

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/129956>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Adaptive persuasive messaging to increase service retention: using persuasion profiles to increase the effectiveness of email reminders

Maurits Kaptein · Aart van Halteren

Received: 7 November 2011 / Accepted: 17 February 2012 / Published online: 11 July 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract In this article, we describe the usage of *persuasion profiles* in a large scale, $N = 1,129$, field trial. Persuasive technologies—technologies intentionally designed to influence user behavior—are emergent and becoming more and more individualized and ubiquitous. Individual differences in people’s responses to often used *persuasion principles*—different psychological means by which to influence users—motivate personalization. We describe how, through identification, representation, and measurement, persuasive technologies can personalize their persuasive attempts. Next, we show that dynamically adapting a persuasive technology to the responses of its users increases the effectiveness of the system. Ubiquitous computing systems are, because of their ability to unobtrusively measure user behavior, very well suited for these types of applications.

Keywords Persuasion · Personalization · Persuasion profiles

1 Introduction

We have entered an era of persuasive technology, of *interactive computing systems intentionally designed to change people’s attitudes and behaviors* [17]. These systems emerged for the first time around the 1980s with a

small selection of research prototypes of computing systems that were designed to promote health or increase workplace productivity [3]. Currently, persuasive technology researchers investigate and design systems in application areas that range from healthcare, to energy consumption, to e-commerce. A number of recent efforts have focussed on motivating people to lead an active and healthy lifestyle. These latter systems have now made it into the public domain: product-service combinations like Fitbug,¹ myZeo,² or Philips DirectLife³ all focus on unobtrusive measurement of users daily (or nightly) activities and providing motivating feedback to ensure a healthy lifestyle.

In this article, we detail the design and evaluation of the *Persuasive Messaging System* (PMS). The PMS is a persuasive system that is designed to increase the effectiveness of reminder emails that are sent out in a commercial activity promotion service. This persuasive application combines an “activity monitor” with active human- and technology-initiated coaching to help users gain a more active lifestyle. The activity monitor is a small and robust 3D accelerometer that users wear either in their pocket or on a necklace. Users can upload the collected data to the service’s backend that analyzes the activity data and calculates the associated activity energy expenditure (AEE). During a multiple-week program, users can set activity goals and monitor their progress on a web site that accompanies the product.

A key success factor for the health promotion service is user engagement: Feedback and progress is reported primarily via the web site, and activity data are only stored

M. Kaptein (✉)
Eindhoven University of Technology, Eindhoven,
The Netherlands
e-mail: maurits@mauritskaptein.com

A. van Halteren
Philips Research, Eindhoven, The Netherlands

¹ <http://www.fitbug.com>.

² <http://www.myzeo.com>.

³ <http://www.directlife.philips.com/>.

and analyzed after it has been uploaded. Uploading takes place via a *physical* connection of the activity monitor to the users computer. Users that fail to upload are thus deprived of feedback and coaching and consequently reduce the benefits of carrying an activity monitor. To encourage docking—the uploading of the activity data to the web service—so-called *docking reminders* are sent via email to users that have failed to upload for a certain number of days.

In this article, we focus on increasing the effectiveness of these docking reminders by using persuasive messages: messages that implement persuasion principles as identified by Cialdini [7]. We test whether personalization of the selection of these messages benefits their effectiveness.

1.1 Persuasion principles

The array of persuasion principles or influence tactics that can be used to change attitudes and behaviors of users can be overwhelming. Both researchers and practitioners have made extensive use of the categorization of persuasive messages as implementations of more general *influence principles*. Theorists have varied in how they individuate these influence principles: Cialdini [7, 8] develops six principles at length, Fogg [11] describes 40 strategies under a more general definition of persuasion, Kellermann and Cole [28] gather 64 groups from several taxonomies, and others have listed over 100 [35]. These different counts result from differing levels of exhaustiveness, exclusivity, emphasis, and granularity [28].

This article focusses on the six persuasion principles described by Cialdini [7]. These principles are:

1. *Reciprocity*: People feel obligated to return a favor; thus, when a persuasive request is made by a person the receiver is in debt to, the receiver is more inclined to adhere to the request Cialdini [8]. People even return a favor that they never asked for [18].
2. *Scarcity*: When something is scarce, people will value it more. Announcing that a product or service is scarce will favor the evaluation and increase the chance of purchase [42].
3. *Authority*: When a request or statement is made by a legitimate authority, people are more inclined to comply or find the information credible [32].
4. *Commitment and consistency*: People do as they said they would. People try to be consistent with previous or reported behavior, resolving cognitive dissonance by changing their attitudes or behaviors to achieve consistency. If a persuasive request aligns with previous behavior, people are more inclined to comply [7, 9].
5. *Consensus*: People do as other people do. When a persuasive request is made, people are more inclined to

comply when they are aware that others have complied as well Cialdini [8]; Ajzen and Fishbein [1].

6. *Liking*: We say “yes” to people we like. When a request is made by someone we like, we are more inclined to act accordingly [7].

While most probably not exhaustive, nor in all cases mutually exclusive, these six persuasion principles provide a concrete means to classify influence attempts. Furthermore, implementations of each of these principles have been shown to be effective in multiple contexts. Interestingly, all of these influence strategies are related to *how* a certain attitude or behavioral change request is made, and not necessarily tied to what the actual request is Kaptein and Eckles [23]. This enables us to distinguish the *end* of a request (e.g., a persuasive application urges you to work out more) from the *means* in which the request is made (e.g., by showing you how your friends are working out, or by giving you expert advice). This property makes persuasion principles useful not just for typifying a specific influence attempt, but more broadly to function as a level of analysis to describe and predict the effects of different implementations of the same principle at later points in time or in a different context.

Investigators in psychology often explain and predict how implementations of persuasion principles affect user attitudes using dual-process models. According to the Elaboration Likelihood Model (ELM) [5, 33, 34], persuasive messages can affect attitudes through both central and peripheral routes. The central route is characterized by elaboration on and consideration of the merits of presented arguments. On the other hand, the peripheral route is characterized by responses to cues associated with, but peripheral to the central arguments of, the advocacy. The latter occurs through the application of simple, cognitively “cheap,” but fallible rules. Frequently, the use of these cognitively “cheap” rules leads to a fast and relatively accurate appraisal of the merits of the appeal: If (e.g.) a product is “almost out of stock,” a large number of prior customers may have bought the product based on product merits and opportunities to buy in the future may be rare or high cost [40]. Thus, without engaging in full and cognitively costly processing, a user can make a choice based on an accurate peripheral cue [12].

