Crowd-Designed Motivation: Motivational Messages for Exercise Adherence Based on Behavior Change Theory

Roelof A. J. de Vries¹, Khiet P. Truong¹, Sigrid Kwint², Constance H.C. Drossaert³, Vanessa Evers¹
¹Human Media Interaction, University of Twente
{r.a.j.devries, k.p.truong, v.evers}@utwente.nl
²Psychology, Health and Technology, University of Twente
s.j.m.kwint@student.utwente.nl, c.h.drossaert@utwente.nl

ABSTRACT
Developing motivational technology to support long-term behavior change is a challenge. A solution is to incorporate insights from behavior change theory and design technology to tailor to individual users. We carried out two studies to investigate whether the processes of change, from the Transtheoretical Model, can be effectively represented by motivational text messages. We crowdsourced peer-designed text messages and coded them into categories based on the processes of change. We evaluated whether people perceived messages tailored to their stage of change as motivating. We found that crowdsourcing is an effective method to design motivational messages. Our results indicate that different messages are perceived as motivating depending on the stage of behavior change a person is in. However, while motivational messages related to later stages of change were perceived as motivational for those stages, the motivational messages related to earlier stages of change were not. This indicates that a person’s stage of change may not be the (only) key factor that determines behavior change. More individual factors need to be considered to design effective motivational technology.

Author Keywords
Crowdsourcing; motivational messages; exercise adherence; behavior change theory; Transtheoretical Model; processes of change; stages of change;

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces - Theory and methods; J.4 Social and Behavioral Sciences: Psychology

INTRODUCTION
In recent years, HCI research has increasingly focused on motivational technology to help people change, for instance, their exercising behavior [10]. The potential benefits of motivational applications in healthcare and well-being are tremendous because of the large number of people who can be reached through mobile technology. This has motivated many researchers to design and develop motivational technology for health.

Design challenges
Challenges arise for HCI researchers aiming to design technology that motivates people to adopt more healthy behaviors long-term. One of these challenges is the evaluation of motivational technology. Although many HCI researchers aim to promote long-term behavior change through their technology, evaluations confirming long-term behavior change are rarely carried out [17]. Practical limitations (e.g., how to track a large number of subjects for more than a few months) and conceptual definitions (e.g., how to measure behavior change, how many months can be considered ‘long-term’) are some of the aspects researchers struggle with. Another challenge is that there is no method available to translate behavior change theories and models into concrete interaction designs to be used in practice. Also a challenge is to increase effectiveness of the motivational technology [33]. One way to increase effectiveness is to go beyond a one-size-fits-all motivational strategy. This can be achieved, for example, by personalizing or tailoring the strategies used in motivational technology to the stage of behavior change a person is in (e.g., still in the stage of thinking about doing regular exercise).

Behavioral change
According to the Transtheoretical Model (TTM) [36], behavioral change consists of five stages (i.e., the stages of change, see Table 1). These stages describe people’s willingness to change their behavior, ranging from long-term inactive (i.e., in the Precontemplation stage) to long-term active (i.e., in the Maintenance stage). Efforts to change behavior should closely match the stage that the person is in [39]. Developers of motivational technology who want people to change their behavior could use theories and models such as the TTM to tailor the information they provide to the stage the user is in. The TTM provides ten strategies to move through these stages of behavior change: the processes of change. Different processes are associated with different stages [27]. For example, Consciousness Raising (i.e., making someone aware of risks) will be
more effective for a long-term inactive person (i.e., someone in the Precontemplation stage), than for a long-term active person (i.e., someone in the Maintenance stage). We expect that when motivational technology uses messages that reflect or represent the processes of change that influence progression to the next stage, the messages will be more effective. This kind of personalization (e.g., tailoring to the user’s stage of change) can have a positive influence on exercise adherence [26].

**Aims**

Our goal is to design technology for long-term behavior change that is tailored to the user, based on well-established principles from behavior change theory. This is set in the context of the development of a smartphone-based application that motivates users to change their exercise behavior (such as running) and become regular exercisers.

In this study, we investigate how we can tailor motivational messages to the TTM’s stages and processes of change, and how people experience these messages. Can we represent the processes of change with motivational text messages? Will these messages relate to the stages of change like the processes of change do? Will users indeed perceive messages that are tailored to the stage of change they are in as motivating?

In the following sections, we report on relevant theory and previous work, methods, results and discussion for the first study, methods, results and discussion for the second study, and we end with an overall conclusion.

**BACKGROUND AND RELATED WORK**

In HCI and healthcare-oriented research, there is an increasing focus on technology that assists or encourages people to change their eating, exercising, or sleeping behavior [10], but also on technology to monitor, assist or change patients’ health-related behaviors. This is shown by recent review papers on mobile healthcare systems [19], obesity management through mobile phone applications [11], assistive technology to support healthy behaviors [15], information and communication technology-based interventions for promoting physical activity [23] and general health interventions using mobile phones [18]. These papers show that the use of text messages has been a common approach, but there is room for improvement in both grounding the technological approach with a theoretical foundation (i.e., behavior change theories or models) and tailoring to the end-user.

