficacy to draw correlations with both longer comments and higher responsiveness. They argue that personal factors could be used to achieve more representative recruiting. In our work, we extend this idea to providing personalized experiences within applications not only for recruiting, but for extended engagement within an application.

Related to personal motivations, Hsieh and Kocielnik [14] investigated how different incentives may attract different people. They tested incentives corresponding to different human values, such as openness-to-change and risk-seeking, and saw correlations between participants' choice of reward and self-reported values. The authors argue that using diverse incentives can improve participation diversity. Other researchers have also shown how incentives make a difference in crowdsourcing participation, both voluntary and under various payment schemes [20]. In our work, we investigate a similar idea of varying incentives and motivations, but based on self-determination. In addition to a paid crowdsourcing experiment, we also investigate a voluntary setting.

2.3 Personality-targeted Design

Personality-targeted design, first introduced by Nov and Arazy [30], is a framework for theory-driven design of personalized user interfaces. In that study, they showed how a user's conscientiousness level can predict their perception of a "critical mass", i.e. the belief that there is a large number of other users who participate; their user interface contained an indicator of high or low participation from others, and they discovered a correlation between a user's conscientiousness level and the level of engagement given one version of the interface or the other.

In a related study, Jia et al. [15] showed how people's personality, as measured with the Big Five test [10], affects their preference for different gamification affordances. As an example, extroverted people thought that collecting points, advancing levels, and leaderboards are helpful and enjoyable. In an older study, Moon [26] already observed similar effects. In their study, participants completed a personality test to be assigned to two groups: dominant and submissive people. Then, they were asked to use an online shopping recommendation agent. Results showed that people were more likely to be persuaded by the recommendation when it was conveyed in a style matching their personality.

These studies indicate new opportunities to use personalization in many parts of user interface design. Whereas increasing engagement through personalization often means providing users with products or content that matches their specific *interests and goals*, it can also be used to adapt interfaces with strategies matching their *personality* and, in our case, *motivation orientations*, regarded by Self-Determination Theory as lasting aspects of one's personality.

3 MOTIVATION-SUPPORTIVE DESIGN

The Self-Determination Theory (SDT) can offer numerous insights into designing with users' motivation in mind. SDT provides a broad, well-established, and empirically validated framework in different domains and cultures to study motivation. In this section, we provide a quick overview of SDT and how its findings relate to the context of crowd-civic systems.

SDT considers different types of motivations: intrinsic motivation, a spectrum of extrinsic motivation, and amotivation. It also addresses the process of *internalization* in which "an individual acquires an attitude, belief, or behavioral regulation and progressively transforms it into a personal

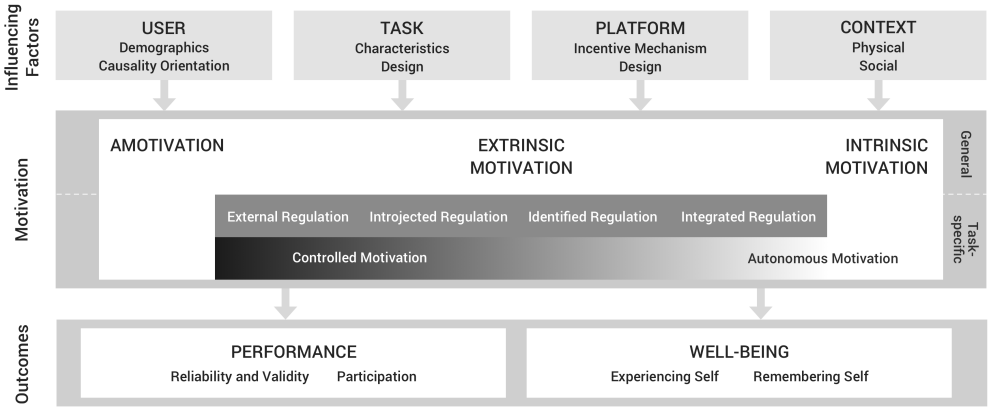


Fig. 1. Taxonomy of motivation in crowd work, after Naderi [29, p. 8]. Our work attempts to personalize the *platform* based on *user factors* to influence *participation* in a given *task*.

value, goal, or organization" [6]. Internalized types of extrinsic motivation share similar properties with intrinsic motivation. In combination, they create *autonomous motivation*, whereas other types of extrinsic motivation merges to create *controlled motivation*.

Within SDT's framework, type and amount of one's motivation do not only depend on the activity, but also on the person him/herself and the environment [32]. SDT claims that certain aspects of the environment can influence autonomous motivation. It also addresses three basic psychological needs and argues that they are universal for all humans: autonomy, competence, and relatedness (see Table 2 for definitions). When an environment supports the satisfaction of these needs, it is called *autonomy-supportive*. SDT claims that autonomy-supportive environments facilitate internalization of extrinsic motivation and positively influence performance and well-being of people [8].

Human behavior can be regulated by a variety of factors that can be mapped on a continuum of self-determination (Table 1). The processes involved in these regulations can be thought in terms of an inverse relationship between self-determined and controlled motivation. However, these factors are environment- and task-specific — they can vary not just between individuals, but also between tasks of the same person. People's motivations are usually a combination of different regulations. In terms of individual orientations, SDT offers the concept of *general causality orientations*. They describe ways people generally and across many settings orient themselves to environments and regulate their behaviour (Table 2).

SDT has been studied in various domains of organizations, including in the context of organizational citizenship behaviors [8]. It has been argued that satisfaction of the three basic needs will enhance workers' motivations and in turn yield the outcomes deemed important in the respective field of work. In a volunteer work setting, perceived autonomy support may relate positively to the amount of volunteering. Interestingly, while studying factors for involvement in an important national referendum, it was found that intrinsic motivation itself is not enough to predict actual voting behavior. "When intrinsically interested in the issues, people became well informed, but only when they were motivated by the importance of the issues to themselves were they likely to actually exert the effort to go out and vote." [8, citing 18] Furthermore, numerous studies have shown the potentially detrimental effects of external rewards on autonomous motivation: payments decrease voluntary pro-social behavior [8]. This leads us to **RQ 1: What is the effect of external regulation on autonomously motivated online volunteer work in a crowd-civic system?**

Table 1. Behavior regulatory styles on the self-determination continuum, associated processes, and perceived locus of causality. Table reproduced from Ryan and Deci [32].

Amotivation	**Extrinsic motivation**				**Intrinsic motivation**
	External regulation	Introjection	Identification	Integration	
Perceived non-contingency, low perceived competence, nonrelevance, nonintentionality	Saliance of extrinsic rewards or punishments, compliance, reactance	Ego involvement, focus on approval from self or others	Conscious valuing of activity, self-endorsement of goals	Hierarchical synthesis of goals, congruence	Interest, enjoyment, inherent satisfaction
Impersonal	External	Somewhat external	Somewhat internal	Internal	Internal
Lack of motivation	Controlled motivation	Moderately controlled motivation	Moderately autonomous motivation	Autonomous motivation	Inherently autonomous motivation

For certain work environments, it may be beneficial to focus on getting the best participants and supporting their autonomous motivations. If diversity and representativeness is the goal, however, as in crowd-civic systems, that might neglect a large part of the population. It is also a matter of perspective of time: in the long term, supporting autonomous motivations may be beneficial for everyone, but for contexts like a crowd-civic system that need to be effective immediately and sustainedly, we have to consider what we can do momentarily to support individuals to achieve the goal of increasing participation. This leads us to **RQ 2: Can individually targeted motivation support increase engagement of participants with diverse motivations in crowd-civic systems?**

SDT has been used to inform design of interactive computer systems before. In his work, Deen [7] investigated how autonomy-supportive game design can improve motivation to learn. The

Table 2. Basic motivational needs [32] (first row) and individual general causality orientations [6] (second row) according to SDT.