1.2 Individual differences

Despite the large body of work investigating persuasion principles and the theoretical models such as the ELM to explain their effectiveness, researchers have had serious difficulties in replicating previous findings. For example, a thorough meta-analysis [20] of the research on the effects of argument strength on persuasion—as frequently used in

ELM research to appeal to either peripheral or central processing—has found mixed results. Because of these, and other, difficulties in replications, researchers have investigated properties of context, messages, and individuals to further understand persuasion processes.

Much of the work on individual differences in persuasion has directly drawn on dual-process models—and the ELM in particular—to work out how new or established traits could moderate persuasion. Many of these studies have examined trait differences in motivations, such as Need for Cognition (NFC [5]), that affect differences in peripheral and central processing of persuasive messages. Thus, NFC predicts differences in the effects of argument strength on attitudes, the degree to which individuals rely on product characteristics versus source liking (e.g., [15]), attitude strength resulting from processing a persuasive message (e.g., [14]), and metacognition in persuasion (e.g., [39]). More generally, for many user choice settings in which personal relevance is neither very low or very high, elaborative processing of stimuli varies with NFC, such that NFC measures an individual difference in propensity to scrutinize and elaborate on arguments via the central route [4].

While NFC is the most widely used trait that operationalizes stable motivational heterogeneity in dual-process models, several relating traits have been identified and studied [13]. Measures of individuals' need for closure [41], need to evaluate [19], and need for affect [31] have all received attention in the persuasion literature. More recently, Kaptein and Eckles [24] have explicitly examined individual-level responses to difference influence strategies. They find that large individual differences exist in people's responses to different persuasive strategies. In their work, it shows that while (e.g.) the authority strategy is effective to increase compliance on average, it can be counter-effective for up to 35% of the population. These findings combined motivate adapting the use of distinct influence strategies—the different means to a common end—to individual users of persuasive systems.

1.3 Overview

In this paper, we describe the development and evaluation of the PMS. This system provides persuasive content to be used in the docking reminder emails that are sent out to remind participants to upload activity data. We use the six persuasion principles described above as our basis to design multiple docking reminders. Next, because of the large individual differences that have been found in people's responses to these persuasion principles, we try to adapt the principle that is used in the docking reminder for a specific user to the behavioral response—a successful docking event—by that user. We thus develop an adaptive

persuasive system (see also [21, 36]). In the remainder of this article, we first describe the requirements of adaptive persuasive systems and their possible implementations. We describe how designers of persuasive systems can adapt the persuasive principles they use to motivate users for certain goals to their effectiveness at an individual level. Next, we describe how the PMS was built according to these specifications. Finally, we describe the results of a 6 months long evaluation of the PMS systems with $N = 1,129$ users.

2 Designing adaptive persuasive systems

Given the large individual differences in response to persuasion principles, we believe that designers should take these individual differences into account. This is especially relevant in the light of the work presented by Kaptein and Eckles [24]. Even more [25] show that the use of multiple influence strategies at the same time is not necessarily beneficial for compliance, and thus, designers of persuasive technology should choose the right influence principle for the right individual. These results emphasize the need for designers of persuasive technologies to attend to, and adapt to, individual differences in response to persuasion. We call this class of systems *adaptive persuasive technologies*.

2.1 Requirements of adaptive persuasive systems

When designers attend individual differences in user response to the use of persuasion principles in persuasive systems, then they will design adaptive persuasive systems: “systems that select the appropriate influence strategy to use for a specific user based on the estimated success of this strategy.” To be able to build adaptive persuasive systems, designers should create systems that are capable of identifying their users, representing different social influence strategies, and measuring their effectiveness (cf. [22, 27, 36]). We detail each of these requirements.

2.1.1 Identification

To be able to adapt to individual differences in response to social influence strategies, a system must be able to uniquely identify individuals.⁴ Only once a user has been identified can the influence strategy that is used to support a persuasive appeal be adapted to this user. Currently, many means of identification exist. In online marketing contexts, cookies are frequently used to tailor appeals, and this practice can easily be extended to tailor the choice of

⁴ It is not necessary to know the true identity of the person, as long as each interaction with the system uniquely distinguishes between individuals.

persuasion principles to specific individuals. However, in a ubiquitous computing scenario, the possibilities of identification are more diverse: Designers have used the unique bluetooth key used by mobile devices [29], face recognition [16], or fingerprints [6] to identify individual users. When these identification mechanisms are combined, individual users can be tailored to both offline as well as online, and this type of personalization be used over a multitude of persuasive applications.

2.1.2 Representation

Adaptive persuasive systems have to be able to represent one *end* via multiple *means*. Thus, the system needs to be able to implement various persuasion principles. For example, a digital exercise coach can influence users to exercise more by having users set targets (commitment principle), coupling users to others (consensus principle), or by providing advice from a fitness instructor (authority principle). To enable personalization, systems should have the flexibility to present their end goal (e.g., work out more) in different ways to users. In the system architecture, designers should distinguish between different persuasion principles and their respective implementations. Thus, if a persuasive system uses the authority strategy, then still different expert sources could be used, via different communication channels, to influence users. In each case, the authority strategy is represented by a different implementation.

2.1.3 Measurement of success

When designers create systems that adapt to users responses *dynamically*—for example, during the lifetime of the product—it needs to be possible to measure the effect of a persuasive principles on an individual user. While this sounds straightforward, it is not always easy to measure whether an appeal was successful, or even to determine what a measure of success would entail. For example, in a digital exercise coach, a prompt by a fitness instructor to run for 30 min that is followed by the user running for 20 min 14 h after the prompt might constitute a partial success—indicating the success of the authority strategy—but might also be due to external causes. Furthermore, technologically not all behavioral responses are easily or reliably measured.

2.2 Realizing dynamic adaptation

Once the three prerequisites identified above are met, and thus a persuasive system is able to identify its users, represent different social influence strategies, and measure the effect of the persuasion principles, then the system can be

made to adapt to user responses. While different machine learning algorithms could be used for such a goal, this section briefly presents a simple self-learning system capable of using individual-level responses by considering an example in which identification, representation, and measurement are relatively easy. The description below details how individual-level estimates of the success of different influence strategies can be used for personalization. We call this collection of estimates, in line with Kaptein and Eckles [24], a *persuasion profile*.

