

Send Me a Different Message: Utilizing Cognitive Space to Create Engaging Message Triggers

Rafal Kocielnik, Gary Hsieh

Human Centered Design and Engineering

University of Washington

{rkoc, garyhs}@uw.edu

ABSTRACT

Social systems and applications often rely on message triggers to promote, remind and even persuade people to perform certain actions. However, repeated exposure to these triggers can lead to boredom, annoyance and decreased engagement. While existing research suggests that diversification of trigger contents may mitigate these issues, no systematic way of introducing it has been proposed. This paper proposes two message diversification strategies based on the use of cognitive spaces: 1) target-diverse – using concepts cognitively close to the targeted action; and 2) self-diverse – using concepts cognitively close to the message’s recipient. Through a controlled experiment we found that the self-diverse strategy reduces annoyance and boredom from repeated exposure and that both strategies increase perceived informativeness and helpfulness of the triggers. In a subsequent 2-week long field deployment focused on assessing the effects of the self-diverse strategy, we found that this strategy results in higher activity completion through supporting awareness, providing more information, and making the triggers more personally relevant. These diverse triggers are perceived as motivators rather than simple reminders. We conclude with insights on how to design and generate diverse messages.

Author Keywords

triggers; messages; behavior change; persuasion.

ACM Classification Keywords

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

INTRODUCTION

Increasingly, technologies-mediated motivators are being used to influence positive behaviors from users, such as nudging users to exercise [14,51], conserve energy [1], or to be civically engaged [42]. These types of nudges can be especially critical in the CSCW and social computing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. *CSCW '17*, February 25-March 01, 2017, Portland, OR, USA
© 2017 ACM. ISBN 978-1-4503-4335-0/17/03...\$15.00
DOI: <http://dx.doi.org/10.1145/2998181.2998324>

context, where users’ participation and continued contributions are critical [47,58]. These systems need to attract users to generate contents (e.g. movie reviews [12]), contribute opinions (e.g. Yelp ratings [43]) and even provide financial support (e.g. Wikipedia donations [3]).

Message-based triggers have been one of the most commonly used approaches for encouraging such participation and sustained involvement [40]. Triggers [22], or cues to action [24] help promote, remind, and even persuade individuals to perform the desired actions. A growing number of behavior change applications have also begun to adopt and apply triggers into their designs. This includes an alarm reminding you to take a break from sitting too long [46], a short message reminding you to eat more fruits and vegetables [55], or a text based intervention for triggering cancer screening visit [8].

However, designing effective triggers is hard [14,22]. One critical problem is that repeated exposure to these triggers is often needed [14]. This is because many of the desired actions from users require continued support and reminders over a period of time to sustain behaviors and participation [22]. However, much prior work has found that such repetitions can lead to annoyance and boredom [10,54], result in purposeful avoidance [28], contents blindness [31] and, in extreme cases, even lower motivation [48] and cause reactance [15].

One proposed solution to help reduce some of the negative effects of overexposure, based on research in advertising, is to increase the diversity of messages [45]. Instead of getting the same messages, e.g., in a social media health campaign, “exercising is good for you” repeatedly, different messages may be used over time. But, while this concept works in theory, applying it to practice is not simple. How might designers diversify the messages in a systematic manner? How can they actually generate these diverse messages? Some recent work have shown the feasibility and added value of using crowds for generating motivational contents [13,57]. We therefore want to explore how we can leverage computational and crowd-based approaches to accomplish message generation in an effective and low-cost manner.

In this work, we aim to explore solutions to address these challenges. Using the concept of cognitive spaces from the domain of persuasion [20], we proposed two strategies that

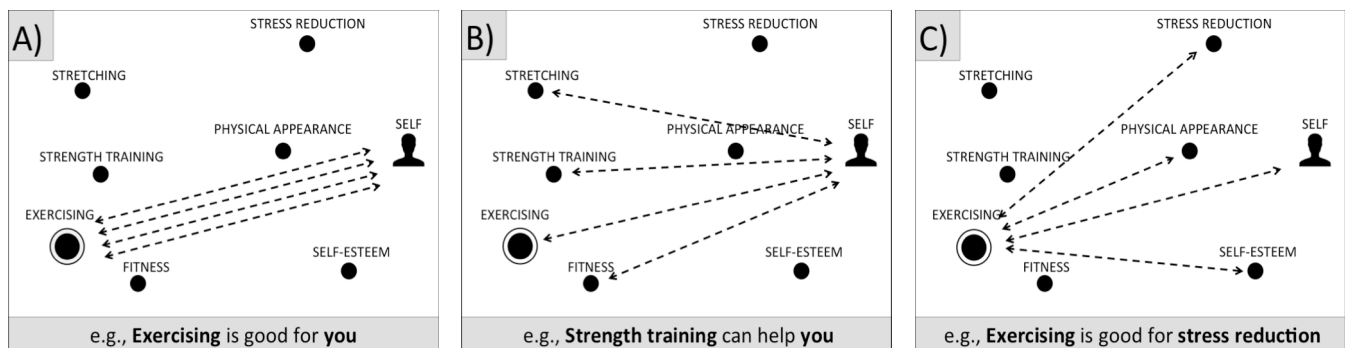


Figure 1 - A non-diverse (baseline) and two diverse strategies as depicted in the cognitive space: A) Non-diverse – messages connect self and exercising, B) Target-diverse – messages connect concepts cognitively close to target (e.g. different types of exercising) and self, C) Self-diverse – messages connect concepts cognitively close to self (e.g. motivations) with exercising.

can be used to systematically structure the generation of diverse messages. One is the target-diverse strategy, in which diverse messages can be generated using alternative concepts cognitively close to the targeted behavior (e.g., for exercising: strength training, aerobics, fitness). Another is the self-diverse strategy, in which diverse messages can be generated using alternative concepts cognitively close to the message's recipient (e.g., benefits people care about or values they hold, such as relaxation, stress reduction, physical appearance and benevolence) (Figure 1).

To test our strategies, we conducted two studies. In a controlled online experiment with 122 participants, we found that both strategies, when compared to a baseline, non-diverse message condition, can improve the perceived helpfulness and informativeness of the message triggers, and that the self-diverse strategy helps minimize annoyance and boredom, mitigating the negative effects of repeated exposure. To further assess the effects of our self-diverse strategy, we conducted a 2-week long field experiment with 28 participants, where participants received 4 message triggers daily asking them to perform one of 4 physical activities. We found that compared to the baseline non-diverse condition, the self-diverse messages resulted in higher exercise completion. Post-study interviews further uncover the strengths of the diverse messages (e.g., they increase attention, provide additional information and enable personally relevant elaboration). The interviews also suggest challenges and future opportunities for generating diverse messages using our proposed strategies.

Our work contributes to the research and design of message triggers in three ways: 1) we show the effectiveness of using two cognitive space modeling based strategies for generating diverse messages that mitigate the negative effects of repeated exposure, 2) we demonstrate a number of processes, using computational and crowd-based approaches, to generate message triggers, 3) we offer guidelines and highlight opportunities, based on our experiences and qualitative feedback, to design more effective message triggers.

REPEATED EXPOSURE AND DIVERSIFICATION

Research on the effects of repeated exposure to persuasive contents suggests an existence of an inverted-U relationship between exposure and liking [5]. Findings suggest this relationship exists due to two opposing psychological processes: positive habituation (a reduction in uncertainty or conflict) and tedium. The relative strength of each varies as a function of exposure to stimulus. With limited exposure, positive habituation has greater impact than tedium, resulting in liking and minimizing negative responses to the novel stimulus. With repeated exposure, tedium increases leading to negative perceptions and reactions to the stimulus. Simple and homogeneous stimulus (e.g., short, similar messages) as well as high exposure, accelerate the appearance of tedium [27].