Need for autonomy Choosing to engage in behavior that is compatible with one's values, out of personal interest or expression of the self.	**Need for competence** Sense of proficiency and feelings of effectiveness in one's work	**Need for relatedness** Feeling connected with others, feeling interdependent with others, and feeling of belonging to a group or with other individuals
Impersonal orientation Believes that attaining desired outcomes is beyond his or her control and that achievement is largely a matter of luck or fate. Likely to be anxious and feel very ineffective. Has no sense of being able to affect outcomes or cope with demands or changes. Tends to be amotivated and to want things to be as they always were.	**Controlled orientation** Oriented toward being controlled by rewards, deadlines, structures, ego-involvements, and the directives of others. Likely to be dependent on rewards or other controls, and may be more attuned to what others demand than to what they want for themselves.	**Autonomy orientation** Oriented towards environments that stimulate intrinsic motivation, are optimally challenging, provide informational feedback, and allow choice. Tends to display greater self-initiation, seek activities that are interesting and challenging, and take greater responsibility for his or her own behavior.

author gives a wide range of examples of how findings from SDT can be translated to concrete design decisions. However, games make use of motivations which are quite distinct from voluntary participation in crowd-civic systems. The primary driver for games tends to be intrinsic motivation, e.g. enjoyment and inherent satisfaction, but voluntary participation is dependent on at least some extrinsic factors, e.g. importance of the issue and expected possible outcomes. Naderi [29] investigated SDT-based motivation factors in the context of paid crowd work. In this paper, using the same taxonomy (Figure 1), we want to contribute to the discussion of supporting individual motivations in a voluntary participation context, crowd-civic systems.

For measuring motivation in the forms of regulatory styles and causality orientations, generally regarded as lasting aspects of personality, we used two standard questionnaires that are directly related to SDT constructs. The Motivation to Volunteer Scale (MVS) developed by Grano et al. [11] is used to score a person on scales related to the behavioral regulatory styles. It has been shown to be highly correlated with self-reported frequency of volunteering, both in the past and future intention [11]. The General Causality-Orientation Scale (GCOS) [6] measures people's orientations on three scales: autonomy, control, and impersonal. Motivation orientation has been used as a measure in other studies regarding voluntary participation, for example in online communities [4] and also in regards to gamification [24].

4 STUDY 1: CONTROLLED STUDY WITH CROWD WORKERS

To gather initial evidence that using Self-Determination Theory (SDT) to inform messaging strategies of crowd-civic systems to support diverse motivations holds merit, we ran an online study with crowdworkers on Amazon Mechanical Turk (MTurk). We designed a user interface mockup for a fictional mobile application which features a prominent call-to-action to voluntarily participate in collecting ideas for a community. We implemented seven different message strategies based on SDT and collected rankings of individual preferences for each version through pairwise comparisons.

The hypotheses for this study were:

H1: different people prefer different motivation strategies, i.e. there is not a clear "winning strategy",

H2: individual preferences correlate with personal differences in motivation or causality-orientation.

In other words, we wanted to see if strategies based on findings from SDT yield any differences in the crowd-civic domain (H1) as well as if the effect correlates with personal differences, manifested in motivation and causality-orientation (H2).

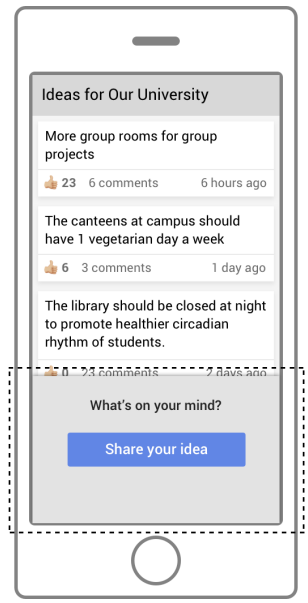


Fig. 2. The baseline (BL) version of the design sketch. The screen includes a feed of recent ideas and a prominent call-to-action at the bottom, for which we designed 6 other versions.

4.1 Design

This fictional application aims to collect a community's issues and ideas. We designed the main screen (Figure 2) which shows the feed of issues and a prominent call for participation.

We designed seven different versions of this call-to-action: a baseline that features only the basic instruction, and six versions that are hypothesized to support different kinds of motivations, as explained in detail in section 3. The versions are shown in Table 3. It is expected that the causality-orientated versions (**A**utonomy, **C**ontrol, **I**mpersonal) exhibit bigger differences overall, as *A* combines multiple aspects of the need-based versions, *C* is the only version with a strong

Table 3. Study 1 design versions and messages for the prototype's call-to-action component following the rationales from Table 2. The messages are inserted into the screenshot depicted in Figure 2.

Need for autonomy (nA)	**Need for competence** (nC)	**Need for relatedness** (nR)
Sharing ideas matters! [Share your idea] Express your opinion, let everyone hear our voice!	Your opinion is needed. [Share your idea] You can have an impact on your surroundings.	Your community needs your input! [Share your idea] 24 other people contributed today.
Autonomous orientation (A)	**Impersonal orientation** (I)	**Controlled orientation** (C)
Make our community a better place! [Share your idea] 24 other people contributed today. Your ideas matter.	Change may be beyond our control... [Share your idea] but there's a chance someone sees your idea and considers it.	Become a Contributor of the Month! [Share your idea] $20 gift cards for Top 5 contributors. Announcement on official website.

extrinsic reward, and *I* also has very unique content. In addition, we included three versions each focusing on a particular basic need (*nA*, *nR*, *nC*) to investigate if there are any significant differences among them.

To verify that our designs align well with SDT, we performed an expert review following the card-sorting procedure introduced by Moore and Benbasat [27]. We created a pool of messages and asked three experts from the field of HCI and psychology to assign each message to one of the 6 groups. Afterwards, any ambiguous message was removed or edited in an iterative procedure.

4.2 Method and Measures

We asked MTurk participants to (1) answer two questionnaires about general-causality orientation and motivation to volunteer, (2) rate their personal preferences between pairs of all interface versions, and (3) answer some questions about the design. The data collected also included basic demographic information (age, sex, education) and the frequency of volunteering, if any. The specific prompt for the pair comparison task was, "in which version would you personally be more likely to contribute a new idea?" Participants were asked to pairwise-compare all of the seven versions.

For measuring relative preferences based on subjective attributes, pairwise comparison was used. Our experiment was fully balanced: all participants made every possible comparison [5], and the order of the comparisons was balanced by the method of a Latin square.

Before seeing the design, participants were given the following scenario to aid their imagination in this hypothetical task, by pointing out the importance and potential impact:

> You joined an application that aims to gather ideas to improve conditions in your community, e.g. the members of your university, workplace, or neighborhood. We want as many people as possible to participate, because not everyone knows about the troubles of everyone else. The officials have agreed to monitor the community's ideas and follow-up with concrete steps once an issue has been identified.