Consider, for example, a ubiquitous persuasive system designed to encourage users to save energy by using a revolving door (which keeps the heat in) instead of a sliding door that is next to it. This setup is common in hotels and office buildings and often one can find a paper sign motivating visitors to indeed take the revolving door. Face recognition, by the use of camera's, could potentially be used for identification in this scenario. This same identification method can also be used to measure the effectiveness of each persuasive attempt: Through face recognition, one could determine which entrance was used by the current visitor. Based on this knowledge about the visitor and records of earlier decisions, a message implementing the right influence strategy can be selected and displayed on a screen instead of the paper sign.

The probability of a single visitor taking the revolving door on multiple occasions can be regarded as a binomial random variable $B(n, p)$, where n denotes the number of approaches the visitor has made to the doors, and p denotes the probability of success: the probability of taking the revolving door. Given M different messages, one can compute for each individual, for each message, probability $p_m = k_m/n_m$, where k_m is the number of observed successes after representation of message m , n_m times to a specific visitor. It makes intuitive sense to present an approaching visitor with the messages with the highest probability of success, thus the message where p_m is highest. However, this will not inform a decision for a newly observed visitor. For a *new* visitor, one would present the message m for which p_m is maximized for previously observed visitors. Actually—given Stein's result [10, 38]—for every user, a weighted average of the p_m for an individual user and those of other users—one where the estimated \hat{p}_m for an individual is “shrunk” toward the population mean—will provide a better estimate than an estimate based on observations of a single visitor alone. For example, if the authority message is effective 70 % of the time over all visitors and only 30 % percent of the time for the specific visitor under consideration, the best estimate of the (real) effectiveness of the authority message \hat{p}_A for this visitor is a weighted average of these two.

2.2.1 Adapting to individual behavior

To include both the known effectiveness of a message for others, and a specific visitors previous responses to that same message, into a new estimate of message effectiveness, p_m , designers can use a Bayesian approach. A common way of including prior information in a binomial random process is to use the Beta-Binomial model [43]. The $Beta(\alpha, \beta)$ distribution functions as a conjugate prior to the binomial. The beta distribution can be re-parametrized as follows:

$$\pi(\theta|\mu, M) = \text{Beta}(\mu, M)$$

where $\mu = \frac{\alpha}{\alpha+\beta}$ and $M = \alpha + \beta$, then the expected value of the distribution is given by: $E(\theta|\mu, M) = \mu_m$. In our specific scenario, μ_m represents the expected probability of a successful influence attempt by a specific message. The certainty of this estimated success probability is represented by:

$$\text{Var}(\theta|\mu, M) = \sigma^2 = \frac{\mu(1 - \mu)}{M + 1}.$$

After specifying the probability of success μ_m of message m and the certainty about this estimate σ_m^2 as the prior expectancy about the effectiveness of a specific message and updating this expectancy by multiplying it by the likelihood of the observations, one obtains the posterior expectation:

$$\begin{aligned} p(\theta|k) &\propto l(k|\theta)\pi(\theta|\mu, M) \\ &= \text{Beta}(k + M\mu, n - k + M(1 - \mu)), \end{aligned}$$

in which $k, 0, 1$, is the outcome of the new observation. The newly obtained Beta distribution, $B(\mu, M)$, functions as our probability distribution with a new point estimate of the effectiveness of the presented message given by:

$$E(\theta|k) = \frac{k + M\mu}{n + M}.$$

2.2.2 Decision rule to choose a persuasive strategy

The Beta-Binomial model described above allows estimation of the effectiveness of message m , including prior knowledge, and updating these estimates based on new observations. As such, one can maintain a record of both the point estimate, μ_m , and its certainty, σ_m^2 , for each specific visitor. To determine which message to present next, one could pick the message which has the highest μ_m . However, if σ_m^2 is large, this decision rule might not be feasible given that—from a frequentist perspective—the difference between effectiveness estimates might not be statistically significant. Thus, while one would like to *exploit* the obtained estimates by selecting the most optimal

strategy, one should avoid making a selection based on too limited or noisy observations. In the case of a small number of observations, one would be better off to further *explore* responses to multiple strategies.

A recent solution to this selection problem is presented by Scott [37]. His proposed *randomized probability sampling* method depends on obtaining a single draw of each of the Beta distributions for each strategy and comparing these draws. At a specific occasion, the strategy representing the highest draw will be shown. Scott [37] shows through simulation that this strategy of selecting from competing random variables with differing levels of uncertainty provides an almost optimal solution to the explore/exploit problem [30]: Randomized probability matching ensure on one hand that a single strategy is not selected too early (and possibly erroneously) while it also ensures an effective usage of all the available information.

2.2.3 Persuasion profiles

The estimates of the effectiveness of different messages—or of the persuasion principles they implement—create a profile for each user. This profile, in prior work called a *Persuasion Profile* [23] can, via a decision rule like the one specified above, be used for persuasion principle selection for individual users. While initial attempts to create these types of systems have already been reported on in the literature [26, 36], no large-scale evaluations of this idea of using persuasion profiles to increase the effectiveness of persuasive technologies exist to date.

3 Design of the PMS

The PMS system was created to test the use of persuasion profiles for docking reminders. In this section, we describe the development and implementation of the PMS. As motivated in the previous section, for any adaptive persuasive system, *identification, representation, and measurements* are necessary requirements to build and use persuasion profiles.

3.1 Identification

The PMS used a unique one-way hashed identifier for each individual user. When a user docks—connects their activity monitor to their computer—the a timestamp of this event and a one-way hash of the user identifier were sent to the persuasive system. For operational reasons, the PMS was implemented on another server that was external to the server of health promotion service. The one-way hash ensured that no personally identifiable information of the participants was stored on the PMS server while at the

Table 1 Persuasion principles and their implementations in the PMS

Principle	Implementation (persuasive paragraph)
Neutral	0.
Scarcity	1. We would like to remind you to connect it to your PC soon and stay in touch with Today is a great day to stay fit so make sure you do not miss out on your participation in ...! 2. Any chance to connect your activity monitor is a chance to learn about your own activities. Take the opportunity to learn about your activities right now.
Authority	3. Experienced coaches recommend frequent uploads of your activity data. This will help you to gain more insight and be more active! 4. Activity experts recommend moderate to high activity on a daily basis and connecting to the ... platform will help you to reach this target!
Consensus	5. People like you who connect their activity monitor frequently with their PC are more likely to benefit from the program and obtain a healthy lifestyle! 6. Thousands of people are participating actively in the ... program and they stay connected at least once a week. Join the group!

same time, the PMS server could log each docking event for each individual user.