Tedium has generally been assessed using annoyance and boredom as measures [10]. But more recent research on short text messages to promote general health [15], smoking cessation [41] and daily physical activity [51], have also indicated the importance of messages being informative and helpful in order to counter tedium. Therefore, in this work to evaluate the quality of our messages we focus on measuring informativeness and helpfulness in addition to annoyance, and boredom.

Mobile message-based triggers in personal informatics

A recent overview of mobile technologies for health & wellness applications [14], emphasizes the value of notifications to sustain behavior and retain users. It also points to the “profound difficulty” of properly designing notifications to make them work for long-term and avoid “alert fatigue”. Another review [21], focuses specifically on the use of SMS mobile messaging for delivering behavior change interventions. Authors report high effectiveness of SMS interventions, but point out the lack of systematic investigations of specific SMS design characteristics and their impact on aspects of participants’ engagement and retention, which authors consider to be a critical factor for successful behavior change. Finally, another recent review [29] focuses on the effectiveness of text messaging for health interventions in general. This review, while

confirming the exposure frequency as important for effectiveness of behavior change interventions, emphasizes the need for researchers to focus more on theories that inform message effects and investigate the relationship between message characteristics and repeated exposure.

We believe that our work on systematic design of diverse messages for mitigating the negative effects of repeated exposure addresses this important gap in research and design knowledge.

Message diversification

To counteract the problems that arise from repeated exposure, one solution that has been proposed and studied is the use of diverse messages. In controlled experiments, diversification was shown to reduce tedium from repeated exposure [27,48], however, the messages used were developed manually by researchers, on a study-by-study basis. There was no process for generating diverse messages in a structured and well-defined way [16,35].

This limitation creates a major barrier for practitioners hoping to generate diverse messages. How might they generate message triggers that are focused around the same topic, but at the same time, diverse?

Research on cognitive structures and persuasion suggests a potential solution. Cognitive structures suggest that our attitudes and beliefs are stored in memory and are interconnected [56]. Activating one element of such structure may also retrieve associated attitudes and beliefs [56]. The distances between concepts define their relatedness [20]. This model has been used for understanding the persuasion process and the design of persuasive messages [6,35]. If the self-referent term (e.g., me) is close to a behavior concept (e.g., exercising), then an intention to exercise is likely. Messages connecting two or more cognitive concepts can act like forces that bring these concepts closer together. For example, to get people to exercise more, we could use a message such as “**exercising** is good for **you**.” or “**exercising** can help **you**,” bringing the concept of exercising closer to the “self.” This is often the most basic form of messages used as current behavior change triggers (e.g., X is good for you) [46,59]. We also used this type of messages in our baseline, non-diverse condition (Figure 1. A).

Because all concepts in the space are interconnected, the pulls and pushes between any two individual concepts would propagate to other concepts in the space, much like the gravity of planets (one of the prevalent models of this cognitive space is aptly named Galileo Theory [20]). This property of cognitive space suggests opportunities for design of multiple diverse persuasive messages. Instead of repeating the same messages using the same concepts (e.g., *exercising* and *you*), we may construct diverse messages using related concepts that are close to the target (exercise) or self (you). Specifically, we propose two such strategies: self-diverse and target-diverse.

Target-diverse strategy

Our proposed target-diverse strategy relies on the use of concepts close to the target concept (e.g. “exercising”) in cognitive space. Based on the properties of cognitive space, by connecting concepts related to exercise (e.g., strength training) to “self” in a message, we can indirectly affect attitude towards “exercising” (Figure 1. B).

Self-diverse strategy

Our second proposed strategy, self-diverse, relies on using different concepts close to “self” for generating the diverse messages. By connecting the concept of “exercising” with concepts that are close to self (those that people care about, e.g., stress reduction) in a message, we can again indirectly pull the target concept to self (Figure 1. C). Thus our first research question is:

RQ1. Can we use cognitive structures to generate diverse message triggers that reduce the negative effects of repeated exposure?

When applying these strategies in practice, we would also need processes and guidelines in generating the actual messages. Researchers and practitioners may be able to use our strategies to help brainstorm related concepts to use in their messages, but the cognitive spaces can be infinitely large (with concepts directly and indirectly related to the targeted behavior and self). With the rise of tools to assess semantic relatedness between concepts [18], as well as the use of crowds [13,57], can we use computing to support the generation of these messages? In addition, with the potential of work done by crowd of people, what is an effective workflow to utilize the crowd to generate diverse message triggers? Thus, another critical part of our research explores potential processes to facilitate the generation of the actual messages.

RQ2. How should we generate target and self- diverse message triggers?

STUDY 1 – CONTROLLED LAB EXPERIMENT

In this first study, our goal was to test if our strategies can mitigate the negative responses from repeated exposure. Following the best practices of invoking the repeated exposure effects in the lab [33,38,48], we designed a between subjects study, with three conditions: target-diverse, self-diverse and the baseline non-diverse.

Participants were told that they will be asked to evaluate a draft of an informational website about health and nutrition. They were asked to read the content carefully as they will be asked questions about it at the end of the study. They were then shown a series of 4 web pages. Each web page contained health and nutrition information (100-150 words) and an associated image (e.g., presenting people running or eating healthy). We used the actual content from a university’s student health services’ webpage.

On each of the 4 pages the participants were shown 1 health tip consisting of an image (always the same for each

message and condition) and a pop-up text message trigger (Figure 2). In both the self- and target-diverse conditions, participants saw four different messages, one on every page. Each message connected different pair of concepts. Whereas in the baseline, non-diverse condition participants saw four messages always connecting the same pair of concepts, i.e. “Exercising can help you”, “Exercising is good for you”. Prior research shows that four messages in controlled study represent a sufficiently high repetition frequency for tedium to occur [48]. To protect against possible ordering effect, we counterbalanced the message order (for the manipulation conditions).

At the end participants received a survey about the triggers they just saw, including reactance and attention check questions and questions about the attitude and intention to exercise from Theory of Planned Behavior (TPB) [2].

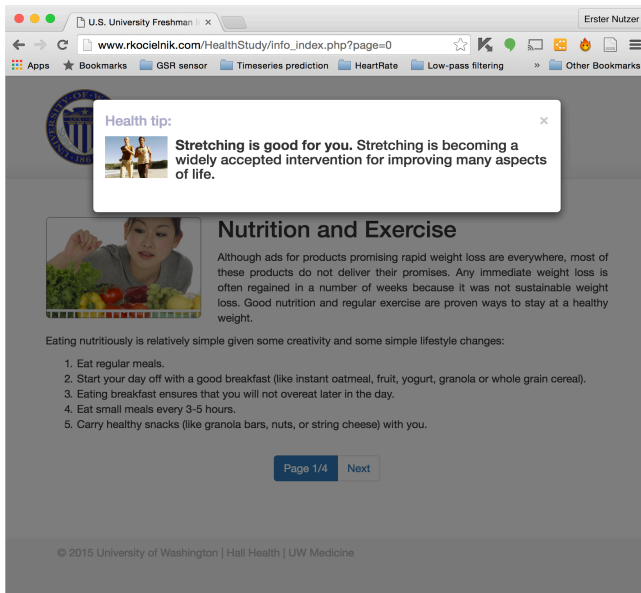


Figure 2. Screenshot of the webpage used in the lab study with a pop-up text message triggers.

Message Trigger Generation

In this study, our trigger generation process involved three steps: concept generation, concept selection and message generation (Table 1). In the first step, “concept generation”, we employed two different approaches to generate a list of concepts to be used in target- and self- diverse strategies.

The target-diverse strategies require a set of concepts related to the target behavior, exercising. We used the “ensemble” semantic relatedness (SR) measure available in WikiBrain to find an initial list of 8 concepts closest to “exercising” in terms of SR measure: “weight training”, “jogging”, “stretching”, “strength training”, “running”, “walking”, “aerobics”, “body building”. WikiBrain is one of the leading SR frameworks based on mining Wikipedia’s text and link structure [52]. This structure represents a shared knowledge of human contributors and curators, which allows the SR measure provided by WikiBrain to

achieve high correlation with manual human relatedness ratings [39]. Such manual relatedness ratings are the basis for building the cognitive structures and therefore use of automated SR score offers an efficient alternative.