After the comparison task, participants were presented with the "winner" according to their ratings and asked to give a one-sentence explanation of why they thought that this was the best version for them (or not).

For the test of general-causality orientation, the General Causality Orientations Scale (GCOS) questionnaire was used along with the recommended scale and ordering [6]. Through 12 questions

with 3 possible responses each, participants rated the likelihood of responding to a scenario in a certain way on a scale of 1 to 7. These ratings are summed into three dimensions which represent orientations towards autonomy, relatedness, and competence.

To test participants' volunteering motivation, we used a questionnaire developed by Grano et al. [11] (Motivation to Volunteer Scale, MVS). It contains 24 prompts with ratings on a scale of 1 to 5 relating to the six types of motivation/regulation identified by the SDT. We also asked participants about how often they participated in volunteering activities in the past three months on average, on a three-point scale (never, less than once a week, about once a week or more).

As is standard practice in conducting research experiments on MTurk, we used a number of measures for quality control [29]. We used a set of trapping questions to check participants' attention. Both the GCOS and MVS questionnaires included bogus items requiring participants to answer with a certain rating. The pairwise comparison task included one repeated measure to check for consistency. We also measured the time-to-complete to detect outliers.

4.3 Participants

We collected responses from 150 U.S. based MTurk users. For completing the task, which was estimated to take up to 20 minutes, they were rewarded USD 3.50 (for a nominal hourly wage of USD 10.50). After eliminating responses that failed the attention checks or that were impossibly quick, 99 responses remained, with 35% female (slightly less than a usual U.S.-based MTurk distribution [21]). Attained education and age were distributed over the complete range with 18% having completed high shool, 35% college, 40% a Bachelor's degree and 6% Master's degree or higher. The age groups were 18% 18-26, 33% 27-32, 23% 33-40, 21% 41-55, 4% over 56. The collected demographic variables (age, sex, education) had no significant effects on the reported preferences.

4.4 Results

The distribution of all subscale scores follows the trend of data reported in previous work [6, 11], indicating validity of the results obtained from MTurk workers. In terms of GCOS, most participants scored highest on the Autonomy scale, followed by Control and Impersonal. The three scales are nearly normally distributed, with the Autonomy scores being skewed towards the upper end of the scale. As for MVS, participants tended towards the autonomous side of the spectrum (high scores on the identified, integrated, intrinsic subscales), but exhibited a broad range in all of the subscales.

To analyze the differences in preferences, we constructed a Loglinear Bradley-Terry model (following the examples of [13]) using the results of the pairwise comparison task. The model predicts likelihoods (worth estimates) for the different versions and can be fitted using a generalized non-linear model (Poisson family).

Looking at the preference ratings without taking into account additional factors, our hypothesis H1 is supported, as there was no strong consensus on the "best" version. All versions were chosen by at least some people. The fitted GNM model shows significant differences of all versions over the baseline ($p<0.001$), except version I ($p=0.146$ n.s.). In absolute numbers, 65 (63.1%) participants ranked version C the highest and 14 (13.6%) version A. The baseline (BL) and Impersonal-oriented (I) versions consistently ranked lowest and the other versions were somewhere in between. Even though there were some individual differences, based on these results, one might conclude that offering extrinsic rewards like in version C is preferable if only designing for the majority of our participants, at least when asking people to choose.

To answer if individual differences in preferences can be explained by motivation orientations (H2), we added subscales scores as subject covariates to the model. We created new factors for each test scale by binning scores into "low" and "high" groups corresponding to the lower and upper quartiles of the recorded scores. The model estimates support our hypothesis H2, with the most

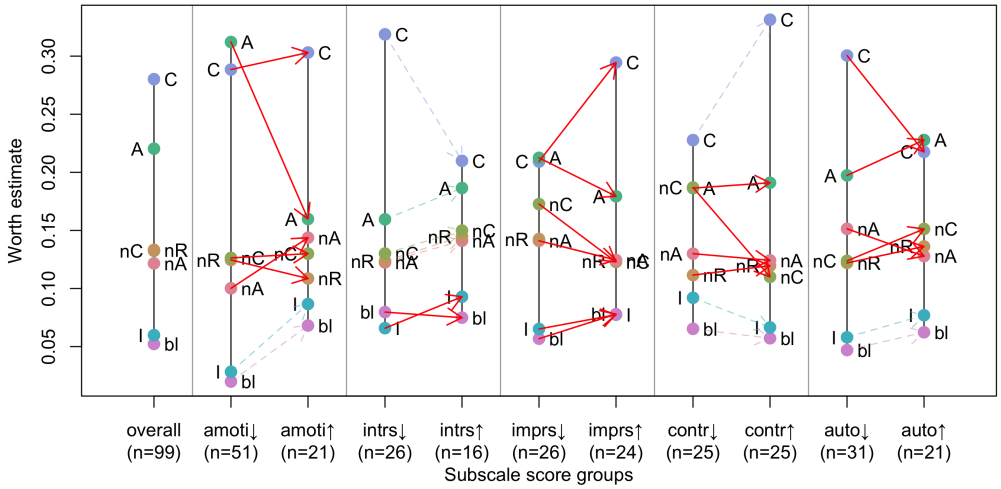


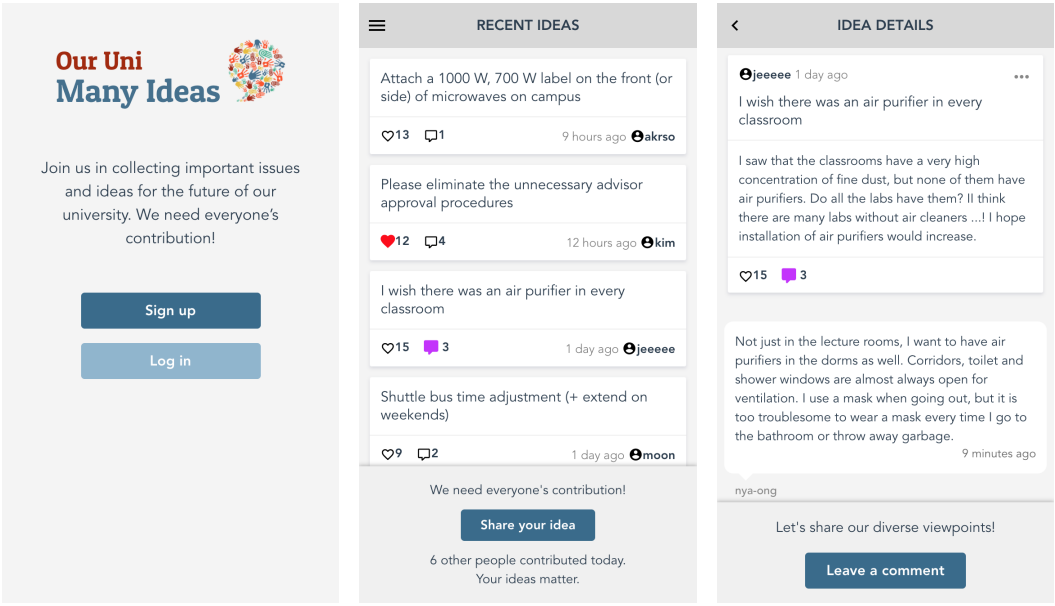
Fig. 3. Version preference estimates of Bradley-Terry model of pairwise comparison data, overall and for groups of low↓ or high↑ scores on the subscales amotivation, intrinsic, impersonal, control, autonomy. Solid red arrows indicate rank changes. For each of displayed scales, the changes in both A and C versions were statistically significant (p<0.01). Omitted subscales showed no significant differences.