**Using text messages to motivate**

Using tailored text messages to influence someone’s behavior has proven effective in various contexts, for example physical activity [31]. Unfortunately, studies describing the development of such technology do not yet explain in detail how the researchers designed the motivational messages used [22]. Recently, studies have started to describe the process of acquiring motivational text messages. For example, Kaptein et al. [14], explained that two researchers thought of 42 text messages for six strategies to be tailored to the user’s susceptibility to persuasion: Patrick et al. [35] explained that they developed 3000 SMS and MMS messages to be tailored to the user’s preferences on timing and frequency of the messages; and Redfern et al. [35] explained that they designed 137 text messages (based on behavior change techniques) tailored to the user’s name. For all these and other studies, it is usually the authors or other experts who designed the messages. However, as is shown by Coley et al. [4], peer-to-peer designed text messages are more engaging and more relevant to the user compared to expert-written text messages. Therefore, we decided to use crowdsourcing to collect motivational messages.

**Crowdsourcing motivational messages**

Crowdsourcing involves employing a large number of people to contribute to a specific task. Over the last few years, researchers have been using online crowdsourcing platforms such as Amazon Mechanical Turk (AMT) for a growing variety of tasks; for example, for user-studies [16], graphical perception tasks [9], parallel prototyping tasks [5], evaluating user interfaces [42], but also for natural language processing tasks [2].

Crowdsourcing written transcriptions, translations or annotations (e.g., [28, 43, 12]) is a relatively common natural language processing task, but we were aware of only two related studies on collecting motivational messages. Coley et al. [4] crowdsourced peer-to-peer text messages to encourage users to quit smoking and compared these text messages to expert-written text messages. They found the peer-to-peer text messages to be more engaging and relevant to the user. They also found that the peer-to-peer designed messages reflected the same key theoretical concepts also addressed in the expert-written messages. To reduce alcohol consumption, Kristan and Suffoletto [20] also looked at peer-written text messages and evaluated expert-written text messages. Overall, they found that there were no universal positive attitudes towards any of their messages, which suggested the need for tailoring. By using crowdsourcing, we can study the message content and effectiveness for a large number of messages designed to motivate people to change their behavior.

**Transtheoretical Model of behavior change**

There are many theories and models on changing and influencing behavior, ranging from more practical, such as Persuasive Design [7] to more theoretical ones, such as the Social Ecological Model [41]. A theory-based model that can easily be used in practice, is the Transtheoretical Model (TTM) from Prochaska et al. [36, 37, 38]. The model has been thoroughly reviewed (e.g., [1]) and also received criticism, specifically the construct of the stages of change [24]. Despite this, we chose the TTM because as articulated in [32], it: “provides a framework for both the conceptualization and the measurement of behavior change, as well as facilitating promotion strategies that are individualized and easily adapted” [32, p. 7].

The TTM is a dynamic, integrative model focused on the individual and can be applied practically, especially the construct of the stages of change [32], which classifies people into (not necessarily linearly) progressing stages of changing behaviors: Precontemplation (PC), Contemplation (C), Preparation (P), Action (A) and Maintenance (M). While the stages of change (see Table 1) are useful in explaining when changes in cognition, emotion, and behavior take place, the
<table>
<thead>
<tr>
<th>Stage of change</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precontemplation (PC)</td>
<td>The individual is not willing to change in the foreseeable future (measured as the next 6 months). Individuals in this stage are mostly uninformed or demoralized.</td>
</tr>
<tr>
<td>Contemplation (C)</td>
<td>The individual is willing to change in the next 6 months. Individuals in this stage are aware of some pros of behavior change, but are still more inclined to value the cons.</td>
</tr>
<tr>
<td>Preparation (P)</td>
<td>The individual is willing to change in the foreseeable future (measured as the next month) and has already taken some small steps towards change (in the past year). Individuals in this stage usually have some plan on how to tackle this inactiveness.</td>
</tr>
<tr>
<td>Action (A)</td>
<td>The individual has changed, but not longer than 6 months. Individuals in this stage have 'changed', but have not reached the duration which exemplifies real behavior change.</td>
</tr>
<tr>
<td>Maintenance (M)</td>
<td>The individual has changed, longer than 6 months. Individuals in this stage have changed and are working not to relapse.</td>
</tr>
</tbody>
</table>

**Table 1. The stages of change with a short description.**

The individual has changed, longer than 6 months. Individuals in this stage are working not to relapse. The individual is willing to change in the next 6 months. Individuals in this stage are aware of some pros of behavior change, but are still more inclined to value the cons. The individual is not willing to change in the foreseeable future (measured as the next 6 months). Individuals in this stage are mostly uninformed or demoralized. The individual is willing to change in the foreseeable future (measured as the next month) and has already taken some small steps towards change (in the past year). Individuals in this stage usually have some plan on how to tackle this inactiveness.

