Personality-Targeted Design: Theory, Experimental Procedure, and Preliminary Results

Oded Nov
Polytechnic Institute of New York University
Brooklyn, New York
onov@poly.edu

Ofer Arazy
University of Alberta
Edmonton, Alberta, Canada
ofer.arazy@business.ualberta.ca

ABSTRACT
We introduce a framework for personality-targeted design. Much like a medical treatment applied to a person based on his specific genetic profile, we make the case for theory-driven personalized UI design, and argue that it can be more effective than design applied equally to the entire population. In particular, we show that users’ conscientiousness levels determine their reactions to UI indicators of critical mass. We created a simulated social recommender system in which participants answer a short personality questionnaire and are subsequently presented with a picture of a pet that purports to be the “best match” for their personality. We then manipulated the UI by providing indicators of the existence and the lack of critical mass. We tested whether the interaction between personality and UI design affects users’ participation. The findings validate our hypothesis, showing that manipulation of the critical mass indicators affect high-conscientiousness and low-conscientiousness participants in opposite directions.

Author Keywords
Social psychology; theory-driven design; personality; conscientiousness; critical mass; diffusion of responsibility.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design

INTRODUCTION
Recent years have seen a substantial growth in the scale and scope of social technologies, which are based on technology-mediated social participation [45]. Consequently, researchers have been stressing the need to develop theories that will guide practice, which is to date, guided to a large extent by intuition, trial and error [29, 45]. Current studies on the factors that drive social participation tend to take one of the following approaches:

- **Experimental, design-based research** in which controlled experiments are used to test the effect of design features (primarily focusing on UI design) on user performance. Typically, such studies build on social psychology theories to inform design choices [6, 12, 13, 17, 30, 47]. This approach enables researchers to draw conclusions about design effectiveness. However, to date, studies based on this approach did not take into account differences in users’ idiosyncratic attributes, such as their personalities or motivations; in other words, to follow a medical metaphor, this is equivalent to testing the effectiveness of a medical treatment by providing it to all patients, regardless of their specific genetic make-up.

- **Individual differences research** in which contributors’ personal attributes, such as motivations or personality traits are identified, often using self-report surveys, and correlated with data on participation levels quality and quantity [25, 36-38, 49, 50]. This approach enables researchers to identify a relatively large number of relevant personal attributes directly, using validated instruments, and then develop and test theories about the relationships between personal attributes, attitudes and participation. A disadvantage of this approach, however, is that the research design is often cross-sectional, and causality is difficult to establish. Furthermore, such studies usually do not involve direct testing of system design effectiveness.

The objective of the present study is to develop, implement and test a novel approach, which builds on the strengths of these two approaches. We combine controlled experiments, design manipulations, surveys, and system data, to test the effectiveness of design features that target participants’ idiosyncratic personal attributes. While the proposed approach is applicable in a variety of settings, in the present study we focus on social recommender systems and participants’ rating and feedback activities.

We address the following general research question: are there differences in the effects of design interventions targeting users’ idiosyncratic personal attributes traits on users’ participation? Using the medical metaphor mentioned above, the proposed approach is analogous to a medical treatment that is applied to an individual based on his specific genetic profile. The extant research on
personalization [7, 18, 19, 21, 31], often focuses on users’ past experiences and interaction. Our approach here focuses on more fundamental user attributes such as personality traits.

To address the research questions, we developed for the present study a simulated online recommender system called PetMatch. PetMatch is presented to its users as a research project in need of feedback from users: its landing page invites participants to answer a very short personality questionnaire, and a simulated recommender system returns a picture of a pet that purports to present the “best match” for the participant’s personality (see Figure 1). The PetMatch environment facilitates a 2X2 factorial experimental design consisting of: (1) measurement of personal attributes (through a short questionnaire), (2) design intervention, and (3) opportunities for users to rate the accuracy of the recommendation and provide verbal feedback (see Table 1). As such, PetMatch serves as a platform on which to address research questions concerning the interactions between personal attributes, design choices, and user participation.

In the present study, we present findings from an experiment focusing on the interaction between the personality trait of conscientiousness (a known predictor of helping behavior in various social settings), design choices involving indicators of high and low levels of critical mass, and their joint effects on user participation.

<table>
<thead>
<tr>
<th>Experimental Interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personal Attributes</strong></td>
</tr>
<tr>
<td><strong>Low level of personal attribute</strong></td>
</tr>
<tr>
<td>Design intervention: low level</td>
</tr>
<tr>
<td>Outcomes for: low attribute X low intervention</td>
</tr>
<tr>
<td><strong>High level of personal attribute</strong></td>
</tr>
<tr>
<td>Outcomes for: high attribute X low intervention</td>
</tr>
</tbody>
</table>

Table 1. Experimental conditions: personal attributes and interventions.