significant differences between the A and C versions (Figure 3). Regarding GCOS, participants with a low Impersonal score preferred $A \succ C \succ nC \succ nR \succ nA$, whereas participants with a high score preferred $C \succ A \succ nA \succ nR \succ nC$. As expected from the theory, the Autonomy scale showed significant effects towards the opposite direction. The Control orientation scale also had a strong correlation with preference for version C. Regarding MVS, the two scales at the end of the spectrum had the strongest effects (Amotivation score low: $A \succ C \succ nC \succ nR \succ nA$, and high: $C \succ A \succ nA \succ nC \succ nR$; Intrinsic score low: $C \succ A \succ nA \succ nC \succ nR$, and high: $C \succ A \succ nR \succ nC \succ nA$). These results lead us to the following claims:

- Overall, external control in the form of rewards (C) is preferred, however for users with high Autonomy-orientation, low Control-orientation, low Impersonal-orientation, or low Amotivation scores, an autonomy-supportive (A) design may be preferable.
- If external rewards (C) are not possible, supporting all three needs for autonomous motivation (A) is generally preferable over supporting only a subset of needs.
- If supporting all three needs is impossible, Competence-need (nC) or Relatedness-need (nR) supportive messages may be more effective for users with low Impersonal, low Amotivation, or high Autonomy scores, whereas Autonomy-need (nA) supportive messages may be more effective for the opposite cases.

The free-form explanations given by the participants serve as further evidence for our hypotheses. Participants cited various reasons for liking different versions, and the reasons were aligned with our expectation. Participants who preferred version A mentioned reasons such as "*It looks more friendly*" and "*Making things better for everyone sounds like the best plan overall.*" A participant who preferred version C said, "*The chance of winning makes me more compelled to participate and try harder*".

Even though only a few people preferred the baseline and version I, they made sense of their choice: "*It's very simple and it doesn't insult the person contributing ideas by talking down to them*" (version BL) and "*I prefer a statement that doesn't try to make me feel guilty for not sharing an*



(a) Home page before login with branding and main goal of the app.

(b) Main feed of ideas with personalized call-to-action to participate.

(c) Idea detail view with ways to engage by commenting or showing support.

Fig. 4. Case study application's main screens. Displayed contents are examples of real user contributions. The university name is replaced with a generic name here.

idea" (version I). These versions might only appeal to a minority of users, but it could be worth investigating supporting them, as well.

4.5 Limitations

The results from Study 1 have a number of limitations. First, the preferences are all self-reported. Asking people what they think they will do in a hypothetical scenario and observing actual behavior can yield very different results, an effect known as *hypothetical bias* [28]. The MTurk population is also possibly non-representative for the general population, as they are paid workers. While this study was controlled for some demographic variables, there are other factors than might have influenced preferences, such as income level and geographical location. Additionally, we only looked at a single component instead of a more complete environment. The manipulations were verified by experts, but additional empirical manipulation checks would be desirable. To address some of these issues, in particular the hypothetical bias and paid nature of MTurk workers, we designed a real-world app and conducted a user study, explained in the next sections of this paper.

5 DESIGN OF A RESEARCH PROBE

We designed a mobile web application to use as an apparatus in a deployment study. It aimed to be a basic example of a crowd-civic system which allows for voluntary participation of the crowd regarding social matters of a community. Some of its main screens are depicted in Figure 4.

5.1 Design Goals

Based on our preliminary study and drawing from previous work on crowd reporting systems [36] and online community design [19], we identified the following minimum requirements:

- Shared, attainable, actionable goals with real impact; empowerment through large numbers
- Small but relevant individual contributions towards these shared goals
- Direct interaction between members of the crowd, for democratic deliberation

Additionally, we formulated the following design goals to make it feasible to run a short-term deployment study using this system:

- User tasks that are general and simple enough to not require special knowledge and enable meaningful interaction within a short period of time, and that are also unrestricted in quantity and quality to be able to measure differences in engagement
- General basic design allowing for addition of different motivation-supportive messaging strategies in multiple places

5.2 Implementation

To satisfy these requirements and goals, we developed Many Ideas, a system supporting communities to report problems and generate ideas to improve conditions for its members. Users share their views through creating issues, discussing, or voting. The system is based mainly on an altruistic motivation to help "the greater good", to improve the lives of every member of that community. Communities can be workplaces, neighborhoods, or other kinds of shared spaces. For our study, we decided to deploy the system for the members of one university. As they often live, study, and work here, students, faculty, and staff spend a considerable amount of time in a shared environment, making it a prime target for this sort of application.

Many Ideas offers four key interactions: contributing your own idea, browsing ideas posted by others, commenting on ideas, and expressing support by pressing a heart button. Browsing is supported by a feed of ideas on the main screen after login, which is sorted like the "What's hot" section on Reddit (reverse chronological, with popular posts boosted to the top for some time). This is to strike a balance between showing popular ideas first and upvote inflation, which would bury unpopular posts too quickly.

The application was implemented as a responsive web application using Vue.js and a backend server using the Django framework. To support open research and enable future experiments using this app as a basis, the source code is available on GitHub at graycoding.com/manyideas.

6 STUDY 2: LIVE DEPLOYMENT

To address the shortcomings of the first study, we ran a real-world campaign to measure effects of designs based on the self-determination theory (SDT) on volunteering behavior. In an open call, members (students, faculty, and staff) of a large technical university in Korea were invited to participate in a community of voluntary participation to gather ideas to improve conditions on campus, following the design introduced in section 5.

6.1 Hypotheses

To understand how engagement in civic participation applications can be affected by design based on SDT, we formulated the following hypotheses.

H1: Both autonomous and controlled motivation strategies are effective in increasing one-time, short-term voluntary participation.

H2: Personalized motivational support is more effective for increasing participation than either strategy applied unconditionally.

H3: External regulation in the form of monetary rewards is effective in increasing one-time, short-term voluntary participation, but detrimental to extended civic engagement.

6.2 Recruitment Campaign

We ran a promotion campaign through multiple on-campus channels, including physical posters, mailing lists, and posts on school online communities and social networks. The public campaign deliberately did not specify any rewards, to minimize the potential confounding of autonomous motivation. We tried to attract attention by using catchy slogans and raise curiosity by including vague messages like "There are various rewards for participation. Sign up to find out more!" After the first 3 days, we also sent an email reminder to every user so far to encourage them to revisit the app. On all promotional material, we indicated that the incentive period will conclude at a specific date (8 days after launching the study).

6.3 Experiment Conditions

In light of our previous results, we narrowed down the space of supported motivation strategies to controlled and autonomous motivation. In Study 1, these two conditions showed the biggest differences. Thus, we designed three different message strategies for the application: a baseline, an autonomous-motivation supportive strategy, and a controlled-motivation supportive strategy.