**Tailoring to stage of change**

Stage-based interventions can be more effective than neutral interventions according to Marcus et al. [26]. In their study, they mailed (at baseline, 1, 3 and 6 months) intervention materials. A tailored intervention (tailored to the participant’s stage of change, associated processes and more) versus a neutral intervention was tested and they found that both increased physical activity levels, but the tailored version increased physical activity levels most. This is a good example of the success of tailoring, but also the success of applying the TTM. This use of theories or models, such as the TTM, in designing interventions to change behavior has been advocated (see [3, 30]) because it will help evaluate the model. However, it is also noted that there is little guidance on how to use theories or models in designing concrete interventions [30]. This could explain why interventions designed with the TTM in mind usually fail to be properly based on the TTM, because not all elements of the model are incorporated [13]. In fact, when tailoring to the stages of change, it is more important than anything else to combine the stages with the processes of change [40], because they are codependent. The importance of combining both is clear from several studies and as Spencer et al. remark in their review of TTM literature “Ensuring that participants use the appropriate processes of change as they move through the stages is essential for their success.” [40, p. 438].

**RESEARCH GOALS AND EXPECTATIONS**

Our long-term goal is to develop technology that motivates and engages people to exercise and adhere to exercising so that long-term behavior change can be achieved. We believe that crowdsourcing peer-to-peer designed motivational text messages capturing the processes of change and then tailoring to the stages of change is an effective method to use theories or models in practice, and can eventually contribute to longer-term exercise adherence.

In the current work, we collected peer-to-peer designed motivational messages and evaluated whether these messages could reflect the processes of change from the TTM. Because previous research [4] showed that peer-to-peer designed messages reflected the same key theoretical concepts as expert-written messages, we expected that:

1) For study 1, when asked to motivate a person in a particular stage of change, participants will design motivational messages that reflect the processes of change related to that stage. To be specific: messages reflecting Experiential processes will be more prevalent in the earlier stages, while messages reflecting Behavioral processes will be more prevalent in the later stages.

And because previous research [27] showed that people evaluated the processes of change differently in the context of exercise than in other contexts, we expect that:

2) For study 2, participants in a certain stage of change will evaluate the motivational messages that reflect the processes of change related to that stage as more motivating, in the context of exercise. To be specific: both the rating of the Experiential and Behavioral processes will peak in the Action stage, but the Experiential processes will be rated as more motivating than the Behavioral processes in the earlier stages and decline in the Maintenance stage, while the rating of the Behavioral processes will not decline in the Maintenance stage.

**STUDY 1: DESIGNING AND CODING MOTIVATIONAL MESSAGES**

In the first study we conducted an online crowdsourced survey where participants designed motivational text messages. Then the collected motivational text messages were coded according to categories based on the TTM’s processes of change.

**Crowdsourcing the design of motivational messages**

The online crowdsourcing survey, to collect motivational messages, was set up on SurveyMonkey through Amazon Me-
The individual substitutes positive behaviors for the individual's problem behavior.

The individual seeks increased knowledge about the causes, consequences and cures for their problem behavior.

The individual has the belief that he can change and commits to it by choosing a course of action.

The impact that the individual's problem behavior has on their environment is reevaluated.

Cognitions and emotions regarding the individual with respect to their problem behavior are reevaluated.

The individual's emotions about the problem behavior and possible solutions are evoked.

The individual seeks trust and open discussion about the problem behavior as well as support for the healthy behavior change.

Attempts are made to increase alternatives for the individual’s former problem behavior.

Steps or changes made by the individual are rewarded when in a positive direction or punished when in a negative direction.

Stimuli that may cue a lapse back to the problem behavior are avoided and prompts for more healthier alternatives are inserted.

<table>
<thead>
<tr>
<th>Experiential processes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consciousness raising (CR)</td>
<td>The individual seeks increased knowledge about the causes, consequences and cures for their problem behavior.</td>
</tr>
<tr>
<td>Dramatic relief (DR)</td>
<td>The individual’s emotions about the problem behavior and possible solutions are evoked.</td>
</tr>
<tr>
<td>Environmental reeval. (ER)</td>
<td>The impact that the individual’s problem behavior has on their environment is reevaluated.</td>
</tr>
<tr>
<td>Social liberation (SOL)</td>
<td>Attempts are made to increase alternatives for the individual’s former problem behavior.</td>
</tr>
<tr>
<td>Self-reevaluation (SR)</td>
<td>Cognitions and emotions regarding the individual with respect to their problem behavior are reevaluated.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioral processes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-liberation (SEL)</td>
<td>The individual has the belief that he can change and commits to it by choosing a course of action.</td>
</tr>
<tr>
<td>Helping relationships (HR)</td>
<td>The individual seeks trust and open discussion about the problem behavior as well as support for the healthy behavior change.</td>
</tr>
<tr>
<td>Counterconditioning (CC)</td>
<td>The individual substitutes positive behaviors for the individual’s problem behavior.</td>
</tr>
<tr>
<td>Reinforcement manag. (RM)</td>
<td>Steps or changes made by the individual are rewarded when in a positive direction or punished when in a negative direction.</td>
</tr>
<tr>
<td>Stimulus control (SC)</td>
<td>Stimuli that may cue a lapse back to the problem behavior are avoided and prompts for more healthier alternatives are inserted.</td>
</tr>
</tbody>
</table>

| Table 2. The processes of change divided in Experiential and Behavioral processes with a short description. |

Sample
The sample consisted of 500 people. Data from 19 respondents was excluded because their questionnaires were incomplete. The final sample we worked with included 481 respondents (250 male and 231 female). The study was conducted in English. All but 5 respondents were native English speakers, their data was not anomalous and was kept.