**BACKGROUND**

The CSCW field has a long tradition of social psychology-driven design research, often using controlled experiments [6, 13, 17, 30, 47]. In such studies, the effectiveness of design features is tested in online environments. For example, drawing on social comparison theory, [6] showed in an online experiment that when users of MovieLens, a movie recommender system, receive information about others’ contribution, users below the median increase their participation, whereas those above the median do not decrease their ratings. Similarly, [30] and [33] found that users contributed more when they were reminded of their contribution’s uniqueness. A different approach to studying user participation involves personal attributes such as motivations or personality traits. In such studies, attributes are identified, often using self-report surveys, and correlated with users’ participation data [25, 49, 50]. For example, the “Big Five” personality factors were found to be useful predictors of internet use [35], and participation in social media sites [4, 8].

Participation in online environments was studied extensively, leading to the identification of various factors that encourage participation. Perceived critical mass – a user’s subjective belief that there is a large number of other users who participate in a community or adopt a new technology – was shown to have a positive effect on the user’s own participation behavior [34, 46, 52]. However, in contrast, extant social psychology research shows that the higher the number of people present in a situation or taking part in a collective effort, there is a higher likelihood of social loafing [28] and diffusion of responsibility [14, 20], such that each one of the people present feels less personal responsibility, and less compelled to help. This phenomenon applies to online settings as well [1, 5, 11].

While prior research focused on factors such as group homogeneity [40] and group interaction [16], we attempt to reconcile the contrasting evidence by focusing on the role of personality. Specifically, we study the role of conscientiousness (being responsible, dependable, planful, organized, and persistent, [2, 23]), one of the Big Five personality factors [9, 22]. Prior studies of personality and social behavior showed the important role the conscientiousness trait plays in explaining helping behavior that is relevant to the present study. Specifically, it was found that conscientiousness was the best personality trait predictor of organizational citizenship behavior [26, 43] (discretionary behavior which promotes the effective functioning of the organization but is not part of the formal reward system [42, 44]), and that it was negatively related to social loafing [26]. These two outcome variables are closely related to the dependent variable in our study, and thus we expect the relationship between consciousness and participation to generalize to the setting of our study.

Building on these prior findings and using our experimental platform, we test the hypothesis that perceived low level of critical mass will discourage diffusion of responsibility among participants characterized by high conscientiousness resulting in increased participation, while a perception of
the existence of critical mass will decrease participation in this population. This is because people characterized by high conscientiousness tend to be responsible and self-disciplined [10, 48], and therefore, we hypothesize, they will act more responsibly in the face of a request for help - such as the researchers’ request to rate users’ pet match as part of a research project - when they see that there are fewer others who may be available to do so (i.e. low critical mass experimental condition). When facing a situation in which there is an indication that others have already provided help (i.e. high critical mass condition), the need for help would seem less important, and highly conscientious people are expected to feel less obliged to help (resulting in increased diffusion of responsibility and decreased participation). Participants characterized by low conscientiousness, on the other hand, are more likely to exert effort when there is social pressure on them from others to do so [3, 24]. Therefore, we hypothesize, they will be more likely to participate when faced with an indication of a large number of other participants who already rated (i.e. high critical mass indicator), but less likely to participate in the absence of such indication (i.e. low critical mass indicator).

METHOD
In what follows we outline the details of the experimental procedure used in the present study.

Personal attributes condition: conscientiousness
Participants were recruited in the authors’ universities among undergraduate and graduate students. In addition, students were asked to share the invitation to participate in the study with their friends and family, and many of them shared the PetMatch link with their contacts via social media. The participants were led to believe that PetMatch is part of an academic research project involving the development of a technique to match respondents’ personality traits and the pet that would suit them most. This was communicated to participants in a recruitment email, and on the PetMatch site itself. Participants were not compensated for their participation.

The PetMatch landing page invites participants to answer a very short personality questionnaire. Self-report surveys are commonly used in social science research to identify personal attributes, and have been used extensively in HCI and CSCW studies [27, 35, 38, 39, 41, 51]. The two conscientiousness questionnaire items used a 7-point Likert scale and were adapted from the Ten Item Personality Instrument [23], a short version of the Big Five instrument which has been validated and tested numerous times in prior studies [15].

After filling out the short questionnaire and answering two demographic questions (age, gender), respondents were presented with their purported “best match” (see Figure 1): an image of an animal based on the responses to the survey questions, such that each combination of responses was consistently associated with a specific pet image. Unbeknownst to the respondents, PetMatch in fact presented pet images arbitrarily, with no real attempt to match images to personalities.