We envisioned two distinct tangible rewards for users, reflecting the internalized and external regulatory styles associated with these kinds of motivations. This choice was a result of an iterative design process in which we evaluated rewards feasible in our campus setting with user interviews. We wanted to offer rewards that are salient and align with SDT. Other rewards would be possible, but what we could offer was constrained by practical constraints, so we settled on these:

- *Idea curation*: For the autonomy-supportive treatment, we offered to curate three "promising ideas" that we will publicize to the whole school and also take active steps towards their implementation.
- *Cash prize*: For the control-supportive treatment, we offered a raffle of 10x KRW 20,000 (ca. USD 18.75) cash prizes among all active users.

The rewards were chosen in accordance with SDT. The *cash prize* is an external reward tied to a single user. It does not create an explicit connection with the amount and quality of a user's participation. Everyone was eligible, as long as they interacted in any way. On the other hand, the *idea curation* reward is related to intrinsic motivations and identified or integrated goals, such as real-world impact and recognition of the importance of one's contribution. Even if there can be some hypothetical monetary value of ideas being implemented, any tangible reward in that condition is not something that the user receives directly, but rather something their idea, or effectively the community, benefits from. We also made sure not to promise that ideas are actually implemented, but only to be considered for further deliberation.

Both rewards were promoted in-app according to the treatment a user was assigned to, i.e. users were only aware of one of the rewards. For the baseline treatment, a section on the app's FAQ page explained that there were rewards for which everyone was eligible, but that we wouldn't disclose them yet. The two other treatment conditions also contained an explanation that even though there might be different rewards, every user will unconditionally be eligible for all rewards.

The two strategies, supporting either controlled or autonomous motivation, were added to the design of our research probe. We selected 6 key elements from the application (e.g. a call-to-action component on the main page like in the first study, as well as several other instructional messages; see Appendix for the exact content) that we created alternative messages for, relating to support of

the different motivations. We aimed to keep these treatments simple and salient to maintain the connection to the theory and to not confound different motivational processes as much as possible.

All users were Korean native speakers, and the complete application including all prompts and motivational messages was translated to Korean by a group of native speakers under the authors' guidance.

6.4 User Assignment to Treatment Groups

We intended to classify users into groups that are likely to be affected by autonomous or controlled motivation based on information gathered at the time of registration to increase control and avoid confusing users with changing messages. Given the voluntary and quick-use nature of the app, we did not want to subject users to completing the complete questionnaires, due to a fear of causing annoyance and early dropouts.

We inferred a decision tree from the data collected in Study 1, using the raw responses to all items as independent variables and the relative preference of the autonomy over the control-supportive version as the dependent variable. As the distribution between the two preference was very imbalanced (ca. 1:3), we used the method of Randomly Oversampling Examples [25] before partitioning the data using a Conditional Inference Tree. K-fold cross validation yielded an accuracy of 70% on a generated tree containing only 4 questions out of the original 63 questions. These four questions happened to be one each from the motivation subscales amotivation (from MVS), control, impersonal, and automonomy (all from GCOS) and all were intuitive. We decided to use this classifier. Later, we will discuss this choice.

6.5 Method and Measures

6.5.1 User study. Upon sign up, users were classified into one of two groups predicted to prefer one design over the over. They were then uniformly distributed between the three experiment conditions implementing different strategies based on SDT, as explained previously. In this setting, we collected several measures of engagement under different motivation-supportive conditions. The assignments were fixed per user (between-subjects design). This resulted in a 2x3 factor design: orientation (Autonomy, Control) x treatment (Baseline, Autonomy, Control). The goal is comparing "successful" (i.e. matching) personalization vs "unsuccessful" (i.e. unmatching) personalization versus Baseline assignment.

To evaluate our hypotheses, we collected interaction data during the deployment, including the number of posts, likes, comments, events, sessions, and time spent per session for each user. We also analyzed the contents users contributed in terms of length. Upon signup, we collected information about sex, age, and status in the university. The study spanned a period of 8 days.

6.5.2 Post-survey. To find out more about the individual differences between users and feelings towards continued voluntary participation, we distributed a questionnaire to every user after the end of the study period. While no general compensation was offered to users of the application, survey respondents were paid KRW 7000 (ca. USD 6.50) for their participation. The survey included the questionnaires for general-causality orientation (GCOS [6]) and motivation to volunteer (MVS [11]) as in Study 1. We collected this data to compare the groups assigned during the deployment, which were based on limited information. In addition, the survey contains open questions about why users joined or stopped using the app following by the AttrakDiff mini questionnaire [12] for general usability and some open-form feedback.

6.6 Results

6.6.1 Participants. Within 8 days of the live deployment, 120 users signed up for our application (94% undergraduate and graduate students). There were less women than men, 29%, however, this is higher than the overall 20% ratio of women in this university. Regarding age distribution, 74% were between 18-26, 19% 27-32, 6% 33-40, and less than 1% over 56. The collected demographic variables had no significant effects on any of the engagement measures.

Users posted 72 ideas (0.6 per users), 62 comments (0.5 per user), and expressed their support using the heart button 357 times (3.0 per user). The majority, 79 users (66%), performed at least one of these interactions. In total, we recorded 528 unique visitors (visit-to-signup ratio of 23%). To track the source of traffic, we provided different URLs for different channels. Approximately half of the traffic came from our posts to the campus-internal discussion forum, 25% through Facebook shares, 10% through the posters, and the rest accessed the app directly, e.g. through word-of-mouth.

Upon signup, users were assigned to groups and treatment conditions using the result from four personality questions and a pre-computed inference tree, as described before. Our model assigned 72 (60%) of the users to the "controlled" group and 48 (40%) to the "autonomous" group. For each preferred design estimated by the system, we aimed to have a balanced distribution between the three treatment conditions. The absolute numbers in each combination were: 1. Autonomy group: Autonomy: 17, Baseline: 17, Control: 14; 2. Control group: Autonomy: 24, Baseline: 24, Control: 24.

We collected post-survey responses from 38 users (who had been assigned to these study treatments: 14 autonomy, 11 baseline, 13 control), a response rate of 32%.

6.6.2 Post-hoc Clustering. We used the post-hoc data from the full motivation questionnaires to cluster respondents into two groups using k-means. The clusters found align with theory intuition: one cluster has participants with relatively high scores on scales related to autonomous motivation, the other on scales related to control or amotivation. Comparing the original groups to the post-hoc ones yielded an accuracy of only 44%, indicating a bad fit of our original classification model.

6.6.3 ***H1****: General effectiveness of Control and Autonomous motivation strategies.* The monetary reward that was offered in the controlled-motivation condition was clearly successful. Users in that condition showed higher engagement compared to the two other conditions. They posted more ideas (posts per user: Control 0.94, Baseline 0.41, Autonomy 0.46; ANOVA $p<0.01$; post-hoc comparisons $p<0.05$ for Control treatment compared to Baseline; GLM for Poisson distributed count data) and wrote longer posts (average characters per post: Control 64, Baseline 33, Autonomy 36; n.s.). Differences between the Autonomy condition and Baseline were also positive, but much weaker and non-significant. This means that hypothesis H1 is partially supported.

6.6.4 ***H2****: Interaction Effects of Treatment and Group.* There were no significant interaction effects between treatments and user groups as classified from the signup questionnaire. None of the collected engagement measures showed statistically significant differences. Comparing the groups with the post-hoc clusters indicates that the classification based on four questions was inaccurate.