The self-diverse strategies require a set of concepts related to the self. Our approach was to consider concepts that individuals care about (which are by definition close to self). To do so, we turned to prior literature on exercising [17], and identified 8 key motivations people have reported for exercising: “stress reduction”, “physical appearance”, “increased vigor”, “relaxation”, “health”, “fitness”, “pleasure”, “self-esteem.”

Then, we narrowed down the set of concepts to be used in developing our messages through a concept selection step. We validated these concepts’ spatial closeness by creating a cognitive space, following the procedure described in [36]. We recruited 159 Amazon Mechanical Turk workers (49% female) with median age range of 24-35 years, where 45.5% reported that they exercise regularly. Each received \$0.50 for evaluating the cognitive distances between 20 randomly selected pairs of concepts, given as an example a distance between “fit” and “obese” as 100.

Generation step	Target-diverse	Self-diverse
1. Concept generation	Semantic-relatedness (WikiBrain)	Past literature (exercising motivations)
2. Concept selection	Distances in the cognitive space (crowd based assessment)	
3. Message generation	Fixed message templates (4 templates, 8 diverse messages)	

Table 1. Message generation process used in the lab study.

Based on the distances between concepts in generated space, we selected the concepts to be used in the messages. For target-diverse strategy 3 concepts closest to “exercising”: “stretching”, “strength training”, “fitness” and for self-diverse strategy 3 concepts closest to “self”: “stress reduction”, “physical appearance” and “self-esteem.”

Condition	Message text
Non-diverse	“ Exercising can help you. Latest research has confirmed many of the anticipated benefits of exercising .”
Self-diverse	“ Exercising can help with improving self-esteem . Latest research has confirmed many of the anticipated benefits of improved self-esteem .”
Target-diverse	“ Strength training can help you. Latest research has confirmed many of the anticipated benefits of strength training .”

Table 2. Example messages used in the lab study. Parts of the messages that changed are made bold for better visibility.

Finally, in the “message generation” step, we created four generic message trigger templates, and swapped in the concepts respective to the strategy used (see Table 2 for example messages with concepts for each strategy). These messages were designed to be simple and similar to minimize any differences in their actual persuasiveness.

Hypotheses

Our proposed strategies make use of the related cognitive concepts for systematic message diversification. These diverse messages should not be less effective as behavior change triggers as they do not undermine the persuasive nature of the messages; the diverse concepts should still (indirectly) pull the behavior target closer to self. The diversity, however, should help minimize boredom and annoyance [27]. Additionally, research has suggested that message novelty may lead to higher perceived informativeness and helpfulness [15,54], which further helps to offset the negative effects of repeated exposure:

- **H1:** Annoyance towards the message-based triggers will be lower when using diverse strategies.
- **H2:** Boredom with the message-based triggers will be reduced when using diverse strategies.
- **H3:** Informativeness of the message-based triggers will be rated higher using diverse strategies.
- **H4:** Helpfulness of the message-based triggers will be rated higher using diverse strategies.

Participants

A total of 150 Amazon Mechanical Turk workers participated in our study (45% male) with a median age of 25-34, where 44% claimed they exercise regularly. Due to pure random assignment to experimental conditions, we had unequal number of participants per condition (non-diverse, n=44; self-diverse, n=52; target-diverse, n=54). We removed 28 participants from the analysis for failing an attention check (10 from non-diverse, 10 from self-diverse and 8 from target-diverse), which left 122 participants (non-diverse, n=34; self-diverse, n=42; target-diverse, n=46). The groups were not different in terms of age, gender, education, level of physical activity or ethnicity. Each participant was awarded \$2.20 for participation.

Measures

We measured annoyance (H1.1), boredom (H1.2), informativeness (H1.3), and helpfulness (H1.4) by asking the participants to estimate the experienced level of each towards the message contents on a 5-point likert scale. We measured behavior intention and attitude using TPB [2]. The reliability of both TPB measures were high, attitude: $\alpha=0.88$, intention: $\alpha=0.77$. We also measured reactance following procedure in [15].

One-way ANOVA was used to compare the effects of message strategies (conditions) on dependent measures. Post hoc comparisons were conducted to compare the conditions (Tukey HSD test).

Results

	Non-diverse (baseline)	Self-diverse	Target-diverse
Annoyance	3.85	3.07*	3.33
Boredom	3.79	3.17*	3.24
Informativeness	2.53	3.57***	3.26**
Helpfulness	2.44	3.57***	3.07*
Reactance	1.98	2.05	1.89
Attitude (TPB)	5.26	5.22	5.03
Intention (TPB)	5.15	5.26	4.88

Sig. compared to non-diverse: *** $p<0.001$, ** $p\leq 0.01$, * $p\leq 0.05$

Table 3. Summary of the results analyzing the differences between different conditions.

Both boredom and annoyance were highest in the non-diverse condition (Table 3), but only the difference between non-diverse and self-diverse conditions reached significance ($p=.04$ for both measures). There was almost a one-point difference between the two, where the 4 on the likert scale means “Agree” and a 3 on the likert scale means “Neutral.” While prior work treated annoyance and boredom as separate constructs, it is important to note that they were highly correlated in our study ($r=.72$, $p<.001$). H1 and H2 partially supported.

For informativeness, the self-diverse ($M=3.57$, $SD=0.97$) and target-diverse ($M=3.26$, $SD=1.10$) conditions both were rated significantly higher than the baseline, non-diverse condition ($M=2.53$, $SD=1.19$; $p<.001$ and $p<.01$ respectively). The difference between diverse conditions was not significant. H3 is supported.

For helpfulness, the self-diverse messages ($M=3.57$, $SD=0.97$) scored the highest, followed by the target-diverse ones ($M=3.07$, $SD=1.16$), and they were both significantly higher than the baseline, non-diverse message ($M=2.44$, $SD=1.21$; $p<.001$ and $p<.05$ respectively). H4 supported.

These results indicate that our message diversification strategies can mitigate the negative effects of repeated exposure. We also found that the self-diverse strategy performed better compared to the target-diverse strategy. It could be that the self-diverse strategy addresses more personally relevant issues and such personal relevance may render it less annoying and boring. Additionally, the particular concepts used in the self-diverse strategy e.g. “self-esteem” may invoke a set of rich experiences in any person, making them seem more contents rich [10].

We should also note that there were no significant differences in attitude and intention across conditions. While we do not want to elaborate on a null result, this might at least suggest the potential for these messages to preserve the persuasive nature of the messages even though they are not the same as the baseline message [48]. In the

case of target-diverse strategy, this is particularly interesting to note, as the majority of the messages actually do not try to persuade the participant towards the specific targeted behavior – exercising (but to related behaviors).

STUDY 2 – CONTROLLED FIELD DEPLOYMENT

There are two main limitations of the first study that we sought to address with a second study. First is the lack of realism. It involved only four messages and participants were not sent these message triggers in context. The second major limitation is that we collected only perceived measures. Do these message triggers actually affect participation? Therefore, we complemented the first study with 2-week-long study in an exercise context. We used a between subject design, comparing the stronger of our two strategies, the self-diverse, to the baseline, non-diverse.

Participants were invited to participate in a daily-challenges application that we built for the study. The application presents 4 daily exercises that the participants were asked to complete: 2-4 push-ups, 12-15 crunches, 12-15 lunges, and 12-15 jumping-jacks (these numbers were chosen using feedback from pilot studies). Participants were notified via a message trigger to perform these activities four times a day (the order of activities were randomized daily). Participants were asked to perform the activities at 9:30 am, 11:30 am, 1:30 pm and 3:30 pm. They also received a daily summary message at 6:00 pm, which linked them to a webpage showing their completions (Figure 3). Participants could mark their completion of the activities either directly through the communication channel (e.g., texting “done” back), or if they forgot, they could manually enter in their completion on our website.