The minimum age was 18 and the maximum was 68. The average age was 31.09 (SD = 9.22) and the median 29. With respect to education, 201 respondents received some college education, 183 had a college degree, 46 completed a master degree, 42 completed high school, 5 obtained a PhD and 4 had other types of qualifications. The AMT requirements for the respondents were that they should have already completed >1000 tasks on AMT, that >98% of these tasks were approved successfully and that the respondents were located in the United States. These requirements ensured that respondents were already familiar with surveys (our survey was quite extensive), that they were serious about filling in the survey (only 19 were not, which is low for online anonymous surveys) and that they had some proficiency in English. In fact, 473 reported ‘very good’ as self-assessed proficiency, 7 ‘good’ and 1 ‘average’, but none ‘bad’ or ‘very bad’.

Although our sample consisted only of AMT workers, which could misrepresent the general population, AMT can be considered to deliver an acceptable representation of general society, especially by online survey standards [29]. However, we are aware of the limitations of AMT.

Task manipulation
To obtain motivational messages from the participants, we developed five scenarios, each about a different stage of change, for which participants designed messages. The participant was given one short paragraph (the scenario) in which a person in a specific stage of change (for instance, Precontemplation) was described. The scenario that the participant received was randomized. We asked the participants to imagine that they had to motivate this person to exercise. In total, participants were asked to come up with six messages. Examples of scenarios and actual participant responses are shown in Table 3.

Procedure
Participants were recruited through AMT. They were informed of their compensation, the goal of the survey and the estimated time to complete the survey. Participants could then decide to accept or decline the survey and proceed to the SurveyMonkey website where the survey was hosted. There, the goal of the survey was summarized and participants were asked to complete a consent form. The participants were then presented with one of the five scenarios, and were asked to come up with motivational messages for the scenario. Afterwards, participants were debriefed about the goals of the survey and given a completion code to fill in on AMT to receive payment. On average, the survey took about 45 minutes to complete. Participants were compensated with 3 US dollars for their participation.

Coding motivational messages
To see how the content of the 2886 (481 x 6) crowdsourced peer-to-peer messages reflected the processes of change, we translated the processes of change from the TTM to coding categories. We used a similar process to [25]. To use these categories, we operationalized the fairly fluid processes of change descriptions into more fixed definitions that considered the perspective of ‘text messages sent by a peer’. The procedure started with two coders (coder 1 was the main investigator) separately coming up with operationalized definitions of the processes of change for a first version of the codebook.

The codebook was developed following the guidelines of Guest and MacQueen [8, 25] who advise structuring the codebook with (at least) six parameters: the code, brief definition, full definition, when to use, when not to use, and examples. We added a seventh parameter, namely the ‘perspective’, which gives an alternative definition of what the process could look like in terms of a text message. Each category (a process of change) for coding was described in the same manner. Because the majority of the messages were only a few words
long, it was decided that only one code per message would be used. The context for the definitions of the processes of change was the TTM in general, but also the TTM specifically in the context of exercising. A unified codebook was defined before coding started.

The messages were coded iteratively, independently and without scenario information by the two coders and afterwards the coders resolved any mismatches. The codebook would then be updated to reflect these resolved mismatches and the next round of coding started. The first round of coding started with 60 messages and increased to a maximum of 300 for the final round of coding (approximately 10% of the data). The final round of coding was determined when the coders felt they had reached a saturation point at which achieving a higher agreement would not be feasible. The final agreement was a Cohen’s kappa of 0.72, which is substantial [21].

For our consecutive study (Study 2), we needed only the most representative messages for each of the process categories. To this end, we did another selection of only the best representations of the processes. A follow-up coding was designed to complement the last round of coding, in which we used a ‘certainty’ measure along with the existing coding. This meant that both coders would add a certainty code to their ‘normal’ coding of the messages if they were 99% sure about the message belonging to the particular coding category (i.e., one of the ten processes) and that the other coder would agree on this. The messages that were coded by coder 1 scored a Cohen’s kappa of 0.86 for the certainty measure.

The remaining messages (2586 messages) were coded by one coder (coder 1), yielding a dataset containing 2886 coded messages in total and a subset of 1433 (49.7%) messages coded with the certainty measure (examples of coded messages are given in Table 3 and a snippet of the codebook in Table 4).

Results: designed and coded motivational messages reflecting the processes

In this section, we present the results of the online survey and how the crowdsourced messages fitted into categories based on the processes of change. We looked at what kinds of messages (process categories) and how many of each type of message would come up for each of the stage of change scenarios, and whether the distribution of messages more or less aligned to our expectation: the messages reflecting Experiential processes were expected to be more prevalent (have more counts) in the earlier stages compared to the later stages, while the messages reflecting Behavioral processes were expected to be more prevalent (have more counts) in the later stages compared to the earlier stages. We also looked whether the same distribution of counts for the messages was shown for each of the separate process categories.

The motivational messages from 481 participants were used in the coding process. It is important to note that the manipulated stages of change scenarios were not completely equally distributed: 91 participants designed motivational messages for the Precontemplation scenario (PC-S), 93 for the Contemplation scenario (C-S), 87 for the Preparation scenario (P-S), 102 for the Action scenario (A-S), and 108 participants for the Maintenance scenario (M-S).