However, H2 is supported when taking post-hoc data into account. For survey respondents, we used the post-hoc groups instead of the ones assigned during the deployment for further analysis. Being in the treatment condition that supported the individual motivation orientation had positive effects on the number of interactions (Figure 5), whereas being in the umatching group showed negative effects (ANOVA for number of interactions $p<0.01$ for treatment, group, and interaction; GLM for Poisson distributed count data). Compared to each group's baseline median, for control-motivated participants, we counted 42% more interaction events when in the control treatment and 20% less events when in the autonomy treatment (Autonomy: 27, Baseline: 33, Control: 47; multiple comparisons with Tukey method, all $p<0.01$). For autonomy-motivated participants, we counted 4%

more events when in the autonomy treatment and 45% less events when in the control treatment (Autonomy: 44, Baseline: 42, Control: 20; multiple comparisons n.s.). Though not significant, other collected measures show similar trends. For example, users in the respective motivation-supportive treatment contributed ideas with more content (median baseline for both groups: 54 characters per post; control group in control treatment +93, in autonomy treatment -54; autonomy group in autonomy treatment +27, in control treatment +6). The overall data suggests that SDT-related personality scores are correlated with actual behavior, but should be regarded with caution due the low number of responses (N=38, with 5-7 users per group-treatment combination).

6.6.5 ***H3****: Detrimental Effects of Controlled Regulation.* Several data collected during the post-survey suggests possible detrimental effects of the control-supportive treatment. Users in the controlled-motivation treatment rated an insufficient reward significantly higher as a reason for dropout (mean 4.15, compared to baseline mean 2.6; ANOVA p=0.13 n.s.), whereas the difference in the autonomy-treatment was non-significant. Taking interaction effects into account, users in the controlled-motivation group under autonomy-supportive treatment had the highest ratings of any group for this dropout reason (mean 5.4; ANOVA p=0.12 n.s.), whereas those under the baseline or control-supportive treatments showed no differences.

6.6.6 Qualitative Data. Ideas submitted by users were generally well written and covered a wide variety of topics, such as facilities, administrative procedures, and inclusion of minorities. Some of the most popular ideas were: *Install air purifiers in every lecture hall*, *Eliminate unnecessary approval procedures*, and *Introduce a vegetarian day in the cafeteria*. The "report inappropriate content" feature that we had implemented in fear of spam or trolling was not used even once.

We thematically analyzed the survey data and highlight only some themes of interest here, as the data is limited in number (N=38, and not every participant gave responses to all questions as some were optional). For the analysis, we first translated the responses to English and then coded them into themes with the help of colleagues not involved in the research via open coding.

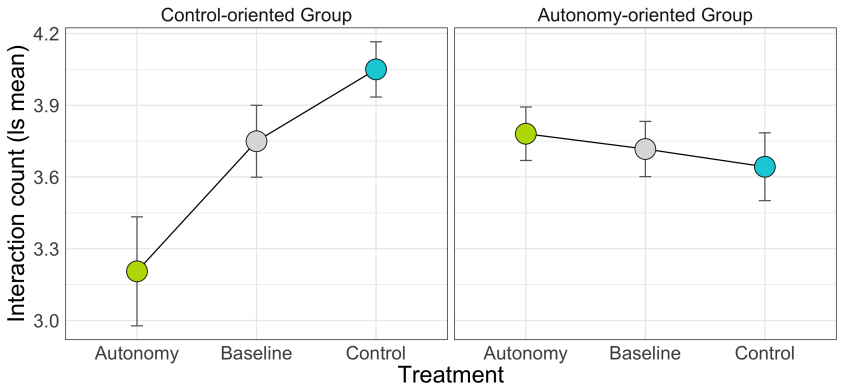


Fig. 5. Interaction counts by post-hoc group and treatment (N=38). Motivation-supportive treatments (i.e. Control treatment for Control-oriented group and Autonomy for Autonomy-oriented) resulted in an increase of interactions whereas being in the mismatching treatment had negative effects compared to the Baseline treatment. Least-squares means with error bars for 95% CI, GLM for Poisson distributed count data. ANOVA p<0.01 for treatment, group, and treatment x group; post-hoc Tukey's test for pairwise comparison: left-hand side (Control-oriented) all p<0.01, right-hand side (Autonomy-oriented) n.s.

Participants reported various reasons why they chose to join the application and why they quit. Answering an open-form question, several people mentioned they were attracted by the opportunity to contribute to real change. Some said they found it interesting to see what others think. One person mentioned that the app could fill a void by offering a place to publicly express ideas and be credited for them. Another person highlighted the importance of listening to diverse opinions, even though they might seem "strange and stupid". A similar sentiment was visible in another respondent, who said it was "interesting to see others' ideas, regardless of how feasible they are." These responses reveal diverse motivations to join. As for dropout reasons (likert-style questions), the highest ranked ones were "It was hard to imagine what my contributions will be useful for" and "I didn't have more time." When comparing the data by treatments, more users in the Control treatment cited "I thought I had contributed enough" as a dropout reason than others, and more users in the Autonomy treatment cited "The content wasn't interesting enough." These differences were not statistically significant, but nevertheless hint at possibly important differences.

Multiple participants mentioned they wanted to see actual evidence of the system's impact, for example updates from the team or feedback from officials. As highlighted in previous research [36], this would be a key factor for ongoing success.

6.6.7 Usability of the App. A poor overall design might have stronger effects on participation than any of the treatment conditions. To rule this out, we asked survey respondents to rate the usability and attractiveness of the app and report any issues. The results imply that usability issues are not likely to have been major source of confounds. In terms of quality as measured using the AttrakDiff questionnaire [12], the application scored medium-highly, with an average pragmatic quality of 5.3 (std 0.8), hedonic quality of 4.5 (std 0.7), beauty of 5.2 (std 1.2), and goodness of 5.7 (std 1.0) on 7-point scales. There is some room for improvement in terms of attractiveness, but generally people liked the application and did not report any major problems.

7 DISCUSSION AND FUTURE WORK

Our results show some evidence for the applicability of the Self-Determination Theory to the design of crowd-civic systems, but there are still a number of open challenges regarding personalization. In this section, we discuss our findings and limitations, along with possible future work.

7.1 Benefits and Challenges of Theory-based Design

The self-determination theory has proven to be a useful perspective for designing applications dealing with voluntary participation. Learning from theory, as opposed to examples, can broaden the designer's view. With this work, we are presenting another avenue of how theory-based design can be applied in the context of crowd-civic design.

However, translating theory into concrete designs has been challenging. Validated theories in controlled laboratory environments may not perform as expected in an uncontrolled field experiment due to the interaction between a variety of factors. Our attempt of translating self-determination theory into distinct motivation-supportive messages was our best effort, but due to the subjective nature of motivation and behavior, it is not guaranteed that every design is perceived by the participants exactly as theorized. It is unlikely we can ever find the "perfect" solution here, and we have to rely on incremental improvements and continuous measurements.