Message Trigger Generation

In the field deployment each message was composed of motivational and exercise parts. The exercise part specified the randomly assigned exercise the participant was asked to perform, which was exactly the same in both conditions. As for the motivational part, in our manipulation condition this part was diversified following the self-diverse strategy as described below. In the baseline condition this part was the same every time: “Exercising is good for your health.”

To generate the self-diverse messages, we employed the message generation process used in study 1, with some modification (Table 4). Again, in the concept generation step, we needed concepts related to the self. However, since the concepts used in study 1 require a prior knowledge of motivation for performing the targeted behavior (i.e., prior research), that approach may not be available for all potential contexts. Instead, we sought for a more generalizable approach.

We explored the use of values as our concepts during concept generation. Values are also motivational constructs that have been shown to guide our behaviors in many facets of our lives [50]. In the same ways that certain motivations (e.g., physical appearance) may be close to the self, certain

values (e.g., achievement) are also close to self. Schwartz’ value theory, which has been widely studied and tested, suggests 10 universal values that all people relate to, to varying degree.







Generation step	Self-diverse
 1. Concept generation	Universal values framework (10 concepts)
 2. Concept selection	skipped (all 10 concepts used)
 3. Message generation	Crowd sourced generation workflow
 Creation	Prompts with pair of concepts (“exercising” and a value)
 Evaluation	“on topic” and “motivational” (3 raters per message)
 4. Message selection	Individual values scoring (survey based)

Table 4. Message generation process used in the filed deployment.

Because these ten values present a manageable number of concepts for message generation, we did not require a “concept selection” step to narrow down the concepts.

Another difference between the two studies’ processes is the “Message generation” step. In the field context, having only four messages (as was done in study 1) may not be sufficient in reducing tedium. We needed many more messages and therefore we explored the use of crowds to generate them. We devised a crowd-sourced workflow with quality assessment in the message generation step.

First, we prompt the crowd-workers on Amazon Mechanical Turk to write short 140 characters-long messages that would encourage an individual to perform a short exercise by connecting two given concepts (exercising and one of the ten values, “Creation” sub-step in Table 4). We then followed-up with quality assessment (“Evaluation” sub-step in Table 4). This step involved rating the messages in terms of being on topic (i.e. mentioning both concepts from the prompt) using a binary scale, and rating their motivational impact using a 7-point likert-scale. Each message has been rated by 3 independent workers and only the messages that were considered on topic by the majority of the raters and scored at least 5 (corresponding to “somewhat agree”) on the motivational impact have been selected for further use (see example messages in Table 5). Based on the described criteria, we selected 45 messages from the initially generated 200 (20 per value). We should note that some values seem to be harder to generate messages for: the number of messages meeting the selection criteria per value was different, with as many as 10 messages being selected for achievement and as few as 1 for universalism.



Figure 3. Screenshot of the exercise completion webpage used in the study and an example SMS message delivered on a mobile phone. These two and email were the two main communication channels used in our study.

Because people hold different values, to use the self-diverse strategy where the concepts are supposed to be close to self, we surveyed participants prior to the study assessing their value orientations, using the Portrait Values Questionnaire [49]. In the “Message selection” step for each participant, we then picked the three values that they most strongly identify with and randomly selected messages from that subset of value-based messages. Due to the fact that different number of messages met the quality criteria for each of the values, the number of unique messages for each participant also varied from as few as 7 to as many as 32 with an average per participant equal to 18.4 ($\sigma=7.46$). We control for this difference in the number of messages in our analyses, but also discuss their implications in the crowd message generation process in discussion.

Hypotheses

In addition to the 4 hypotheses from study one we also tested the messages’ effect on participation. Through mitigating the negative effects of repeated exposure and encouraging people to pay more attention to the message contents, the diverse messages may result in higher exercise completion.

- **H5:** Diversification will increase exercise completion.

Participants

We recruited participants online and through fliers distributed at a university campus. We also prescreened the participants based on the reported activity level to exclude people that have no interest in exercising and also the ones that have already been exercising regularly for at least 6 months. The reason for this is that people who have no motivation to exercise may need a motivation campaign rather than triggers [30]. As for the people that do exercise

regularly, they already have an established routine and may not need such triggers. As a result of these steps a total of 28 participants were recruited for our study (18% male) with a median age of 31 (range: 20-45), where 32% claimed

Value	Message text
Achievement	“To have a healthy and intelligent mind, one must have a healthy body.”
Benevolence	"Physical activity promotes general happiness and self love."
Conformity	"You can take care of ageing parents better by being healthy yourself."
Hedonism	"Exercising will make you feel and look better!"
Power	"People respect someone who makes a commitment to exercise!"
Security	"Exercising is extremely helpful for your body and mind."
Self-direction	"I exercise so that I have the freedom to do whatever I want to."
Stimulation	"Smell the air, touch the world, experience life. Exercise!"
Tradition	"Even modest effort can have measurable results."
Universalism	"Moving the body is activation of the brain & being one with nature."

Table 5. Example messages used in the field deployment.

they exercise regularly, but have only started doing so within the last 6 months. Each participant was awarded \$35 for participation and an additional \$20 if she agreed to participate in a follow up interview. We removed 1 participant from the analysis, who, due to health related problems, was unable to accomplish any exercise. This left 27 participants (non-diverse, n=14; self-diverse, n=13).

A stratified randomization was used to assign participants into two groups (self-diverse and non-diverse). We used the level of physical activity, evaluated by RM 1-FM: Physical Activity Stages of Change Questionnaire [37], as the stratification factor. This ensured that the people with different levels of reported physical activity are equally represented across the two conditions.

Measures

Most of the measures used in study 2 were the same as the ones used for study 1. This includes the questions on annoyance, boredom, informativeness, helpfulness, attitude ($\alpha=0.89$) and intention ($\alpha=0.74$). Additionally, we collected the self-reported exercise completion rating. Finally, at the end of the study, we asked all our participants to fill in an online survey and conducted one-hour semi-structured interviews with 14 of our participants that expressed interest in an interview.

In our analysis of pre-survey data, we found that despite stratified randomization, there was a weakly significant difference in pre-study behavior intention between the control and manipulation conditions ($M_1=5.83$, $M_2=6.42$, $p=0.06$). This may be a potential confound as it could be that it was participants' pre-study intention, and not our messages, that resulted in differences in exercise. Thus, we included the intention measure as a predictor in all our analyses. We found that its effect was not significant.

One-way ANOVA was used to compare the effect of message strategies (conditions) on self-reported dependent measures. To compare exercise completion between conditions, we constructed a number of different mixed-effects logistic regression models, using pre-test TPB intention, age, gender and exercise day as fixed effects and participant id as the random effect in all the models. For specific models we also included the condition, number of unique messages, message repetition count, and value orientation scores as fixed effects (Table 7).

We conducted analysis of the qualitative feedback from survey and interviews by assigning each sentence or paragraph an initial code following an open coding approach [44]. We then performed affinity diagramming in order to group the similar codes into themes. In this process we also revised the initial codes in order to merge the same codes and subdivide the codes that had more nuanced differences. We repeated this process a few times until we arrived at a stable set of minimal and unambiguous codes and themes following qualitative analysis steps outlined in [9].