3.2 Representation

Representation of the persuasion principles was done in the email reminders that were sent to users that had refrained from docking for either three or 6 days. To create the persuasion principle implementations, five persuasive technology researchers brainstormed a large number of messages. Messages were created that implemented the scarcity, authority, and consensus principles. After the brainstorm, a card-sorting test was used to classify messages according to their strategies, and for each principle, two messages were selected for use in the trial. The persuasive messages consisted of text snippets containing persuasion principles that were added to the standard docking reminder email. This standard reminder mail looked as follows:⁵

Dear (first-name⁶),

How are you doing? We hope all is well. It is 3 days since the last time you connected your Activity Monitor.

[Persuasive paragraph]

We would like to remind you to connect it to your PC soon and stay in touch.

Sincerely,

The ...⁷ Team

⁵ The email reminders and social influence implementations were translated into several languages to facilitate the various languages of the users of the service.

⁶ The first-name of the user was filled out by the activity promotion service.

When a reminder was due, the health promotion service would request the PMS server for the next social influence text snippet to be used for the current user (identified by their hashed ID). The PMS server, upon receipt of the request, looked up the persuasion profile for that user (all stored using the hashed user ID's) and returned the appropriate persuasive text snippet.

The text snippet was inserted at the *[Persuasive paragraph]* location of the reminder email. Table 1 gives the implementations of the social influence strategies as used in the PMS. Since the original docking reminder was also used, there are four different types of messages in use (one not containing a persuasion principle). By combining the hashed user ID with the message ID, the docking reminder server was able to dynamically construct a message for a specific individual user's of the activity promotion service.

To enable estimation of the possible effect of these messages, each of the messages was presented to $N = 80$ participants in a pretest. Participants were instructed to read each of the (full) messages and answer the question *This message would motivate me* on a seven-point (Totally Disagree (1) to Totally Agree (7)) scale. Scores over two implementations of the social influence strategies were averaged, and mean scores for each strategy were subsequently used to estimate the successfulness of the different social influence strategies at an average level. The neutral message had the lowest evaluation: $\bar{X} = 3.46, SD = 1.44$. The messages implementing social influence strategies scored only slightly higher, with authority scoring highest, $\bar{X} = 4.21, SD = 1.59$, before consensus, $\bar{X} = 3.96, SD = 1.54$, and scarcity, $\bar{X} = 3.81, SD = 1.52$. Given the range of the scale, the persuasive messages seem not too convincing.

⁷ The reference to the name of the service has been removed for confidentiality. We are grateful for their participation in this trial, but since the PMS is no longer in use, we do not want to associate the service itself with the presented research.

However, they do score significantly higher than the neutral message ($p < .05$ using paired t-test for each pair).

3.3 Measurement

The docking reminder server, after consulting the PMS, sends emails to remind users to dock their activity monitor. Hence, the reminder message containing a specific persuasion strategy is successful if, within a certain time period after reading the email, the activity monitor is indeed docked. To measure this effectiveness, a small image was inserted into the email message body that allowed the PMS to log the fact that a user with a specific hashed ID opened an email. If, and only if, within 24 h after opening the email the user with that ID docked her activity monitor the message was considered a success, and thus, the persuasion principle that was implemented in the message (neutral, scarcity, authority, or consensus) was regarded successful for that user. The PMS ran a cron-job every 24 h to match all opened emails with the recent docking behavior. Next, the PMS updated the individual-level persuasion profiles according to the responses to messages send the last 24 h.

4 Evaluation of the PMS system

To evaluate the PMS system, an experimental comparison was setup in which the system was deployed for a selection of *new* users of the activity promotion service in the period the 1st of January 2011 until the 1st of July 2011.

4.1 Method

4.1.1 Procedure

From January 1st onwards, new users of the activity promotion service were included in the experimental evaluation of the PMS. Upon the first upload of their activity data, users were randomly assigned to one of the four conditions (see Sect. 4.1.2) by performing a random draw from a four-level multinomial with equal probabilities for each level. The assignment to one of the four conditions was decisive for the messages that users received later on. All users that joined the service between the start-date and the 1st of June 2011 were considered for inclusion in the trial. Users received a reminder messages after 3 days of inactivity and after 6 days of inactivity (not uploading their data). After these, two messages had been send and found unsuccessful users would not receive any new messages unless they had been active again (uploaded their activity data). After a new activity upload, the counters for the reminders were reset and a new reminder message was sent out after 3 days

of inactivity. For the analysis of the data, only users who received *at least* 3 email reminders during the trial period were included. Given this selection scheme, participation in the trial ranged from 1 to 6 months.

Besides differences in the email messages that were sent out to users, their usage of the activity promotion service as a whole was similar in each of the four experimental conditions. The data collected in the PMS evaluation consisted solely of a description of the email that was opened, with a record of the *Condition* that the user was in as well as the persuasion *Principle* that was used in the message, and a record of the subsequent response (*Success or failure*). We further recorded the timestamp of opening of the email as well as the *Number* of the reminders: This figure indexed how many reminders a user had received. The timestamp allows us to compute the date of the first message that was received and from there compute the *Time in Trial* (in days) for each participant. Finally, at the level of users, we marked those users that had not docked during the last 30 days of the trail as *Dropout*. This definition for dropout is also used by the management of the activity promotion service.

4.1.2 Conditions

To test the performance of the PMS as opposed to different message selection scheme's, users were distributed over four conditions:

1. *Control*: Users assigned to this condition received the standard docking reminder. This message did not contain any implementations of persuasion principles. This condition was included to be able to compare the PMS to the current reminder message.
2. *Best Pretested*: Users assigned to this condition received randomly one of the two messages implementing the authority advice—this message was judged most motivating in the pretest evaluation of the messages. This condition was included to compare adaptive selection of social influence strategies to the “best” average strategy.
3. *Random*: Users assigned to this condition received randomly one out of the seven versions of the message (with probabilities equal for each of the principles). This condition was included to compare adaptive messaging with alternating messages.
4. *Adaptive*: Users assigned to this condition received messages suggested by the adaptive persuasive system algorithm as in the previous section. Thus, for the first few messages, the selection was random. If users displayed a clear preference for one of the persuasion principle after receiving multiple reminder emails, the

reminder message was adapted to include only those strategies users were susceptible to.

Comparison of the adaptive condition to the control condition allows estimating the applied value of using the PMS to personalized messages over the current use of reminder messages in the activity promotion service system. Comparison of the adaptive condition with the random condition serves to examine the benefit of using a self-learning system, as opposed to merely using different influence strategies in the reminder messages. Finally, comparison of the adaptive condition with the best-pretested condition allows estimation of the benefits of using the PMS over selecting the most promising message based on a questionnaire.