Moreover, theory is often manifested in more than one concrete implementation. In this work, we used the Motivation to Volunteer (MVS) and General-causality Orientation Scale (GCOS) questionnaires to capture two different constructs from the same theory. According to our results, both of these showed some correlations to both self-reported and actual behavior. While previous studies implied that both constructs should be applicable to voluntary work [8, 11], in Study 2, MVS

has proven to be more useful to post-hoc classify users into two groups. It is possible that the MVS scale worked better here because its context is volunteering, whereas GCOS scale's context is mostly professional work. We had expected the GCOS' *autonomy* and *control* orientations to be closely related to a user's susceptibility of our two motivational treatments, but perhaps the tendency towards certain regulatory styles is more indicative in the context of voluntary participation. It is also possible that this is a matter of the population that we studied: well educated university members might differ more in their volunteering motivations than in their general orientations.

7.2 Limitations and Alternatives for Message Personalization

According to the results from our studies, personalization of messages may be possible: treatment conditions showed significant differences in both studies, and personality data obtained through full questionnaires showed possible interaction effects between treatment and individual motivation. However, our attempt in Study 2 to personalize based on a subset of questions, backed by findings from Study 1, was unsuccessful. A likely explanation is that the classification using a subset of questions (4 out of 63) did not accurately capture individual motivations and orientations. Reducing the number of questions was a design trade off we chose to reduce the risk of causing dropout or biasing users to think about their motivation. Successful personalization may depend on choosing the right method of data elicitation to feed the personalization algorithm. It is also possible that app users did not respond accurately to the personality test questions as they were taken out of a traditional survey context. The interface included an explanation to users that the questionnaire mattered, but we cannot be sure that that had the intended effect.

Further investigation of different ways of eliciting sufficient data at signup time to accurately classify users is needed. Alternatively, there may be ways to gather the data necessary to personalize without the need of the user's explicit input, for example by automatically inferring users' personalities from their social media profiles [9]. Furthermore, instead of one-shot personalization, we can use a dynamic system that learns the user's preferences while they are using the application. Similar attempts have been introduced in the context of online education [35]. This will require more different motivational strategies, as one wouldn't be able to change affordances in the mutually exclusive way we did in this study after the user has started using the system. We might assign users to a motivation-neutral treatment at first, and only introduce motivation support once the system learned enough about the user and is confident enough about its prediction.

However, automatic classification and personalization does not only have advantages. Users may be concerned about the involved privacy risks and scared by computers' increasing ability to judge and predict human behavior, and lament the lack of autonomy [34], something that self-determination theory clearly stipulates. Thus, personalization itself can be limiting, and its unintended side effects have to be investigated.

We may also consider a combination of personalization and customization by offering users a choice between different motivational affordances. This may mitigate the limitations of personality assessment while providing a form of autonomy to users. For example, a system might offer a choice between various affordances that support different kinds of motivations (similar to [15]). In turn, the user's choice could be used to infer their orientations and provide supportive messages in other parts of the system. However, this can introduce risks of undermining motivation: users who were originally autonomously motivated might be negatively affected by being offered a choice of an extrinsically motivated regulation.

7.3 Studying Crowd-civic Systems

Researching design that affects behavior of voluntary crowds is challenging. In a controlled setting like Study 1, effects may be more apparent, but at the risk of lacking ecological validity. The

reliability of self-reported intentions is also questionable due to the hypothetical bias [28]. In a natural environment like Study 2, we can expect greater validity, but measuring specific effects is harder due to a lack of control. By investigating our research questions in both settings, we attempted to provide a more balanced view. Research in crowd-civic systems is likely to benefit from both kinds of experiments. Furthermore, the complexity and intertwining of motivation factors in voluntary work make empirical studies in this field difficult. Civic engagement is also a long-term effect affected by people's firmly held beliefs. We targeted a minimum of one week for the study to balance this to some extent, but running longitudinal studies will be beneficial.

We have gathered substantial evidence addressing our hypotheses, but we need additional experiments to verify them. For example, survey data indicates some support for the theory that providing too strong external regulation can diminish autonomous motivation, which will lead to satisficing and higher dropout. We need a more long-term and larger scale experiment to measure if people actually behave as they said.

In some ways, our research probe has succeeded in exemplifying a basic crowd-civic system. Users joined our platform with diverse motivations and they interacted with each other as a crowd to achieve civic goals. Both the MTurk and the campus community populations, though being biased each in their own way, showed a wide range of motivations. As the underlying theory of motivation is universal, we argue that the idea of motivation-supportive design is likely to generalize to other populations and other types of systems in this domain. However, the concrete implementation will vary and require separate testing, as motivation-support is not only dependent on the individual, but also on the specific task and overall goal.

Our results make a contribution to the increasing calls for designing with diverse crowds in mind [16]. Civic platforms should support multiple motivations, not only in terms of different "factors" [2] (such as moral obligations, learning, fun, and recognition), but in terms of different regulatory styles relating to self-determination. We argue that this can increase not only overall participation, but also representativeness and inclusiveness. In this work, we showed how people with different motivations might be treated more supportively. Our methodology did not account for other diversity factors such as balanced participation of people of different genders, income levels, or from different geographical locations or cultural backgrounds. Future work could investigate how motivation support affects representativeness by these other factors, or how it can help gathering a representative variety of opinions.

8 CONCLUSION

In this paper, we investigated the effects of motivation-supportive messages in crowd-civic systems. Informed by the Self-Determination theory, we created several messages for a number of tasks related to idea generation for social communities. In two studies, we gathered evidence for the applicability of this theory as well as a possibility of personalization. This work has implications for the design of systems that benefit from engagement from diverse groups. Especially in crowd-civic systems, this may increase not only overall participation, but also representativeness and inclusiveness.

ACKNOWLEDGMENTS

This work was supported by the X-Project Program (NRF-2017R1E1A2A02082072) and the Next-Generation Information Computing Development Program (NRF-2017M3C4A7065960) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT.

REFERENCES

[1] Tanja Aitamurto and Hélène Landemore. 2016. Crowdsourced Deliberation: The Case of the Law on Off-Road Traffic in Finland. *Policy & Internet* 8, 2 (2016), 174–196.

[2] Tanja Aitamurto and Jorge Saldivar. 2017. Motivating Participation in Crowdsourced Policymaking: The Interplay of Epistemic and Interactive Aspects. In *Proceedings of the 21st ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM.

[3] Kirsten Boehner and Carl DiSalvo. 2016. Data, Design and civics: an exploratory study of civic tech. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2970–2981.

[4] Irma Borst. 2010. *Understanding Crowdsourcing: Effects of motivation and rewards on participation and performance in voluntary online activities.* Number EPS-2010-221-LIS.

[5] Herbert Aron David. 1959. The method of paired comparisons. *Fifth Conference on the Design of Experiments in Army Research Development and Testing* (1959).

[6] Edward L Deci and Richard M Ryan. 1985. The general causality orientations scale: Self-determination in personality. *Journal of research in personality* 19, 2 (1985), 109–134.

[7] Menno Deen. 2015. *GAME Games Autonomy Motivation & Education: How autonomy-supportive game design may improve motivation to learn.* Ph.D. Dissertation. Technische Universiteit Eindhoven, Eindhoven, NL.

[8] Marylène Gagné and Edward L Deci. 2005. Self-determination theory and work motivation. *Journal of Organizational behavior* 26, 4 (2005), 331–362.

[9] Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems*. ACM.

[10] Lewis R. Goldberg. 1992. The Development of Markers for the Big-Five Factor Structure. *Psychological assessment 4* (1992).