Quantitative results

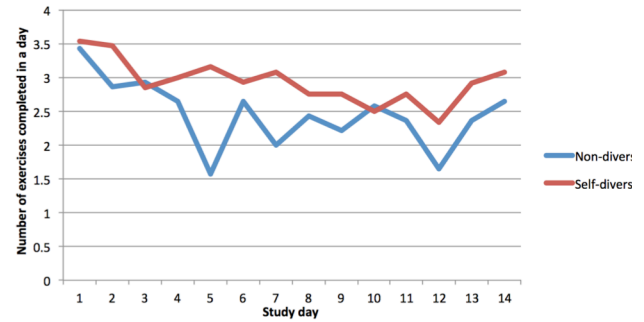


Figure 4 - Average self-reported exercise completion per study day per condition. For almost every day participants in the self-diverse condition completed more exercises. Big drops around day 5 and 12 represent weekends when participants reported traveling or meeting friends.

The main hypothesis for the field deployment was that self-diverse messages could increase exercise completion. We found that this hypothesis was supported. First, through visual inspection of the data, it appears that those in the self-diverse conditions completed more exercises in 12 out of the 14 days and the other 2 days the differences were negligible (Figure 4).

We explored two ways to analyze our data. First, we constructed a mixed-effects logistic regression model for predicting the completion of each prompted exercise (Model 1 in Table 7). We found that those in the self-diverse condition are 3.7 times more likely to exercise, but this result was only weakly significant ($p=0.09$). We also created another model, analyzing exercise completion per day. We binned the number of daily exercises completed into two levels using median split (0-2 completed as *fewer* and 3-4 completed as *more*) and used a similar mixed-effects logistic regression to predict likelihood to complete more exercises (Model 2 in Table 7). This model also shows that participants in the self-diverse condition

	Non-diverse (baseline)	Self-diverse
Exercise completion	2.45	2.90†
Annoyance	2.79	3.08
Boredom	3.21	3.08
Informativeness	2.93	2.62
Helpfulness	3.93	2.69***
Reactance	2.43	2.21
Perceived as different	2.00	3.00*

Significance: *** $p<0.001$, ** $p\leq 0.01$, * $p\leq 0.05$, † $p\leq 0.1$

Table 6. Summary of the results analyzing the differences between the conditions based on the post-study measures.

	Exercise completion		
	Model 1 (per exercise)	Model 2 (per day, 2 tiers)	Model 3 (per exercise, self-diverse only)
	Exp(B)	Exp(B)	Exp(B)
Condition	3.68†	6.46*	
Intention (TPB) - pre-test	0.63	0.57	1.76
Age	1.10	1.13	1.17
Gender	0.84	0.84	0.88
Exercise day	0.91***	0.89**	1.06
Repetition count			0.52***

Significance: *** p<0.001, ** p<0.01, * p<0.05, † p<0.1

Table 7. Mixed-effects regression models for predicting exercise completion.

exercised more (about 6.5 times more likely to complete 3-4 exercises, daily; p=0.04).

The fact that we had message repetitions in our self-diverse condition (we did not have the 4*14 messages needed for the full duration of the study), allowed us to more specifically test whether message repetition affects exercise completion. Using the self-diverse only dataset, we coded up how many times a specific message has been sent to an individual participant and used that as a predictor variable in our model (Model 3, repetition count in Table 7). We found that, controlling for the time progression of the study (exercise day), the number of repetitions was indeed a significant factor affecting exercise completion (p<0.001) with participants being 0.5 times less likely to complete the exercise with each message repetition. Interestingly, in this model, the effect of exercise day was not significant, suggesting that the decline in the self-diverse condition is mostly due to the message repetition than potential novelty effects associated with study participation.

We next turned to our post-study self-report measures (Table 6). We found that our initial hypotheses (H1-H4) were not supported. In fact, for helpfulness, we found the opposite effect as hypothesized, where participants in the non-diverse condition reported the messages as generally helpful (3.93) whereas those in the self-diverse conditions reported that the messages were in between neutral and not helpful (2.69, p<0.001). To help explain these potential conflicting results, and to explore whether and how these triggers helped, we turned to our qualitative results.

Qualitative results

Across both conditions, all the participants appreciated that the triggers reminded them to regularly perform exercises during the day. With some even pointing out that they would not have done any exercises during the day if they were not reminded of them. Even if the participants felt that

the triggers were pushing them a little, they still considered it to be a positive push.

“I liked that it was annoying me to do exercise. Kept nagging at me to do the few workouts needed (...)” – P5 (non-diverse)

Vast majority of participants explicitly mentioned that they liked that the messages were short and to the point. Many also commented they felt positive and encouraging. This shows that regardless of trigger type, having triggers was helpful in general.

Diversity helps

Diversification of messages in the experimental condition was generally appreciated. In-depth analysis of user feedback allowed us to identify three most common ways in which diversification was considered helpful: attracting attention, providing information and personal relevance.

Attracting attention: All the participants noticed that the contents of the messages were changing. These constant changes built an expectation of novelty each time a message arrived, which in turn increased attention, sustained engagement and interest in reading the messages: *“I would definitely skip over them if they were all the same message.” – P26 (self-diverse)*

Furthermore, the diversity of the messages led to increased curiosity and introduced a certain element of “fun”.

“With all different messages it's fun to read what they say. (...) That's a fun element to it I guess. I remember a few of them” – P22 (self-diverse)

Providing information: About half of our participants in the self-diverse condition liked the fact that the messages delivered small informational pieces about the different benefits of exercising.

"I liked that they talked about all the different benefits of exercises. One of them was learn about yourself, which I thought was cool. Something about endorphins." – P2 (self-diverse)

Also some of the information in the messages was not necessarily obvious to the participants and was therefore somewhat revealing. On top of that, the participants perceived such informational pieces as inspirational.

"I like that it seemed informative. Some of the things weren't really as obvious. It's kind of inspirational (...) Yeah, so like, there were some that were kind of more factual." – P15 (self-diverse)

Personal relevance: Most of the participants also resonated with specific messages or specific keywords in them. These participants specifically remembered selected messages that had some sort of personal value to them either by relating to their past or current experiences.

"I remember one, 'Being healthy yourself helps your aging parents better, as they get older.' (...) I think that one was the one that I picked up most, because my parents are getting old." – P24 (self-diverse)

In that sense the message diversity was very valuable for helping the participants cognitively elaborate on the personal value of exercising. It is worth noting that we have also proposed the ease of making the diverse messages personally relevant as an explanation of the higher informativeness and helpfulness ratings in our lab study.

Additionally, what also seems to have contributed to the sense of personal relevance was the fact that the messages felt as if a person wrote them.

"I did think that they were written by a person. Certainly, there's a sense of someone writing them for you at some point" – P3 (self-diverse)

Non-diverse worked as reminders; Self-diverse worked as reminders and motivators

Further analysis of participants' comments revealed that the triggers were perceived and evaluated differently between conditions. In the non-diverse condition, the participants almost immediately noticed that the motivational part of the message is fixed and would subsequently focus on just the part that changes – the exercise to complete.

"My first reaction was 'Yes, exercise is good for me'. Then because they never changed my brain just looped past entirely. (...) I stopped really paying attention to it so much and was going straight to what is the exercise." – P4 (non-diverse)

Consequently, they perceived the messages as simple reminders about exercising, which they considered helpful.

"It was actually really helpful to have little reminders! They were short enough that I didn't need a lot of motivation to do them." – P9 (non-diverse)

"The message is very straightforward. It's short. 'Exercising is good for you and then do something now'." – P25 (non-diverse)

"I think it was good, because it was quick. It was easy, very quick and easy to read, so you knew exactly what you had to do real quick'." – P13 (non-diverse)

On the other hand, those in the diverse condition employed a more critical evaluation of the contents. The changes from message to message solicited more attention from the participants. They scrutinized the messages more than those in the non-diverse condition to see whether the particular contents they received this time is actually appropriate and helpful to them given their context. They expected the messages to be somewhat intelligent and meaningful motivators rather than just simple automated reminders.