In the *adaptive* condition, the prior expectancy of the success of the different social influence strategies had to be set. Before the trial, no information was available about the effects of the reminder message, and thus, the estimates were (a) set close together and (b) set with large uncertainty to be updated quickly by new data. The prior for the neutral (no social influence) message was set to: $\bar{X} = 0.39$, $\text{Var} = 0.1$. In line with the pretest of the messages, the authority strategy prior was set the highest, $\bar{X} = 0.52$, $\text{Var} = 0.1$, before consensus, $\bar{X} = 0.50$, $\text{Var} = 0.1$ and scarcity, $\bar{X} = 0.47$, $\text{Var} = 0.1$. Randomized probability matching was used to select messages in the adaptive condition.

4.1.3 Participants

Since the company that markets the activity promotion service on which this trial was ran is, understandably, very careful with the personal data of their users, we did *not* gain access to any personal information of the users of the system. Hence, while we could identify users based on a unique ID, we could not link this ID to any demographic or actual activity data. We were only able to record the emails that were sent as well as the docking behavior within 24 h after opening the email. This limits the possibilities of exploring possible interesting demographics that influenced our results and also inhibits us from reporting background information about the participants in the trial such as age or gender. However, given the random assignment over the conditions, we believe that the results presented below do provide a valid test of the effectiveness of the reminder messages in the different conditions.

4.2 Results

For the period of the evaluation, this led to a data set describing the upload frequency and responses to reminders of 1,129 users. Figure 1 shows a histogram of the

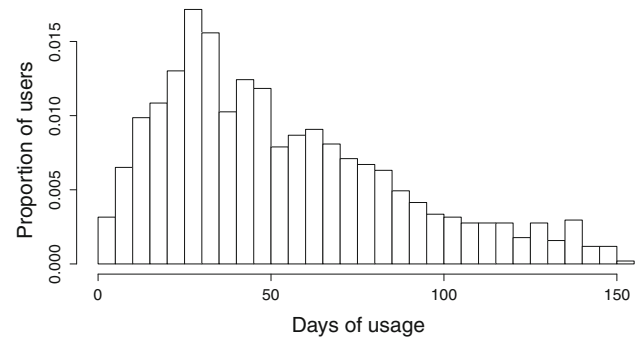


Fig. 1 Overview of the number of days that users who were included in the experiment were active using the persuasive application

Table 2 Overview of the raw data from the PMS evaluation

Condition	Average No. Rem. (S.E.)	% Success (S.E.)	% Dropouts (S.E.)
Control	9.13 (0.33)	28.5 (1.7)	25.1 (2.6)
Best pretest	8.97 (0.32)	24.2 (1.5)	26.0 (2.6)
Random	8.92 (0.33)	25.6 (1.6)	20.8 (2.4)
Adaptive	9.41 (0.35)	26.3 (1.6)	18.9 (2.3)

Presented are the average number of reminders, the success rate of the reminders, and the percentage of dropouts in each condition

number of days users included in the trial spend using the application. Since users were added continuously to the experiment as they started using the service, this histogram shows both those who entered late and those who dropped out. To give further insight into the raw data, Table 2 gives an overview of the average number of reminders sent, the mean success percentage, and the number of dropouts in each condition.

4.2.1 Effectiveness of the messages

To analyze the data obtained in the PMS evaluation, a series of multilevel models is fit to the data predicting the successes of each of the reminders sent to users included in the trial. We first fit a “null” model [2] to the data, which models the success of emails via a logit link using an overall intercept and individual-level intercepts that are distributed normally. This “null” model can be written as:

$$Pr(y_{ij} = 1) = \text{logit}^{-1}(\alpha_{j|i})$$

for $i = 1, \dots, n$ messages and where $\alpha_j \sim (\mu_\alpha, \sigma_{\text{subject}}^2)$. Thus, the probability of a success for each individual message is modeled with a logit link using a overall intercept μ_α and individual-level intercepts for each participant. This means that for each participant, multiple observations—multiple responses to messages—are included as a level in the model.

Table 3 Table showing the model comparisons used to select the analysis model

	<i>Df</i>	<i>AIC</i>	Log Lik.	χ^2	<i>p</i>
<i>Mod. A:</i> null model	2	10,903.18	−5,449.59		
<i>Mod. B:</i> <i>A</i> + time in trial	3	10,435.32	−5,214.66	469.86	<.001
<i>Mod. C:</i> <i>B</i> + number	4	10,428.11	−5,210.06	9.21	.002
<i>Mod. D:</i> <i>C</i> + ind. strategy rand. eff.	13	10,423.14	−5,198.57	22.97	.006
<i>Mod. E:</i> <i>D</i> + condition × number	16	10,420.51	−5,194.26	8.62	.034

From the null model, which models the probability of success only using a overall intercept and individual-level intercept, time in trial, the number of messages, and individual differences in response to persuasion principles are added (random effect). Finally, condition × number is added

From this null model, we proceed by fitting a series of multi-level models where we use χ^2 tests to examine the increase in model fit of each of the subsequent models. Adding a fixed effect of principle to this null model does not improve model fit, $\chi^2 = 4.75$, $Df = 3$, $p = .19$, neither does adding condition $\chi^2 = 2.433$, $Df = 3$, $p = .49$ ⁸. However, adding the Time in Trial does significantly increase model fit, see also comparisons in Table 3, Model *A* and *B*. This comparison shows that the success rate of the email reminders declines when users are using the activity promotion service for a longer period (see also the coefficients of the final model presented in Table 4). Adding a fixed effect of the number of messages received, Model *C*, again significantly improves model fit. Here, the interpretation is similar: the more reminders are sent, the less likely they are to be successful.

Next we add, in accordance with earlier findings on large individual difference in the effects of influence strategies (see [24]), random persuasive principle effects by participant. Thus, we allow for different intercepts for each principle for each of the individuals in our study, and we constrain the prior distribution over these principles to be distributed Normal with mean zero. Allowing for these individual differences in the effects of persuasion principles significantly improves model fit, which replicates the findings presented in Kaptein and Eckles [24] (see model *D*, Table 3).