A post-hoc median split was performed such that respondents whose conscientiousness score was above the sample median were classified as high-conscientious and those below the median as low-conscientious respondents.

Design intervention: perceived critical mass
In addition to the image, respondents were also presented with information about the average rating for the particular pet image presented to them, as well as the number of ratings received so far (along the lines of UI design common on popular recommender systems such as Amazon or Netflix). Unbeknownst to the respondents, the average rating was fixed at 3.5 (out of 5), and the count and average rating indicators were in effect an experimental manipulation. The number of ratings received served as an indicator of critical mass. Perceived critical mass is a subjective and context-specific concept [32, 52], and we therefore set two levels of critical mass in the experiment, assigned randomly to participants: a high value of prior ratings (2,127) represented a high level of critical mass, and a low value (26 ratings) represented a low level of critical mass. These low and high values were established based on a pre-study in which we described the experimental design to three of our colleagues, and asked for their opinions on critical mass values which would be believable in the context of this study. The values of 26 and 2,127 represented the averages for the responses we received.

To further validate the low and high values in the experiment, we administered an additional experiment, this time using Amazon Mechanical Turk. In this experiment, participants were directed to a webpage describing a simple scenario fairly similar to PetMatch: in this scenario, the participant heard about a web site where users rate movies,
and visits it to check out a movie they are curious about. The participant then finds that the movie received the rating of 3.5 stars out of 5, based on X reviews (where X is manipulated by the researchers and is randomly assigned the values of either 2,127 or 26 ratings – the low and high critical mass PetMatch values). Having seen these ratings, the subject is asked to what extent they agree with a statement that the movie reviews website has reached a critical mass of users. Responses range from 1 (strongly disagree) to 4 (Neutral) to 7 (strongly agree) on a Likert scale. 78 people took part in this validation experiment. The low critical mass anchor (26 ratings) received the average score of 2.44 out of 7, and the high critical mass anchor (2,127 ratings) received the average score of 5.11. A t-test was used to compare the means, and difference between them was found to be significant (p<0.001). Furthermore, both scores were significantly (p<0.001) lower and higher (respectively) than the “Neutral” perceived critical mass value. The results therefore support the assumption that the two anchors represent high and low levels of perceived critical mass anchors.

In summary, at this stage respondents were assigned to one of the four experimental conditions described in Table 1.

Respondents’ participation (outcome variables):

In addition to the image indicators described, respondents were presented with two participation opportunities that would help the project: (1) a request to rate the quality of the match on a five-star scale, and (2) a request to provide verbal feedback: a comment or a link to a better match (see Figure 1). Respondents’ decision on whether to perform these actions or not served as a measure of the participation outcome variables.

To test the hypothesis, we performed a comparison between the average participation rates in all four conditions. In addition, to validate the hypothesis that the interaction between the dimensions (interventions x personal attributes) explains participation rather than just its separate dimensions, we performed two distinct bivariate correlations, one between participation and the critical mass manipulation, and the other between participation and conscientiousness. We also performed logistic regressions for both outcome variables, without the independent variables’ interaction.

RESULTS

Overall, 459 people used PetMatch, of which 53.2% were women. The average age was 31.2 (stdev=10.7). Cronbach’s alpha was 0.72, representing satisfactory reliability. Participation data of respondents in the experimental conditions were compared in order to identify the differences between the proportions of respondents who participated (by rating or by providing verbal feedback) vs. those who did not. Overall, 46.8% of the participants provided rating and 21.7% provided verbal feedback. However, patterns of participation differed across experimental conditions. These are discussed below.

Logistic regression analyses were performed to test a full model with two independent variables, the interaction between them, as well as demographics (age and gender). The dependent variables were (i) whether respondents rated or not (Rated=1, Not Rated=0), and (ii) whether respondents provided verbal feedback (Feedback=1, No Feedback=0). The independent variables were conscientiousness level (high and low - above and below the median, respectively) and perceived critical mass (high and low). The results of running the full regression are presented in Table 2. Logistic regression: rating The overall model for Rated was significant (χ²(5)=13.0, p<.05). The main effects of conscientiousness and perceived critical mass on rating were both positive but only conscientiousness was significant. The interaction between the independent variables was negative and significant.