[11] Caterina Grano, Fabio Lucidi, Arnaldo Zelli, and Cristiano Violani. 2008. Motives and determinants of volunteering in older adults: An integrated model. *The International Journal of Aging and Human Development* 67, 4 (2008), 305–326.

[12] Marc Hassenzahl and Andrew Monk. 2010. The inference of perceived usability from beauty. *Human–Computer Interaction* 25, 3 (2010).

[13] Reinhold Hatzinger, Regina Dittrich, et al. 2012. Prefmod: An R package for modeling preferences based on paired comparisons, rankings, or ratings. *Journal of Statistical Software* 48, 10 (2012), 1–31.

[14] 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, 823–835.

[15] Yuan Jia, Bin Xu, Yamini Karanam, and Stephen Voida. 2016. Personality-targeted gamification: a survey study on personality traits and motivational affordances. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM.

[16] Eunice Jun, Gary Hsieh, and Katharina Reinecke. 2017. Types of Motivation Affect Study Selection, Attention, and Dropouts in Online Experiments. In *Proceedings of the 21st ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM.

[17] Rafal Kocielnik and Gary Hsieh. 2017. Send Me a Different Message: Utilizing Cognitive Space to Create Engaging Message Triggers.. In *Proceedings of the 2017 ACM Conference on Computer-Supported Cooperative Work and Social Computing*.

[18] Richard Koestner, Gaëtan F Losier, Robert J Vallerand, and Donald Carducci. 1996. Identified and introjected forms of political internalization: Extending self-determination theory. *Journal of personality and social psychology* 70, 5 (1996).

[19] Robert E Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. 2012. *Building successful online communities: Evidence-based social design.* Mit Press.

[20] Andrew Mao, Ece Kamar, Yiling Chen, Eric Horvitz, Megan E Schwamb, Chris J Lintott, and Arfon M Smith. 2013. Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing. In *First AAAI conference on human computation and crowdsourcing*.

[21] David Martin, Sheelagh Carpendale, Neha Gupta, Tobias Hoßfeld, Babak Naderi, Judith Redi, Ernestasia Siahaan, and Ina Wechsung. 2017. Understanding the Crowd: Ethical and Practical Matters in the Academic Use of Crowdsourcing. In *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*. Springer, 27–69.

[22] Brian McInnis, Alissa Centivany, Juho Kim, Marta Pobet, Karen Levy, and Gilly Leshed. 2017. Crowdsourcing Law and Policy: A Design-Thinking Approach to Crowd-Civic Systems. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'17 Companion)*. 355.

[23] Brian James McInnis, Elizabeth Lindley Murnane, Dmitry Epstein, Dan Cosley, and Gilly Leshed. 2016. One and Done: Factors affecting one-time contributors to ad-hoc online communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 609–623.

[24] Elisa D Mekler, Florian Brühlmann, Alexandre N Tuch, and Klaus Opwis. 2017. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior* 71 (2017),

525–534.
[25] Giovanna Menardi and Nicola Torelli. 2014. Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery* 28, 1 (2014), 92–122.
[26] Youngme Moon. 2002. Personalization and personality: Some effects of customizing message style based on consumer personality. *Journal of Consumer Psychology* 12, 4 (2002), 313–325.
[27] Gary C Moore and Izak Benbasat. 1991. Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information systems research* 2, 3 (1991).
[28] James J Murphy, P Geoffrey Allen, Thomas H Stevens, and Darryl Weatherhead. 2005. A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics* 30, 3 (2005), 313–325.
[29] Babak Naderi. 2018. *Motivation of Workers on Microtask Crowdsourcing Platforms*. Springer.
[30] Oded Nov and Ofer Arazy. 2013. Personality-targeted design: Theory, experimental procedure, and preliminary results. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 977–984.
[31] Anne C Rouse. 2010. A preliminary taxonomy of crowdsourcing. *ACIS 2010 Proceedings* 76 (2010), 1–10.
[32] Richard M Ryan and Edward L Deci. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology* 25, 1 (2000), 54–67.
[33] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. 2016. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 813–822.
[34] Eran Toch, Yang Wang, and Lorrie Faith Cranor. 2012. Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction* 22, 1-2 (2012).
[35] Joseph Williams, Na Li, Juho Kim, Jacob Whitehill, Samuel Maldonado, Mykola Pechenizkiy, Larry Chu, and Neil Heffernan. 2014. The MOOClet framework: Improving online education through experimentation and personalization of modules. (2014).
[36] Dieter Zinnbauer. 2015. Crowdsourced Corruption Reporting: What Petrified Forests, Street Music, Bath Towels, and the Taxman Can Tell Us About the Prospects for Its Future. *Policy & Internet* 7, 1 (2015), 1–24.

Received April 2018; revised July 2018; accepted September 2018

Table 4. Case study app components with motivation-supportive message treatment versions, in the order that users are likely to perceive them.

Baseline	Autonomy-supportive	Control-supportive
	(Main feed) Call-to-Action	
[Share your idea] We need everyone's contribution!	[Share your idea] <N> other people contributed today. Your ideas matter.	Become a Contributor of the Week! [Share your idea] Your chance to win $20!
	(New idea) Intro text	
(No text)	Our voice matters! Among all ideas submitted until April 11, we will handpick three promising ideas, present them to the whole school, and follow up with concrete steps to support their implementation! Together, we can have a real impact.	Among all contributors until April 11th, we will randomly select 10 members to win $20.
	(Main feed) Tutorial Message	
Welcome to Many Ideas for [Uni]! We're happy to have you here. On this page, you can see ideas by other members. Do you see anything you are interested in? Try tapping on a post.	Welcome to Many Ideas for [Uni], **where our combined voice can have a real impact**! We're happy to have you here. On this page, you can see ideas by other members. Do you see anything you are interested in? Try tapping on a post.	Welcome to Many Ideas for [Uni]! We're happy to have you here. On this page, you can see ideas by other members. Do you see anything you are interested in? Try tapping on a post. **Participating actively can get you a chance to win $20!**
	(View others' idea) Tutorial Message	
Here you can see details about the idea and comments that other people left. If you agree with this issue, how about showing your support by tapping the heart? If you have another opinion, try writing a short comment.	Here you can see details about the idea. **Discussing ideas and showing your support is important for our community.** If you agree with this issue, how about showing your support by tapping the heart? If you have another opinion, try writing a short comment.	Here you can see details about the idea. If you agree with this issue, how about showing your support by tapping the heart? If you have another opinion, try writing a short comment. **Remember, all active users have a chance to win $20!**
	(My ideas) Empty State	
Nothing yet. Post your first idea now!	Nothing yet. Share your ideas with the community now! Everyone's voice counts.	Nothing yet. Share your ideas for a chance to win $20!
	(FAQ) Why should I participate? section	
You can have an impact on improving conditions for everyone. We are also preparing several rewards for good ideas and active contributors. We will share the details with you later, but rest assured all active participants are always eligible for all rewards.	You can have a real impact on improving conditions for everybody! **[Reward description text similar to corresponding Intro text]** As part of our study, we also consider other rewards. We cannot show you that right now, but all active participants will always be eligible for all rewards.	

Tutorial Messages are only shown until the prompted action has been performed by the user for the first time.
Bold text in this table was also highlighted similarly in the application to make the treatment messages more salient.