To see whether the distribution of coded messages follows our expectation, we first looked into the higher-order Experiential and Behavioral processes (see Table 2 for an explanation). The results are shown in Table 5. Of the total of 2886 messages, 800 (27.7%) are coded as Experiential processes, and 2086 as Behavioral (72.3%) processes. From Table 5 it can be seen that the distribution of the higher-order processes over the stages resembles our expectation: Experiential processes are more prevalent in the earlier stages and Behavioral processes are prevalent in the later stages (see the “count” row of Table 5). A Chi-square test was carried out to see if the condition (i.e., the stages) had an effect on the counts within the higher-order processes. The results show that there is a significant association between the stages and higher-order processes ($\chi^2(4) = 223.179, p < .001$). The values of the standardized residuals are used to further interpret the results of the Chi-square test. The standardized residuals represent the error between the observed frequency (i.e., what the data actually shows) and expected frequency (i.e., what the model predicts). A positive value indicates an overrepresentation and a negative value points to an underrepresentation. A $z$-value higher than 1.96 or lower than -1.96 for either the over or underrepresentation is considered to be significant at $p < 0.05$ [6].

The residuals show that for the Experiential processes in the Precontemplation stage, there is a significant overrepresentation of the processes ($z = 8.9$), and in the Action and Maintenance stage there is a significant underrepresentation of the processes ($z = -6.0$ and $z = -6.5$). For the Behavioral processes this is reversed, in the Precontemplation stage, there is a significant underrepresentation of the processes ($z = -5.5$), and in the Action and Maintenance stage there is a significant overrepresentation of the processes ($z = 3.7$ and $z = 4.0$). Overall, this means that there are more Experiential messages...
in the earlier stages and fewer in the later stages than the Chi-square model predicts. Also, this means there are fewer Behavioral messages in the earlier stages and more in the later stages than the Chi-square model predicts.

At a lower level, we also looked at the distribution for the ten separate processes. In Table 6 it can be seen that the distribution of the processes over the stages somewhat resembles our expectation for some, but not all, of the processes. A Chi-square test was again performed to see if the condition (i.e., the stages) had an effect on the counts of the processes. The results show that there is a significant association between the stages and processes ($\chi^2(36) = 437.851, p < .001$). The values of the standardized residuals are used to further interpret the results of the Chi-square test.

For the Experiential processes, Conscientiousness Raising, Dramatic Relief, Environmental Reevaluation and Self-reevaluation show a significant trend of more counts in the earlier stages and fewer in the later stages. The Experiential process of Social Liberation has too few counts to interpret any results.

For the Behavioral processes, only Reinforcement Management shows a significant trend of fewer counts in the earlier stages and more in the later. Self-liberation and Counterconditioning both show some inclination toward this trend near the Action stage, but decline for the Maintenance stage. Helping Relationships is more or less equal throughout the stages and Stimulus Control displays the opposite (Experiential) trend significantly.

Overall, the distribution over the stages for both higher-order processes is in line with our first expectation, but when looking at the distribution of the separate processes, the same trend is not present for all ten processes.

**Discussion**

We found that the messages that people design for different stage of change scenarios can be reliably coded into categories of processes of change. We used scenarios based on the stages to collect a broad range of unbiased messages. The scenarios also included general context (middle-aged, steady personal life, solid friend foundation) to make it realistic, which could have potentially biased the participants, but this text was kept the same for all scenarios and concerned less text than the actual stage description (see Table 3). Also, the coding was carried out without the stage information of the messages. Although a Chi-square test is not optimal to test the expected distribution of messages, together with the message counts, this test does show that the scenarios have an influence on the message content. There is no ratio for distribution over the different stages yet, so we could not do a rate comparison. We hope this research proved to be a first step in dealing with a ratio for distribution of messages.

The kinds of messages (process categories) are different between the stages in the same way as expected: Experiential processes are more prevalent in the earlier stages while Behavioral processes are more prevalent in the later stages. Although the distribution within the higher order processes aligned quite well with our expected distribution, we did not have an expectation for the distribution of the number of messages between the higher order processes. It is important to note that the distribution between the higher-order processes themselves was skewed (2086 of 2886 were coded as Behavioral processes). We found that people came up with mostly Behavioral-themed-processes (72, 3%). This could mean that, in our study setup, people found it much easier to think of more ‘action-oriented’ Behavioral messages than more ‘thinking-oriented’ Experiential messages, or that in general people find it easier to think of more ‘action-oriented’ messages.
For the separate processes of change distributions over the stages, quite a few processes aligned to our expectation, although also several processes did not. Interestingly, most of the Experiential processes followed the expected distribution, but not many of the Behavioral processes. The counts for the Behavioral processes (except Reinforcement Management) are reasonably stable across all stages. One interpretation could be that when someone is motivated to learn a behavior it is more natural to start earlier with Behavioral as well as Experiential messages. Also interesting is that the significant results are mostly for the Precontemplation and Maintenance stages: an interpretation could be that this is where the behaviors are stable, and the differences in the processes that are used largest.

Overall, crowd-designed TTM-informed motivational messages seems to be a promising method. The messages reflected the processes of change. However, we did not yet know whether people in a specific stage of change would indeed find the TTM-tailored messages motivating as we expected. This was addressed in Study 2.