<table>
<thead>
<tr>
<th>IV</th>
<th>B</th>
<th>S.E.</th>
<th>Wald χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.012</td>
<td>0.011</td>
<td>1.165</td>
<td>.281</td>
</tr>
<tr>
<td>Gender</td>
<td>-2.80</td>
<td>0.246</td>
<td>1.292</td>
<td>.256</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>1.023</td>
<td>0.383</td>
<td>7.131</td>
<td>.008</td>
</tr>
<tr>
<td>Critical mass</td>
<td>0.518</td>
<td>0.395</td>
<td>1.716</td>
<td>.190</td>
</tr>
<tr>
<td>Conscientiousness X Critical mass</td>
<td>-1.115</td>
<td>0.506</td>
<td>4.860</td>
<td>.027</td>
</tr>
</tbody>
</table>

Table 2. Logistic regression: rating.

As expected, the interaction (see Figure 2) was such that participants characterized by high conscientiousness were more likely to rate when there was perceived low critical mass (i.e. when participants perceived that few others were available to rate). Low conscientiousness participants, on the other hand, were more likely to rate when faced with an indication of a large number of other participants (i.e. perceived high critical mass).

![Figure 2. Likelihood of rating in each of the experimental conditions (horizontal lines represent error bars).](image-url)
An alternative logistic regression model using verbal feedback as a dependent variable yielded similar results (see Table 3): the overall model was significant ($\chi^2(5)=16.9, p<.01$). The main effects of conscientiousness and perceived critical mass were positive, but conscientiousness was weakly significant and perceived critical mass was not significant. The interaction between the independent variables was significant and negative.

<table>
<thead>
<tr>
<th>IV</th>
<th>B</th>
<th>S.E.</th>
<th>Wald $\chi^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.041</td>
<td>.012</td>
<td>10.837</td>
<td>.001</td>
</tr>
<tr>
<td>Gender</td>
<td>-.098</td>
<td>.291</td>
<td>.114</td>
<td>.736</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.805</td>
<td>.451</td>
<td>3.191</td>
<td>.074</td>
</tr>
<tr>
<td>Critical mass</td>
<td>.430</td>
<td>.469</td>
<td>.840</td>
<td>.359</td>
</tr>
<tr>
<td>Conscientiousness X Critical mass</td>
<td>-1.200</td>
<td>.603</td>
<td>3.965</td>
<td>.046</td>
</tr>
</tbody>
</table>

Table 3. Logistic regression: verbal comment.

Here too, the interaction was such that participants characterized by high conscientiousness were more likely to provide verbal feedback when there was perceived low critical mass, and low conscientiousness participants were more likely to contribute when faced with an indication of a large number of other participants (see Figure 3).

Overall, for both outcome variables the interactions were significant in the expected direction, thus supporting our hypotheses.

The bivariate correlations, performed between participation and the critical mass manipulation, and between participation and conscientiousness showed no significant correlations. Furthermore, logistic regressions performed for both outcome variables without the interactions, revealed insignificant main effects for both independent variables. Both analyses confirmed the hypothesis that it is the interaction rather than the main effects that explains participation.

**DISCUSSION AND CONCLUSIONS**

The design intervention we tested was shown to have a different effect on people with different levels of the conscientiousness personality traits, illustrating the potential value of personality-targeted design. Future research in this direction may help transforming the way we understand technology-mediated social participation and help improve the ways social systems attract, retain and encourage members’ participation.

The findings of the present study have implications for developing and managing social systems, and in particular, recommender systems. A direct implication from our findings is that in recommender systems, when considering the user population at large, we cannot necessarily expect indicators of community’s activity (e.g. critical mass) to affect participation for the entire population. This insight puts into question the common practice by website designers to provide indicators of others’ behavior. Only when we consider users’ personal attributes do we see the effects of critical mass indicators. Interestingly, the effects for highly-conscientiousness and less-consciousness users were in opposite directions, stressing further the need to get an indication of users’ personality before presenting community activity indicators. More broadly, our proposed approach to personalized design highlights the need to tailor design features to personality, such that the front end and back-end design features fit users’ particular personal attributes. Such tailoring may simply involve making visible, or not, information such as the number of other users who already participated. Personalized content delivery is now common practice, yet few – if any – online communities provide personalized UI based on user’s personal attributes. A practical implication of our approach is the need to get an indication of users’ personality. While the survey approach used in this study is possible, there may be less intrusive methods for automatically detect aspect of users’ personal attributes. Such methods can try to measure invariable traits and dispositions (i.e. creating a user profile), or even capture users’ transient preferences and attitudes such that the UI is not only personalized across users but also tailored to a users’ particular attitude at particular point in time.

**ACKNOWLEDGMENTS**

This work was partially supported by NSF award IIS 1149745. We would like to thank Erez Heiman for his contribution to the site design, and Durga Bhavani and Manasa Ranjitha Alla for research assistance.

**REFERENCES**