Finally, to test the effects of the PMS, we add an interaction term of the number of messages send with the condition that participants were in. We setup the contrast such that the PMS system, the adaptive condition, is the reference category. This also significantly increases model fit showing that the effect of the messages differs between conditions when modeled over the number of messages that is send (see model *E*, Table 3). This latter interaction is justified since the adaptive system will take time to adapt, and hence, one would not expect a large difference

Table 4 Table showing the fixed effects of model *E* (see also Table 3)

	β	S.E.	<i>z</i>	<i>p</i>
Intercept	−0.855	0.066	−12.875	<.001
Time in trial	−0.016	0.002	−7.618	<.001
Number (adaptive)	−0.015	0.016	−1.037	.300
Random × number	−0.023	0.015	−1.518	.129
Control × number	−0.039	0.016	−2.389	.017
Best pretested × number	−0.044	0.016	−2.737	.006

Note that the decline of the both the adaptive condition (used as reference) and the random condition is not significantly different. The adaptive condition does outperform the control and the best-pretested conditions

in the effectiveness of the messages send out in the different conditions at the first messages. Only after a period of adaptation—which depends on the number of messages send—will the adaptive condition be able to distinguish itself from the other messaging conditions.

To be able to interpret the model and test whether indeed the adaptive condition is more successful than the other messaging conditions, Table 4 presents the estimated coefficients of Model *E*. The average effectiveness of the messages is rather low (Intercept = −.86), and the effectiveness of messages declines both as participants are longer in the service, $\beta_{\text{Time in Trial}} = -.16$ as well as when they receive more messages, $\beta_{\text{Number}} = -.15$. Inspection of the coefficients of the condition × number interactions allows for the estimation of the effects of the different conditions. Both the best-pretested condition and the control condition perform significantly worse than the adaptive condition (see Table 4). The random condition is also estimated to be less successful over time than the adaptive condition. This latter difference, however, is not statistically significant, $p = .13$.

Figure 2 shows that decline of the estimated success rates for each of the experimental conditions as the number of messages that is send per participant increases. The plot shows the estimated lines for the median of Time in Trial,

⁸ Both in this section and the subsequent section, the order of model building is determined by the largest model fit improvements of each subsequent term.

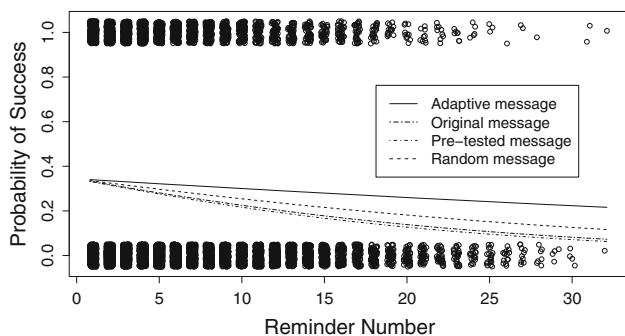


Fig. 2 Overview of the success rates of the reminder messages over the number of times a message was sent out to users for each of the different conditions (both jittered). To reduce complexity, the imposed lines are representative for the median Time in Trial, $\bar{X} = 28$

$\bar{X} = 28$. From the figure, it is clear that the decline in estimated effectiveness of the messages is the slowest in the adaptive condition.

4.2.2 Dropout rate

Besides examining the success rate of each message, we also explore the dropout rates in each condition. Participants are marked as a dropout when they have not been active for 30 days. While this signal is understandably noise due to the fact that some participants that are likely to drop out have not dropped out yet given the limited time-frame of the study, this metric is worth exploration. Dropout is one of the key metrics for the success of the activity promotion service, and thus, a reduction in dropout rates would not only be theoretically interesting but would also be of commercial importance.

To examine dropout rates, a logistic regression is fit on the level of individuals. For each individual in the study the condition, and the total number of reminder send, and the days in the program are examined. Again, a series of nested model comparisons using χ^2 difference tests is used to select the final model to examine the effects of the conditions. First, adding the number of reminders significantly increases model fit, $\chi^2 = 13.44$, $Df = 1$, $p < .001$. When accounting for the number of reminders, adding the days in the trial does not significantly improve model fit, $\chi^2 = 4.89$, $Df = 1$, $p = .53$, and this term is thus omitted. Finally, condition is added as a main effect. In this comparison, condition does not interact with the number of messages since only one row for each subject is included in the model comparisons. The addition of condition does not significantly improve model fit $\chi^2 = 5.35$, $Df = 3$, $p = .15$ (See also Table 5).

Given the importance of the dropout measure, we further inspect the estimates of model C_{drop} . Table 6 shows the coefficients. From the table, it is clear that the dropout

Table 5 Model comparisons for the logistic regression on dropout

	Df	χ^2	p
<i>Mod. A_{drop}</i> : Intercept only			
<i>Mod. B_{drop}</i> : $A_{drop} + \text{Number}$	1	13.44	<.001
<i>Mod. C_{drop}</i> : $B_{drop} + \text{Condition}$	3	5.35	.148

The number of reminders received has a significant effect on the dropout rate. The effect of the messaging condition is not statistically significant

Table 6 Fixed effects of model C_{drop}

	β	S.E.	z	p
(Intercept)	-1.0084	0.1952	5.17	<.001
Number	-0.0501	0.0143	3.49	<.001
Random	0.0938	0.2117	-0.44	.657
Control	0.3539	0.2083	-1.70	.089
Best pretested	0.3907	0.2043	-1.91	.056

The probability of dropout decreases as more reminder messages are send. Furthermore, the estimated probability of dropping out is slightly lower in the adaptive condition than in the other messaging conditions

rate decreases slightly with the number of reminders: People who receive more reminders are less likely to drop out. This is probably due to the fact that those who receive large number of reminders are committed to the program and just forget to upload. The reminders are then effective and make these users keep using the application. Those users that do not respond to the early docking reminders are likely to drop out altogether.

Looking at the effects of condition, it is clear the best-pretested condition and the random condition score worse than the adaptive condition (which is the reference category). However, these differences are only marginally significant. No clear difference between the random condition and the adaptive condition is found, although the estimated effect of the random condition is slightly higher—leading to more dropouts—than that for the adaptive condition. A graphical overview of the estimated effects is presented in Fig. 3.