**STUDY 2: EVALUATING MOTIVATIONAL MESSAGES**

The second study was designed to validate the first study and to see whether the coded motivational messages would be rated according to our expectations. This study was also carried out on SurveyMonkey through AMT. In the survey, people evaluated a selection of the best representations of the previously coded messages. We asked people to rate how motivating they found the messages to be. We also measured the participants’ self-assessed stage of change in relation to exercising.

**Sample**

The sample consisted of 350 people. No data from any respondent was excluded. The study was conducted in English. All but 5 respondents were native English speakers, their results were not anomalous and were left in the sample.

The minimum age was 20 and the maximum was 71. The average age was 36.53 (SD = 11.83) and the median 34. With respect to education, 106 respondents received some college education, 142 had a college degree, 35 completed a masters, 59 completed high school, 5 obtained a PhD and 3 had other types of qualifications. To ensure consistency and a high quality of responses, the AMT requirements for the respondents were the same as those in the first survey.

**Questionnaire measures**

To measure participants’ stage of change, we used the validated 1-item stage of change measure for exercise where participants were given a description of regular exercise and of the five stages and rated their stage based on that description.

**Task manipulation**

We presented participants with a selection of fifty messages and asked them to rate each according to how motivating they thought the message was (“Please rate how motivating or demotivating you specifically find these messages for yourself.”). From the subset of messages that were coded with the certainty of responses, the AMT requirements for the respondents were the same as those in the first survey.

We presented participants with a selection of fifty messages and asked them to rate each according to how motivating they thought the message was (“Please rate how motivating or demotivating you specifically find these messages for yourself.”). From the subset of messages that were coded with the certainty of responses, the AMT requirements for the respondents were the same as those in the first survey.

**Procedure**

Participants were recruited through AMT. They were informed of their compensation, the goal of the survey and the estimated time to complete the survey. Participants could then decide to

---

Table 6. The distribution of all the codes over the ten process categories and five stages of change scenarios with their respective counts for the processes and stages. 

<table>
<thead>
<tr>
<th>Categories/Scenarios</th>
<th>PC-S</th>
<th>C-S</th>
<th>P-S</th>
<th>A-S</th>
<th>M-S</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>67</td>
<td>18</td>
<td>36</td>
<td>10</td>
<td>7</td>
<td>138</td>
</tr>
<tr>
<td>Expected</td>
<td>31.0</td>
<td>26.7</td>
<td>25.0</td>
<td>29.3</td>
<td>26.1</td>
<td>138</td>
</tr>
<tr>
<td>Std. residual</td>
<td>6.5³</td>
<td>-1.7</td>
<td>2.2¹</td>
<td>-3.6¹</td>
<td>-3.7³</td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>33</td>
<td>8</td>
<td>13</td>
<td>4</td>
<td>1</td>
<td>59</td>
</tr>
<tr>
<td>Expected</td>
<td>13.2</td>
<td>11.4</td>
<td>10.7</td>
<td>12.5</td>
<td>11.2</td>
<td>59</td>
</tr>
<tr>
<td>Std. residual</td>
<td>5.4³</td>
<td>-1.0</td>
<td>0.7</td>
<td>-2.4¹</td>
<td>-3.0³</td>
<td></td>
</tr>
<tr>
<td>ER</td>
<td>33</td>
<td>13</td>
<td>19</td>
<td>14</td>
<td>10</td>
<td>79</td>
</tr>
<tr>
<td>Expected</td>
<td>17.7</td>
<td>15.3</td>
<td>14.3</td>
<td>16.8</td>
<td>14.9</td>
<td>79</td>
</tr>
<tr>
<td>Std. residual</td>
<td>3.6³</td>
<td>-0.6</td>
<td>-1.4</td>
<td>-0.7</td>
<td>-1.3</td>
<td></td>
</tr>
<tr>
<td>SOL</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Expected</td>
<td>2.5</td>
<td>2.1</td>
<td>2.0</td>
<td>2.3</td>
<td>2.1</td>
<td>11</td>
</tr>
<tr>
<td>Std. residual</td>
<td>-0.3</td>
<td>2.7²</td>
<td>-0.7</td>
<td>-1.5</td>
<td>-0.1</td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>164</td>
<td>132</td>
<td>101</td>
<td>64</td>
<td>52</td>
<td>513</td>
</tr>
<tr>
<td>Expected</td>
<td>115.2</td>
<td>99.2</td>
<td>92.8</td>
<td>108.8</td>
<td>97.1</td>
<td>513</td>
</tr>
<tr>
<td>Std. residual</td>
<td>4.5³</td>
<td>3.3³</td>
<td>0.9</td>
<td>-4.3³</td>
<td>-4.6³</td>
<td></td>
</tr>
<tr>
<td>SEL</td>
<td>178</td>
<td>168</td>
<td>190</td>
<td>221</td>
<td>182</td>
<td>939</td>
</tr>
<tr>
<td>Expected</td>
<td>210.8</td>
<td>181.6</td>
<td>169.8</td>
<td>199.1</td>
<td>177.6</td>
<td>939</td>
</tr>
<tr>
<td>Std. residual</td>
<td>-2.3¹</td>
<td>-1.0</td>
<td>1.5</td>
<td>1.6</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>44</td>
<td>53</td>
<td>31</td>
<td>59</td>
<td>51</td>
<td>238</td>
</tr>
<tr>
<td>Expected</td>
<td>53.4</td>
<td>46.0</td>
<td>43.0</td>
<td>50.5</td>
<td>45.0</td>
<td>238</td>
</tr>
<tr>
<td>Std. residual</td>
<td>-1.3</td>
<td>1.0</td>
<td>-1.8</td>
<td>1.2</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>49</td>
<td>65</td>
<td>49</td>
<td>83</td>
<td>43</td>
<td>289</td>
</tr>
<tr>
<td>Expected</td>
<td>64.9</td>
<td>55.9</td>
<td>52.3</td>
<td>61.3</td>
<td>54.7</td>
<td>289</td>
</tr>
<tr>
<td>Std. residual</td>
<td>-2.0¹</td>
<td>1.2</td>
<td>-0.5</td>
<td>2.8²</td>
<td>-1.6</td>
<td></td>
</tr>
<tr>
<td>RM</td>
<td>40</td>
<td>67</td>
<td>72</td>
<td>146</td>
<td>187</td>
<td>512</td>
</tr>
<tr>
<td>Expected</td>
<td>115.0</td>
<td>99.0</td>
<td>92.6</td>
<td>108.6</td>
<td>96.9</td>
<td>512</td>
</tr>
<tr>
<td>Std. residual</td>
<td>-7.0³</td>
<td>-3.2²</td>
<td>-2.1¹</td>
<td>3.6³</td>
<td>9.2³</td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>38</td>
<td>28</td>
<td>20</td>
<td>11</td>
<td>11</td>
<td>108</td>
</tr>
<tr>
<td>Expected</td>
<td>24.2</td>
<td>20.9</td>
<td>19.5</td>
<td>22.9</td>
<td>20.4</td>
<td>108</td>
</tr>
<tr>
<td>Std. residual</td>
<td>2.8²</td>
<td>1.6</td>
<td>0.1</td>
<td>-2.5¹</td>
<td>-2.1¹</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>648</td>
<td>558</td>
<td>522</td>
<td>612</td>
<td>546</td>
<td>2886</td>
</tr>
</tbody>
</table>