"It would be like 'Oh, nature is the same as exercise.' It's like what does that mean? (...) I love the self-help stuff. I love motivation, but it needs to actually make sense to me or seem like somewhat logical I guess" – P22 (self-diverse)

This effect, coupled with the fact that people tend to remember the negative more than the positive [4], resulted in the self-diverse participants recalling and focusing on incidents in which the messages were less helpful. They evaluated the messages not just as a reminder for which exercise they need to complete, but also for their motivational component.

Challenges in designing diversity

Despite many benefits, we also identified a number of challenges that designers need to consider to improve on the use of diverse messages. These are: quality of diversification, perpetual novelty and contextual relevance.

Opportunities to increase diversity: Despite the diversification, a number of participants still felt that the messages were not that different. These participants commented that they indeed noticed that the messages were technically different, but felt that they were also very similar in terms of tone and framing.

"(...) They seemed pretty similar in terms of being encouraging of exercise and talking about the different benefits. It seemed like they were the same in tone but certainly, each one was different." – P7 (self-diverse)

Many participants also commented about the practical value of information in the messages. Although the messages were generally perceived as presenting diverse information about the benefits of exercising, which was appreciated, a number of participants felt that the provided information

was not necessarily very revealing for them. They generally felt that they already knew most of the information.

"I think very few of the things were really novel to me, once they're saying, "Exercise to look and feel better," you kind of know that." – P16 (self-diverse)

Repetition is a problem: Making sure that the messages stay novel is important. For this 2-week study we did not have enough different messages to ensure that the participants will not receive the same message twice. Unfortunately, almost all the participants noticed this fact and it led to disappointment in each single case. The fact the messages were repeating, led the participants to lose interest in reading them. Majority reported that they started skipping the motivational part at and focused on the exercise they have to complete once they noticed that the messages are not novel.

"At first, you know ... I think I would have liked it a lot more if you guys had a whole new set of different messages. Make the reading more enjoyable (...) Later when the messages were ... Seems to be repeating I stopped reading them." – P11 (self-diverse)

It is worth noticing that this qualitative feedback is consistent with our qualitative analysis showing that the message repetition had a significant negative impact on exercise completion.

Contextual relevance is expected: All the participants also expected the diverse messages to fit the context of their activity, location and schedule. When this did not happen, participants immediately noticed such discrepancies.

"Sometimes the ... I feel like, at first the exercises and the quotes kind of go together. But later I realized it doesn't really have to do with anything. So, "Reach for the sun." type of message could have gone with jumping jacks or whatever." – P19 (self-diverse)

Discussion

In this research, we sought to evaluate and understand the feasibility of our two proposed strategies for diversification of messages. Through both studies, we found benefits in using our approaches. In a controlled setting, our strategies resulted in messages that were perceived to be more informative and helpful and the self-diverse strategy also reduces annoyance and boredom from repeated exposure. Applied in the field, the self-diverse strategy also led to an increase in behavior change compliance. These diversification strategies may be extended to help a number of social systems that use triggers to attract participation or contribution, and it can be more broadly applied to applications that encourage and facilitate self-improvement and behavior change.

Message Triggers Design

Through this research, we proposed two strategies and a general process to generate diverse trigger messages. These strategies provide different benefits making them useful under different circumstances. The self-diverse strategy seemed more effective in mitigating negative effect of repeated exposure. But the target-diverse strategy may also be useful in settings when the target concepts should not be addressed directly, e.g., antismoking campaigns talking about smoking may actually induce more smoking from smokers [26].

In relation to the messages contents itself, the qualitative feedback from our field deployment helped us identify that despite the measurable effectiveness of our diversification strategy, there are opportunities for improvement. One such direction relates to the possible use of different framings or tones for the messages. Currently all the messages are generally positively motivational. We could imagine, the prompts in our crowd sourced process for generating messages that are more focused on challenging the recipient, pointing out negative consequences of inaction or that employ social comparison for the purpose of triggering competitiveness or cooperation. This would increase the syntactic diversity of the messages within the framing of the prompted concepts.

Another aspects of the diversification that could further be improved relates to the repetition of the messages. Both quantitative and qualitative data indicated that when the messages started repeating at some point it had a measurable negative impact on exercise completion and participants' overall experience. Unfortunately, it may be impossible (and too costly) to generate an infinite number of diverse messages. There are, however, a number of options that we could explore. One, we could increase the perceived novelty of the messages following some of the techniques discussed in the previous paragraph. Another possibility is that given a sufficiently large set of messages, people might start forgetting past messages. There might be an optimal threshold for total message count dependent on use frequency. This will require additional work, but effectiveness of varying exposure has already been indicated in [29].

Finally, the current diversification and message delivery did not take into account the context in which the message is delivered. Many participants pointed out that they expected the messages to "match" the activity they are expected to perform or change in respect to the time of day or social setting they are in. Lack of such matching negatively affected the perception of personal relevance and introduced a sense of artificiality. We could address this mismatch, by prompting message generation for specific contexts in advance and then try to match these messages to the appropriate context; this would unfortunately increase generation costs. Another approach would be to

Generation step	Goals	Specific approaches explored	
		Self-diverse	Target-diverse
 Concept generation	<ul style="list-style-type: none"> Diverse, but still on topic 	<ul style="list-style-type: none"> Past literature Universal values 	<ul style="list-style-type: none"> Semantic-relatedness (WikiBrain)
 Concept selection ¹	<ul style="list-style-type: none"> Lowering cost (narrowing down the cognitive space) 	<ul style="list-style-type: none"> Distances in the cognitive space (crowd sourced) 	
 Message generation	<ul style="list-style-type: none"> Balancing cost and quality 	<ul style="list-style-type: none"> Fixed message template Manual writing (crowd based) 	
 Creation ²	<ul style="list-style-type: none"> Creativity, formulation diversity and natural feel 	<ul style="list-style-type: none"> Crowd prompts with concepts 	
 Evaluation ²	<ul style="list-style-type: none"> Ensuring quality 	<ul style="list-style-type: none"> Evaluation criteria: “on topic” and “motivational” 	
 Message selection ³	<ul style="list-style-type: none"> Personalization Context matching 	<ul style="list-style-type: none"> Individual values scoring 	

1 – this step is optional and mostly applicable if the number of concepts is too large to generate messages in a cost-effective manner.

2 – these steps are specific to the crowd-sourced message generation approach and may not be applicable to other methods.

3 – this step is optional and relates to selecting personalized messages or matching messages to the delivery context.

Figure 5. Summary of the proposed diverse message generation process based on the approaches explored in both studies.

automatically modify the already existing messages to make them more appropriate for the specific context [53].

Diverse Message Generation Process

One of our main research questions (RQ2) focused on exploring processes and workflows for generating diverse message triggers. Based on our experiences through study 1 and 2, we propose a four-stage process of systematic message trigger generation (Figure 5): 1) Concept generation, 2) Concept selection 3) Messages generation and 4) Message selection. We summarize the value and importance of each stage and discuss the benefits and limitations of using different approaches for realizing each stage in the following paragraphs.

Concept generation

The first step of the process with a goal to generate a diverse set of concepts related to the target concept. We proposed two generation strategies: target- and self-diverse.

We demonstrated the use of semantic relatedness measure for generating concepts in the target-diverse strategy. For this purpose we used semantic-relatedness measure available in WikiBrain as it uses the knowledge on Wikipedia to help assess concepts that are related to the target concept (in our studies, exercising). Other approaches that assess relatedness can be used to accomplish the same goals, specifically recent developments in NLP [23] and deep learning [32], such as word embeddings [34] could be used for that purpose.

For the self-diverse strategy, we used past-literature (study 1) and values framework (study 2) for generating personally relevant concepts. For contexts where the motivations are broad or there is no clear set of motivations to target, the

values framework offers an alternative strategy. The values framework offers a manageable set of universal values that people across cultures care about, just at varying degrees [49]. Generating messages triggers using these values can then result in a number of personally relevant messages.

Concept selection

The goal of this step is, if needed, to narrow down the size of the concept space. This is optional and mostly important for reducing the costs involved in executing the next stages.