4.3 Conclusions

This section presented the empirical evaluation of the persuasive messaging system for the docking reminder messages. In this adaptive persuasive system, users are identified by a unique one-way hashed identifier, which is an integral part of their usage of the persuasive service. After inactivity—failure to dock—for a period of 3 or 6 days, users received a reminder email. In this email, the authority, consensus, and scarcity strategy were implemented to increase compliance. The persuasion principles

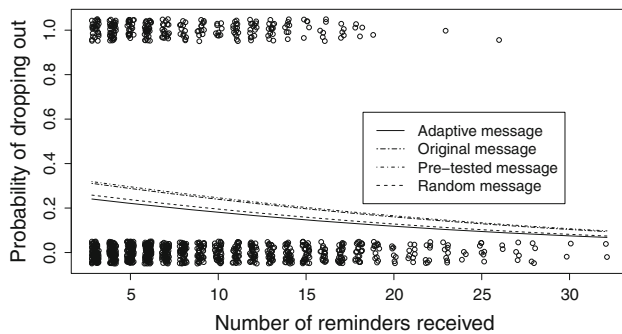


Fig. 3 Overview of the dropout rates of the reminder messages over the number of times a message was sent out to users for each of the different conditions (*both jittered*). The figure compares the estimated dropout rates over the number of messages for each condition

were added to the email messages in such a way that they were interchangeable and could thus be personalized. Finally, the effect of the messages was measured by combining logging of the opening of the email messages via a dynamic image in the content of the email and users logged docking behavior.

Results of the evaluation of the PMS system partly show the benefits of using persuasion profiles: The (repeated) success of the reminder messages is higher when using personalized persuasion than when using the default message (control condition) or the best-pretested message. The estimated success of the adaptive messages was also higher in the adaptive condition than in the random condition, but this latter difference is not statistically significant, $p = .13$.

A similar pattern was found when examining the number of dropouts. Despite the noise signal, the estimated effects of the adaptive condition are marginally better than that of the best-pretested condition, $p = .06$, and of the control condition, $p = .09$. Again, no clear difference was found between the random messaging condition and the adaptive condition even though the estimated dropout rate in the random condition was slightly higher than that in the adaptive condition.

4.4 Limitations

The evaluation presented above has several limitation. First of all, the constraints put up by the company that markets the activity promotion service led us to collect a data set that is of variable length for each of the participants in our evaluation. This makes that the number of reminders send is variable, as well as the number of days in the trial. These restrictions lead to noise estimates of the effects of the conditions and thus a lowered precision. Regrettably, this enables us to conclude only with certainty that the adaptive condition outperforms both the best-pretested as well as the control condition, but not necessarily the random

messaging condition. Studies in which an adaptive persuasive system runs for a longer time period, or at least with more observations, would allow for a more precise estimate of the difference that emerges over messages between the adaptive and the random condition. We believe that the consistently more successful estimates of the adaptive condition—both in increasing success rates as well as decreasing dropout—warrant such further investigation into the effects of adaptive persuasive messaging.

Another limitation—or point of caution—for the presented results is the general low effectiveness of the use of persuasion principles in the messages. This was already clear from the initial questionnaire evaluation but is even more eminent by the lack of a main effect of persuasion principle in the model comparisons. This could imply two things: One, the messages are not powerful enough to make a large difference, or two, the success rate and dropout rates of the emails are determined by several other factors making the estimated effects of the principles small. Both are likely in play. However, in favor of the idea that principle selection should be adapted to individuals, we did find a significant improvement in model fit when principle was added as a random effect over users. Hence, allowing for different effects for each of the individual principles increased model fit. Subsequently, the adaptive system outperformed the best-pretested principle, showing that at least a difference could be made by selecting the right principle for individual users over selecting the on average most effective principle.

5 General discussion

In this paper, we described the development and evaluation of the PMS system. The PMS system is an adaptive persuasive system that uses persuasion profiles to adapt to individual differences in response to persuasion principles. We detailed how, through identification, representation, and measurement, designers can create systems that attend to these individual differences. The empirical evaluation shows the benefits of the use of persuasion profiles: The decrease in effectiveness of the reminder emails is lower when individual differences in response to persuasion principles are adapted to, leading overall to a lower dropout rate of the service.

The system presented here should inspire designers of ubiquitous technologies to create adaptive persuasive technologies that adapt their usage of distinct influence strategies to the responses of users. While large individual differences in the responses to influence strategies have already been shown in experiments, the in-the-field evaluation of the PMS showed promising results for adaptive persuasion: Adaptive persuasion outperforms the two static

messaging conditions. The adaptive nature of the system, for which an implementation was detailed in this paper, however, needs further evaluation to convincingly demonstrate the applied benefits of adaptation over random message selection.

This article is, however, intended not only to demonstrate the effectiveness of adaptive persuasion in ubiquitous technologies, it is also meant to introduce the concept of persuasion profiling for designers of ubiquitous technologies. We believe that the ubiquitous technology scenario particularly satisfied the design requirements of identification, representation, and measurement as imposed by adaptive persuasive technologies: Ubiquitous applications can often track individuals, represent messages in multiple ways, and are able to measure the effects of such representations using sensors. Therefore, we think that persuasion profiles, and the exemplar use described in this paper, are valuable for designers of persuasive ubiquitous applications.

5.1 Future work

This article presents a new view on the study of persuasion strategies and their usage in persuasive technologies. The article builds on the idea that there are large individual differences in response to persuasion principles. We propose a class of technologies created to address these individual differences, *adaptive persuasive systems*, which inspire new questions about human behavior and decision making as a function of persuasion strategies. Questions, through the deployment of ubiquitous sensing technologies to measure user behavior, can hopefully be answered in the future. Adaptive persuasive technologies in ubiquitous applications can be a *tool* for further psychological research and should address the effects of persuasive strategies at an individual level and over time. By and large researchers of influence strategies, persuasion principles and persuasive technologies have till now focussed on average effects of the one term use of a persuasive intervention or manipulation. This article shows how individual-level responses over time can be incorporated in the design of adaptive persuasive applications.

The difference between the average-level effects and the individual-level effects of persuasion principles warrants future research: persuasive applications should deliver on their promise to change behavior of their users, not of *other users on average*. With its focus on a healthy lifestyle, ubiquitous persuasive technologies frequently are designed to influence *individuals*. These individuals use persuasive applications to change *their own* attitudes or behavior and as such this is what the systems should be designed for. The ubiquitous computing paradigm enables the unobtrusive