For the separate processes of change distributions over the stages, quite a few processes aligned to our expectation, although also several processes did not. Interestingly, most of the Experiential processes followed the expected distribution, but not many of the Behavioral processes. The counts for the Behavioral processes (except Reinforcement Management) are reasonably stable across all stages. One interpretation could be that when someone is motivated to learn a behavior it is more natural to start earlier with Behavioral as well as Experiential messages. Also interesting is that the significant results are mostly for the Precontemplation and Maintenance stages: an interpretation could be that this is where the behaviors are stable, and the differences in the processes that are used largest.

---

accept or decline the survey and proceed to the SurveyMonkey website were the survey was hosted. The goal of this survey was summarized and participants were asked to complete a consent form. First, the participants were given fifty motivational messages to rate and then participants were asked to fill out the 1-item stage of change measure. Afterwards, participants were debriefed about the detailed goals of the survey and given a completion code to fill in on AMT to receive payment. On average, the survey took about 40 minutes to complete. Participants were compensated with 3 US dollars for their participation.

Results: coded motivational messages evaluated
In this section, we present the results of our evaluation survey and how the evaluated messages fitted into the categories of processes of change. The evaluations of 350 participants were analyzed. We looked at whether the selection of our coded messages were evaluated as representative of our developed process categories. We also looked at whether the motivational messages were evaluated in the same way that the processes relate to the stages of change (in the context of exercise): the Experiential processes were expected to be rated as more motivating in the earlier stages of change, peak in the Action stage and decline in the Maintenance stage, while the ratings of the Behavioral processes were expected to also peak in the Action stage, but not decline in the Maintenance stage.

It is important to note that for this study, we measured the self-assessed stages of change, which were not equally distributed: 120 participants rated themselves to be in the Maintenance stage (M), 94 in the Preparation stage (P), 73 in the Contemplation stage (C), 46 in the Action stage (A), and 17 participants rated themselves in the Precontemplation stage (PC).

The coded messages we selected for each process (five for each process) were representative of our developed categories, as is shown in Table 7 through the reliability of the coded message categories. The reliability of the measures was overall very good. The only disputable measure was that of Counterconditioning, with a Cronbach’s alpha of .68 which we still found acceptable (and also comparable to other relevant work [27] where they had this score for Social Liberation). Otherwise, the reliability scores were between .72 and .88.

The coded messages somewhat followed the expectation of the processes aligning to the stages in a similar way to Marcus et al. [27], as is shown by the ratings given to the process categories for each stage of change (see Figure 1). We found that there are peaks for all Experiential and Behavioral processes in the Action stage and that the Behavioral processes are rated as more motivating in the later stages. But, the Experiential processes are not rated as more motivating in the earlier stages (Dramatic Relief and Environmental Reevaluation are even rated as demotivating in the earlier stages) and although there is a decline for Experiential processes in the Maintenance stage, the same decline is also found for the Behavioral processes.