To select the set of concepts one can focus on those most closely related to the target concept or self. This was done in study 1, where we selected the 3 most relevant concepts out of 8 initial ones based on the crowd-generated cognitive space. One thing to note, however, is that generating cognitive space required laborious comparisons of pairs of concepts to construct the cognitive space. In the near future, this process may be done algorithmically by selecting the set of concepts most closely related to either the target concept, or oneself, through some sort of semantic relatedness measures. We should point out that WikiBrain already offers a semantic relatedness rating [52].

Message generation

This step is where the generation of the messages using the concepts takes place. One appropriate approach is to use fixed sentence templates as we did in study 1, where related concepts replaced each other in different versions of the messages. This, however, can produce messages that are artificial. It also has limitation is terms of scaling to larger audiences in which case, the fixed template contributes to the feel of artificiality. Existing fully automated methods of natural text synthesis are still in their infancy [7].

In study 2 we explored a crowd based generation approach. This approach results in messages that seem much more natural. The messages are also likely to be more creative and semantically diverse, as crowd-workers have the freedom to weave in other concepts, aside from the ones prompted, into the message. Such tendency has already been observed in previous work, where crowd-workers introduced topics from personal experience [13]. Further, the crowdsourced approach is much more feasible when a for large number of messages, i.e. long-term exposure.

Message selection

This step focuses on selection of messages, from the generated message corpus, that are the most relevant for a particular participant or particular context of delivery. The goal of this step is to further increase the “natural feel” and personal relevance of the messages. In our work, we focused on personalization based on individuals’ values. We asked participants to fill out a short survey to assess their value orientations. Then we sent them a set of the messages that are more personally relevant. To reduce cost, future versions may be able to utilize recent NLP advancements in social media based personality profiling [11,25] as a replacement for such survey.

In addition to personalization, context matching can also be important as we have learned in our field deployment. We discovered that the pre-generated messages might not necessarily go well with some activities or specific social or time-based context. Such mismatch has been picked up by our participants and affected the perceived quality of the messages. We envision the use of automated context recognition in combination with message-context relevant evaluation to address this issue. It might also be valuable to include context information already in the “message generation” step so that the number of messages can be generated for the set of expected contexts. Another approach would be to use automated natural language generation techniques, to slightly alter the messages on the fly to make them fit the context better.

CONCLUSIONS & FUTURE WORK

In this work we show that the diverse message-based triggers generated using cognitive structures can effectively mitigate negative effects of repeated exposure and lead to measurable behavior change in the real-life conditions. We demonstrate two practical strategies for informing diversification: self-divers and target-diverse. Each with its distinct properties and scenarios of use. We offer practical support for informing the design of behavior change tools in a systematic and predictable manner. We also provide a set of practical suggestions for aspects to consider when designing mobile behavior change triggers based on qualitative feedback from our field study. Finally, our work proposes a trigger generation workflow that systematizes the creation of a diverse trigger contents. In the future we would like to improve our diversification approach following the feedback obtained in the field deployment.

We would also like to further explore and systematize the workflow for message generation, specifically in respect to the use of crowds.

REFERENCES

1. Wokje Abrahamse, Linda Steg, Charles Vlek, and Talib Rothengatter. 2005. A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology* 25, 273–291. <http://doi.org/10.1016/j.jenvp.2005.08.002>
2. Icek Ajzen. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50, 2: 179–211. [http://doi.org/10.1016/0749-5978\(91\)90020-T](http://doi.org/10.1016/0749-5978(91)90020-T)
3. Azy [Ed] Barak. 2008. *Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia*. Cambridge University Press.
4. Roy F. Baumeister, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. 2001. Bad is stronger than good. *Review of General Psychology* 5, 4: 323–370. <http://doi.org/10.1037/1089-2680.5.4.323>
5. D. E. Berlyne. 1970. Novelty, complexity, and hedonic value. *Perception & Psychophysics* 8, 5: 279–286. <http://doi.org/10.3758/BF03212593>
6. Elena Bessarabova, Edward L. Fink, and Monique Turner. 2013. Reactance, Restoration, and Cognitive Structure: Comparative Statics. *Human Communication Research* 39, 3: 339–364. <http://doi.org/10.1111/hcre.12007>
7. Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, and Samy Bengio. 2015. Generating Sentences from a Continuous Space. <http://arxiv.org/abs/1511.06349>
8. Rosalind A. Breslow, Ralph J. Coates, Jon Kerner, et al. 2008. Client-Directed Interventions to Increase Community Demand for Breast, Cervical, and Colorectal Cancer Screening: A Systematic Review. *American Journal of Preventive Medicine* 35, 1: S34–S55. <http://doi.org/10.1016/j.amepre.2008.04.002>
9. P. Burnard, P. Gill, K. Stewart, E. Treasure, and B. Chadwick. 2008. Analysing and presenting qualitative data. *BDJ* 204, 8: 429–432. <http://doi.org/10.1038/sj.bdj.2008.292>
10. John T Cacioppo and Richard E Petty. 1979. Effects of message repetition and position on cognitive response, recall, and persuasion. *Journal of Personality and Social Psychology* 37, 1: 97–109. <http://doi.org/http://dx.doi.org/10.1037/0022-3514.37.1.97>
11. Jilin Chen, Gary Hsieh, Jalal U. Mahmud, and Jeffrey Nichols. 2014. Understanding individuals’ personal values from social media word use. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*, 405–414. <http://doi.org/10.1145/2531602.2531608>