measurement of individual responses to persuasive principles and thus enables designers to build persuasive systems that are effective for each individual user.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Ajzen I, Fishbein M (1980) Understanding attitudes and predicting social behavior. Prentice-Hall, Englewood Cliffs, NJ
2. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59(4):390–412
3. Bosworth K, Gustafson DH, Hawkins RP, Chewing B, Day T (1983) Adolescents, health education, and computers: the body awareness resource network (BARN). *Health Education* 14(6): 58–60
4. Cacioppo JT, Petty RE, Feinstein JA, Jarvis WBG (1996) Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. *Psychol Bull* 119(2): 197–253
5. Cacioppo JT, Petty RE, Kao CF, Rodriguez R (1986) Central and peripheral routes to persuasion: an individual difference perspective. *J Pers Soc Psychol* 51(5):1032–1043
6. Caplan R (1990) How fingerprints came into use for personal identification. *J Am Acad Dermatol* 23(1):109–114
7. Cialdini R (2001) Influence, science and practice. Allyn & Bacon, Boston
8. Cialdini R (2004) The science of persuasion. *Sci Am Mind* 284: 76–84
9. Deutsch M, Gerard HB (1955) A study of normative and informational social influences upon individual judgment. *J Abnorm Soc Psychol* 51:629–636
10. Efron B, Morris C (1975) Data analysis using {Stein's} estimator and its generalizations. *J Am Stat Assoc* 70(350):311–319
11. Fogg BJ (2003) Persuasive technology: using computers to change what we think and do. Morgan Kaufmann, San Francisco, CA
12. Goldstein DG, Gigerenzer G (2002) Models of ecological rationality: the recognition heuristic. *Psychol Rev* 109(1):75–90
13. Haugtvedt CP, Liu K, Min KS (2008) Individual differences: tools for theory testing and understanding in consumer psychology research. In: Handbook of consumer psychology. CRC Press, New York, pp 1161–1176
14. Haugtvedt CP, Petty RE (1992) Personality and persuasion: need for cognition moderates the persistence and resistance of attitude changes. *J Pers Soc Psychol* 63(2):308–319
15. Haugtvedt CP, Petty RE, Cacioppo JT (1992) Need for cognition and advertising: understanding the role of personality variables in consumer behavior. *J Consumer Psychol* 1(3):239–260
16. Hazen TJ, Weinstein E, Park A (2003) Towards robust person recognition on handheld devices using face and speaker identification technologies. In: Proceedings of the 5th international conference on Multimodal interfaces ICMI 03, 289
17. IJsselstein W, de Kort Y, Midden C, Eggen B, van den Hoven E, Atkinson B (2006) Persuasive technology. Lecture notes in computer science Series, vol 3962. Springer, Berlin, Heidelberg
18. James JM, Bolstein R (1992) Effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys. *Public Opin Q* 54:442–453

19. Jarvis WBG, Petty RE (1996) The need to evaluate. *J Pers Soc Psychol* 70(1):172–194
20. Johnson BT, Eagly AH (1989) Effects of involvement on persuasion : a meta- analysis effects of involvement on persuasion: a meta-analysis. *Prevention*
21. Kaptein MC (2011a) Adaptive persuasive messages in an E-commerce setting: the use of persuasion profiles. In: *Proceedings of ECIS 2011*. Helsinki
22. Kaptein MC (2011b) Building persuasion profiles in the wild: using mobile devices as identifiers maurits kaptein (CHI 2011 PINC workshop)
23. Kaptein MC, Eckles D (2010) Selecting effective means to any end: futures and ethics of persuasion profiling. In: Ploug T, Hasle P, Oinas-Kukkonen H (eds) *Persuasive technology*, Springer, Berlin /Heidelberg, pp 82–93
24. Kaptein M, Eckles D (2012) Heterogeneity in the effects of online persuasion. *J Interact Market* 26(3):176–188. doi:10.1016/j.intmar.2012.02.002. <http://www.sciencedirect.com/science/article/pii/S1094996812000035>
25. Kaptein MC, Eckles D, Davis J (2011a) Envisioning persuasion profiles: challenges for public policy and ethical practice. *ACM Interact* 18(5):66–69
26. Kaptein MC, Duplinsky S, Markopoulos P (2011b) Means based adaptive persuasive systems. In: *Proceedings of the 2011 annual conference on Human factors in computing systems*. CHI '11. ACM, New York, NY, USA, pp 335–344
27. Kaptein M, De Ruyter B, Markopoulos P, Aarts E (2012) Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Trans Interact Intell Syst* 2(2):10:1–10:25. doi:10.1145/2209310.2209313. <http://doi.acm.org/10.1145/2209310.2209313>
28. Kellermann K, Cole T (1994) Classifying compliance gaining messages: taxonomic disorder and strategic confusion. *Commun Theory* 4(1):3–60
29. Kostakos V (2008) Using Bluetooth to capture passenger trips on public transport buses. *CoRR abs/0806.0*
30. Macready WG, Wolpert DH (1998) Bandit problems and the exploration/exploitation tradeoff. *IEEE Trans Evol Comput* 2(1):2–22
31. Maio GR, Esses VM (2001) The need for affect: individual differences in the motivation to approach or avoid emotions. *J Pers* 69(4):583–615
32. Milgram S (1974) *Obedience to authority*. Tavistock, London
33. Petty RE, Cacioppo JT (1986) The elaboration likelihood model of persuasion. *Adv Exp Soc Psychol* 19(1):123–205
34. Petty RE, Wegener DT (1999) The elaboration likelihood model: current status and controversies. In: Chaiken S, Trope Y (eds) *Dual-process theories in social psychology*, Guilford Press, New York, p 4172
35. Rhoads K (2007) How many influence, persuasion, compliance tactics and strategies are there? <http://www.workingpsychology.com/numbertactics.html>
36. Sakai R, van Peteghem S, van de Sande L, Banach P, Kaptein MC (2011) Personalized persuasion in Ambient intelligence: the APStairs system. In: *Proceedings of Ambient intelligence (AmI) 2011*. Amsterdam
37. Scott SL (2010) A modern Bayesian look at the multi-armed bandit. *Appl Stoch Models Bus Ind* 26(6):639–658
38. Stein C (1955) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proceedings of the 3rd Berkeley symposium on mathematical statistics and probability*. University of California Press, Berkeley, pp 197–206
39. Tormala ZL, DeSensi VL (2009) The effects of minority/majority source status on attitude certainty: a matching perspective. *Pers Soc Psychol Bull* 35(1):114–125
40. Verhallen TMM, Robben HSJ (1994) Scarcity and preference: an experiment on unavailability and product evaluation. *J Econ Psychol* 15(2):315–331
41. Webster DM, Kruglanski AW (1994) Individual differences in need for cognitive closure. *J Pers Soc Psychol* 67(6): 1049–1062
42. West SG (1975) Increasing the attractiveness of college cafeteria food: a reactance theory perspective. *J Appl Psychol* 60: 656–658
43. Wilcox RR (1981) A Review of the beta-binomial model and its extensions. *J Educational Stat* 6:3–32