Because these results were unexpected, we investigated further and analyzed whether there was a significant difference in the rating of the process categories between the different stages of change. We carried out separate univariate analyses of variance (ANOVA) with the stages as predictor variable and the process categories as separate outcome variables (i.e., CR, DR, ER, SOL, SR, SEL, HR, CC, RM and SC). As can be seen in Figure 1, all process categories differ significantly along the stages of change (all nine \( p < .05 \)) except for the Helping Relationships category. This indicates that for these process categories, the scores differ significantly when compared between the stages of change. In Figure 1 we can observe the direction of the relations in the reported means for all the stages. To see between which stages these processes of change differ significantly, we performed post hoc tests. The post hoc test results for the significant process categories (all but the Helping Relationships category) are shown in Table 8. As can be seen, it is mostly the Precontemplation stage that differs from the Action and Maintenance stage and the Contemplation stage that differs from the Action stage.

Overall, when looking at the average ratings over the stages it is clear that: the peak for all processes is in the Action stage, which is what we expected. There are no higher ratings for the Experiential processes than for the Behavioral processes in the earlier stages; in fact, two Experiential processes (Dramatic Relief and Environmental Reevaluation) have negative ratings in the Precontemplation stage, and the decline in the Maintenance stage is there for Experiential but also for Behavioral processes, which is not what we had expected.

Discussion
The selection of coded messages to represent the process categories showed very good reliability. Therefore, the five messages we selected for each category were a good fit for the process of change they represented. This shows that using crowdsourcing to design motivational messages for behavior change is feasible.
We also found that the motivational rating of coded messages of the participants did not entirely match our expectation of processes over stages. Specifically, that the messages representing the Experiential processes were not rated as more motivating in the earlier stages of change, although they did peak in the Action stage. And although the messages representing the Behavioral processes were rated as more motivating in the later stages of change and peaked in the Action stage, they did decline in the Maintenance stage. To investigate further, we looked at where the significant differences between the stages were for all the processes.

Even though the Experiential processes were not rated more motivating than the Behavioral processes in the earlier stages of change and the Behavioral processes did decline in the Maintenance stage, the broader trend for Experiential and Behavioral processes was found (both peak in the Action stage). This general trend is also shown in the post hoc test results (see Table 8), where four of the five Experiential processes (not Environmental Reevaluation) show significant differences between the Action stage and the Precontemplation stage, and between the Action stage and the Contemplation stage. Moreover, the post hoc test results on Experiential processes showed twelve significant mean differences and eight of those were in relation to the Action stage. The Behavioral peak was also found in the Action stage. The post hoc test results for the four significant Behavioral processes (not Helping Relationships), showed differences between the Action stage and the Precontemplation and Contemplation stage, but also between the Maintenance stage and the Precontemplation and Contemplation stage. The latter could be an indication that, even though there is a decline in the Maintenance stage for the Behavioral processes, overall, the Behavioral processes are still relevant in the Maintenance stage.

It is important to note that for Study 2, it was possible for people to rate messages on the demotivating end of the spectrum. We did this on purpose, to see if there might be messages we definitely should avoid for certain stages. The results for Dramatic Relief, Environmental Reevaluation and Stimulus Control actually show that they were rated as demotivating in

![Figure 1. Bar charts describing the averages (M), standard deviations (SD), F-statistics and p-values for all coded message categories. On the left, all Experiential process categories on the right, all Behavioral process categories. Messages were rated on a scale from 1 (“Very demotivating”) to 5 (“Very motivating”) with a 3 as neutral (“Neither demotivating nor motivating”). N = 350, PC = Precontemplation (N = 17), C = Contemplation (N = 73), P = Preparation (N = 94), A = Action (N = 46) and M = Maintenance (N = 120).](image-url)
Although there is a general consensus on the value of most behavior change theory (the TTM) in behavior change, there is still much room to increase the effectiveness and the application of such theories by designing practical implementations of whole theories and testing these in various contexts, such as the exercise domain.

As part of a larger study, we sought to leverage HCI practices, to come to useful and practical insights about how to translate theoretical constructs (i.e., the processes of change) to text messages. We showed a method by which theoretical constructs of a behavior change theory or model (the TTM) can be represented by crowd-designed motivational text messages. Although this result is context-dependent, it could prove to be a valuable method for other theories, models, or contexts. The motivational text messages will be used in motivational technology. For example, in a smartphone application, where users can receive these messages to get motivated to exercise or to be reminded to exercise regularly. The ratings of the motivational text messages (representing the processes of change) can inform the application which messages would be most effective, so that they can be tailored to the specific user’s stage of change. Overall, the findings in this paper can help inform researchers designing motivational technology for long-term behavior change who: 1) look for a method to translate theoretical insights to practical text messages; and who 2) want to go beyond a one-size-fits-all strategy and design effective motivational technology that tailors to the stage of change a user is in, but also to more individual factors.

ACKNOWLEDGMENTS
This research was funded by COMMIT/ and is part of the P3 project SenseL: Sensor-Based Engagement for Improved Health. We would like to thank Cristina Zaga and Maartje de Graaf for their input.

REFERENCES
5. Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. ACM


http://dx.doi.org/10.1186/1479-5868-7-36


34. GJ Norman, SV Benisovich, CR Nigg, and JS Rossi. 1998. Examining three exercise staging algorithms in two samples. In *19th annual meeting of the Society of Behavioral Medicine*.