12. Yan Chen, F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li. Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens.
13. Heather L. Coley, Rajani S. Sadasivam, Jessica H. Williams, et al. 2013. Crowdsourced Peer- Versus Expert-Written Smoking-Cessation Messages. *American Journal of Preventive Medicine* 45, 5: 543–550. <http://doi.org/10.1016/j.amepre.2013.07.004>
14. Sunny Consolvo, Predrag Klasnja, David W. McDonald, and James A. Landay. 2014. Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness.
15. James Price Dillard and Lijiang Shen. 2005. On the Nature of Reactance and its Role in Persuasive Health Communication. *Communication Monographs* 72, 2: 144–168. <http://doi.org/10.1080/03637750500111815>
16. Leslie D. Dinauer and Edward L. Fink. 2005. Interattitude Structure and Attitude Dynamics. *Human Communication Research* 31, 1: 1–32. <http://doi.org/10.1111/j.1468-2958.2005.tb00863.x>
17. W Ebben and L Brudzynski. 2008. Motivations and barriers to exercise among college students. *Journal of Exercise Physiology Online*. <http://www.asep.org/asep/asep/EbbenJEPonlineOctober2008.pdf>
18. Ofer Egozi, Shaul Markovitch, and Evgeniy Gabrilovich. 2011. Concept-Based Information Retrieval Using Explicit Semantic Analysis. *ACM Transactions on Information Systems* 29, 2: 1–34. <http://doi.org/10.1145/1961209.1961211>
19. Douglas Ezzy. 2013. *Qualitative analysis*. Routledge (2013)
20. Edward Luis Fink and Joseph Woelfel. 1980. *The measurement of communication processes: Galileo theory and method*. Academic Press.
21. Brianna S. Fjeldsoe, Alison L. Marshall, and Yvette D. Miller. 2009. Behavior Change Interventions Delivered by Mobile Telephone Short-Message Service. *American Journal of Preventive Medicine* 36, 2: 165–173. <http://doi.org/10.1016/j.amepre.2008.09.040>
22. Bj Fogg. 2009. A behavior model for persuasive design. *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09*: 1. <http://doi.org/10.1145/1541948.1541999>
23. Shalini Ghosh, Oriol Vinyals, Brian Strope, Scott Roy, Tom Dean, and Larry Heck. 2016. Contextual LSTM: A Step towards Hierarchical Language Modeling.
24. K Glanz, BK Rimer, and K Viswanath. 2008. *Health behavior and health education: theory, research, and practice*.
25. Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, 253. <http://doi.org/10.1145/1979742.1979614>
26. Joseph Grandpre, Eusebio M. Alvaro, Michael Burgoon, Claude H. Miller, and John R. Hall. 2003. Adolescent Reactance and Anti-Smoking Campaigns: A Theoretical Approach. *Health Communication* 15, 3: 349–366. http://doi.org/10.1207/S15327027HC1503_6
27. Albert A. Harrison and Rick Crandall. 1972. Heterogeneity-homogeneity of exposure sequence and the attitudinal effect of exposure. *Journal of Personality and Social Psychology* 21, 2: 234–238. <http://doi.org/10.1037/h0032314>
28. Matthias R. Hastall and Silvia Knobloch-Westerwick. 2013. Severity, Efficacy, and Evidence Type as Determinants of Health Message Exposure. *Health Communication* 28, 4: 378–388. <http://doi.org/10.1080/10410236.2012.690175>
29. Katharine J. Head, Seth M. Noar, Nicholas T. Iannarino, and Nancy Grant Harrington. 2013. Efficacy of text messaging-based interventions for health promotion: A meta-analysis. *Social Science & Medicine* 97: 41–48. <http://doi.org/10.1016/j.socscimed.2013.08.003>
30. Lois Jane Heller, Celette Sugg Skinner, A. Janet Tomiyama, et al. 2013. Transtheoretical Model of Behavior Change. In *Encyclopedia of Behavioral Medicine*. Springer New York, New York, NY, 1997–2000. http://doi.org/10.1007/978-1-4419-1005-9_70
31. Guillaume Hervet, Katherine Guérard, Sébastien Tremblay, and Mohamed Saber Chtourou. 2011. Is banner blindness genuine? Eye tracking internet text advertising. *Applied Cognitive Psychology* 25, 5: 708–716. <http://doi.org/10.1002/acp.1742>
32. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553: 436–444. <http://doi.org/10.1038/nature14539>
33. Joowon Lee, Jae-Hyeon Ahn, and Byungho Park. 2015. The effect of repetition in Internet banner ads and the moderating role of animation. *Computers in Human Behavior* 46: 202–209. <http://doi.org/10.1016/j.chb.2015.01.008>
34. Omer Levy, Yoav Goldberg, and Ido Dagan. 2015. Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Transactions of the Association for Computational Linguistics* 3, 0: 211–225.
35. Yon Soo Lim. 2008. Persuasive Message Strategy for International Development Campaigns: A Theoretical Application of The Galileo Spatial-Linkage Model.
36. Yon Soo Lim. 2015. Exploring persuasive message strategy for international aid campaigns: potential donors' inter-attitudinal structure using Galileo model. *Quality & Quantity* 49, 4: 1397–1415. <http://doi.org/10.1007/s11135-014-0065-0>

37. Bess H Marcus and LeighAnn H Forsyth. 2013. *Motivating people to be physically active*. Human Kinetics.
38. Scott McCoy, Andrea Everard, Dennis F. Galletta, and Gregory D. Moody. 2016. Here we go again! The impact of website ad repetition on recall, intrusiveness, attitudes, and site revisit intentions. *Information & Management*. <http://doi.org/10.1016/j.im.2016.03.005>
39. David Milne and Ian H Witten. 2008. An Effective, Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links.
40. Sarah Milne, Sheina Orbell, and Paschal Sheeran. 2002. Combining motivational and volitional interventions to promote exercise participation: Protection motivation theory and implementation intentions. *British Journal of Health Psychology* 7, 2: 163–184. <http://doi.org/10.1348/135910702169420>
41. Felix Naughton, A Toby Prevost, Hazel Gilbert, and Stephen Sutton. 2012. Randomized controlled trial evaluation of a tailored leaflet and SMS text message self-help intervention for pregnant smokers (MiQuit). *Nicotine & tobacco research: official journal of the Society for Research on Nicotine and Tobacco* 14, 5: 569–77. <http://doi.org/10.1093/ntr/ntr254>
42. Pippa Norris. 2003. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*.
43. Anish Parikh, Carl Behnke, Mihaela Vorvoreanu, Barbara Almanza, and Doug Nelson. 2014. Motives for reading and articulating user-generated restaurant reviews on Yelp.com. *Journal of Hospitality and Tourism Technology* 5, 2: 160–176. <http://doi.org/10.1108/JHTT-04-2013-0011>
44. Michael Quinn Patton. 1990. *Qualitative evaluation and research methods, 2nd ed.* Sage Publications, Inc.
45. Cornelia Pechmann and David W. Stewart. 2012. Advertising Repetition: A Critical Review of Wearin and Wearout. *Current Issues and Research in Advertising*.
46. Reinhold Penner and Richard Caelers. 2012. Workrave. <http://www.workrave.org>
47. Mirko Tobias Schäfer. 2011. *Bastard Culture! How User Participation Transforms Cultural Production*. Amsterdam University Press, Amsterdam. <http://doi.org/10.5117/9789089642561>
48. David W. Schumann, Richard E. Petty, and D. Scott Clemons. 1990. Predicting the Effectiveness of Different Strategies of Advertising Variation: A Test of the Repetition-Variation Hypotheses. *Journal of Consumer Research* 17, 2: 192. <http://doi.org/10.1086/208549>
49. Shalom H. Schwartz, Jan Cieciuch, Michele Vecchione, et al. 2012. Refining the theory of basic individual values. *Journal of Personality and Social Psychology* 103, 4: 663–688. <http://doi.org/10.1037/a0029393>
50. Shalom H. Schwartz. 2012. An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture* 2, 1. <http://doi.org/10.9707/2307-0919.1116>
51. Andreas R. Schwerdtfeger, Catalina Schmitz, and Matthias Warken. 2012. Using Text Messages to Bridge the Intention-Behavior Gap? A Pilot Study on the Use of Text Message Reminders to Increase Objectively Assessed Physical Activity in Daily Life. *Frontiers in Psychology* 3: 270. <http://doi.org/10.3389/fpsyg.2012.00270>
52. Shilad Sen, Toby Jia-Jun Li, WikiBrain Team, and Brent Hecht. 2014. WikiBrain: Democratizing Computation on Wikipedia. In *Proceedings of The International Symposium on Open Collaboration - OpenSym '14*, 1–10. <http://doi.org/10.1145/2641580.2641615>
53. Alessandro Sordani, Michel Galley, Michael Auli, et al. 2015. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. Retrieved August 9, 2016 from <http://arxiv.org/abs/1506.06714>
54. Victor J. Strecher, Saul Shiffman, and Robert West. 2005. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction* 100, 5: 682–688. <http://doi.org/10.1111/j.1360-0443.2005.01093.x>
55. Esther Toet, Bernt Meerbeek, and Jettie Hoonhout. 2009. Supporting mindful eating: imbalance chopping board. *Eat, Cook, Grow: Mixing* 34, 51: 1–3.
56. Roger Tourangeau, Kenneth A Rasinski, and Roy D'Andrade. 1991. Attitude structure and belief accessibility. *Journal of Experimental Social Psychology* 27, 1: 48–75. [http://doi.org/10.1016/0022-1031\(91\)90010-4](http://doi.org/10.1016/0022-1031(91)90010-4)
57. Roelof A.J. de Vries, Khiet P. Truong, Sigrid Kwint, Constance H.C. Drossaert, and Vanessa Evers. 2016. Crowd-Designed Motivation: Motivational messages for exercise adherence based on behavior change theory. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 297–308. <http://doi.org/10.1145/2858036.2858229>
58. Bo Xu and Dahui Li. 2015. An empirical study of the motivations for content contribution and community participation in Wikipedia. *Information & Management* 52, 3: 275–286. <http://doi.org/10.1016/j.im.2014.12.003>
59. 2015. Aqualert. <http://aqualertapp.com/>