SAGE Open
October-December 2013: 1–16

DOI: 10.1177/2158244013517245
sgo.sagepub.com
SAGE

A Methodology for Quantifying Effects and Psychological Functioning of Behavior-Change Techniques

Robert Tobias[1], Alexandra C. Huber[1], and Andrea Tamas[1]

Abstract
We present a methodology to quantify the effects of behavior change techniques (BCTs) that allows forecasting campaign effects on behavior and psychological constructs. The approach involves the gathering of longitudinal data during actual campaigns in which different combinations and sequences of BCTs are applied to different groups. Approximate metric data are gathered by asking for simple and specific evaluations. The data are analyzed using regression models that consider the value range of the dependent variable as bounded (bounded linear regression). Based on these models, forecasts of the intervention effects are calculated, considering the uncertainty of the parameter estimates. The methodology is applied to investigate the effects of prompts (external memory aids), public self-commitments, and implementation intentions on affective and instrumental attitudes, injunctive and descriptive norms, forgetting, perceived behavior control, and behavior in a health-promotion campaign in Bolivia. Prompts and public self-commitments reached more than half of the target population but only showed relevant effects when combined or repeated. The effects of both BCTs on behavior were mainly mediated by forgetting. Implementation intentions were not well received by the promoters and the population. From the few cases that implemented this BCT, no clear psychological effects could be derived.

Keywords
quantifying effects of behavior change techniques, forecasting campaign effects, health-promotion interventions, longitudinal bounded linear regression, psychology

Introduction

Many of the most urgent problems humanity faces could be mitigated at least partly by changing individual behavior. For example, the health conditions of hundreds of millions of people in developing countries could be improved by making people disinfect their drinking water and wash their hands with soap (Prüss-Üstün, Bos, Gore, & Bartram, 2008). Psychology could play a crucial role in solving problems related to individual behaviors by providing behavior change theories and evidence of the impact of specific behavior change techniques (BCTs) on psychological constructs and behavior. Based on such information, large-scale behavior change campaigns could be planned, applying the most efficient combination of BCTs for the given conditions. Unfortunately, most psychological studies fail to provide the information relevant for planning behavior change campaigns, namely, how much certain BCTs change different psychological constructs and how much the behavior changes due to the changes in the psychological constructs.

Many studies do not investigate real-world intervention campaigns, but are more or less artificially set up as laboratory or field experiments (e.g., Sheeran et al., 2005; Webb, Ononaiye, Sheeran, Reidy, & Lavda, 2010). While such studies produce high-quality data on isolated effects of BCTs, to the study participants, the situational context of the study is often unfamiliar, the induced behaviors are of low relevance, and the social effects normally induced by large-scale campaigns are absent. A second type of study investigates real-world behaviors and settings but only compares a group in which a single intervention is undertaken to a control group without any degree of intervention (e.g., Cox, Cox, & Cox, 2005; Hill, Abraham, & Wright, 2007). Studies that only compare intervention and control groups for statistically significant behavioral differences are of limited value, as it is not very surprising that doing something has a greater effect than doing nothing. What is of genuine interest is the size of intervention effects when comparing different single BCTs, as well as their combinations and repetitions. A third type of study seems to overcome the issue and uses complex

[1]Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

Corresponding Author:
Robert Tobias, Department Environmental Social Sciences, Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, P.O. Box 611, 8600 Dübendorf, Switzerland.
Email: robert.tobias@eawag.ch

intervention campaigns, comprising many different BCTs. Unfortunately, often only the overall effect of the entire campaign is investigated (e.g., Cairncross, Shordt, Zacharia, & Govindan, 2005), without clarification of how each technique works.

The type of data that many intervention studies collect constitutes a fourth shortcoming. Often, only behavioral outcomes are measured, with no information provided regarding psychological constructs (e.g., Cairncross et al., 2005; Mosler, Kraemer, & Johnston, 2013). To understand how BCTs affect behavior (and which psychological factors determine behavior), data on psychological constructs are needed. Finally, many studies limit their analysis to linear covariance structures of cross-sectional data (e.g., Hill et al., 2007; Kraemer & Mosler, 2010). However, to investigate intervention effects and how determinants of behavior are changed, longitudinal data that represent changes over time in absolute terms are required.

Success (or lack thereof) in delivering the BCTs to the study participants is another aspect that is usually ignored during studies on intervention effects. In many studies, the researchers or their assistants directly intervened in the investigated cases. Thus, a 100% successful delivery is guaranteed, but no information about the success of delivering the BCT in a large-scale campaign is obtained. Investigations of actual large-scale campaigns mostly assume that the BCTs reached every person in the intervention group. However, this assumption is usually wrong and the delivery is only partially successful, for many reasons. Most interesting are the factors that lead to different delivery success rates for different BCTs, such as their attractiveness to the target persons (affecting refusal rate and the motivation of the promoters) or the effort and time required to apply the BCT correctly.

To develop theories and models of behavior change that support planning, guiding, and evaluating behavior change campaigns such as health promotion interventions, the effects of combined techniques on behavior and mediating psychological constructs under real-world conditions need to be quantified. Or, as stated by Michie, Rothman, and Sheeran (2007), "We need to move beyond assuming the theory indicates how to change behavior to studying behavior change techniques in their own right" (p. 252). The present study takes one step in this direction by presenting an exemplary procedure regarding how the effects of BCTs can be quantified and applying this method to the investigation of prompts, public self-commitments, and implementation intentions. Furthermore, the study investigates the success rate of delivering these different BCTs. While the main goal is to present a methodology that allows quantifying the effects of BCTs, a number of substantial results are derived regarding the effectiveness of the mentioned BCTs to change behavior and the psychological mechanisms that lead to these changes. Before explaining the methodology, we present the BCTs used in this study.

The Investigated Behavior Change Techniques

Prompts (reminders or external memory aids; Intons-Peterson & Fournier, 1986) are objects indicating that a certain behavior should be performed. Implementation intentions (Gollwitzer, 1999) are simple plans that associate a certain behavior with a specific situation. Public self-commitment (Kiesler, 1971) is a more or less formal statement of the intent to perform a behavior that is made visible to other people. All three techniques are similar in their focus on the implementation of an intended behavior. These techniques do not change infrastructure to enable behaviors or persuade persons to increase their intention to perform a behavior. Rather, the techniques facilitate performance at the right moment of an already intended behavior in its correct form. All three techniques are widely used, and their effects are often investigated (e.g., Armitage, 2007; Guynn, McDaniel, & Einstein, 1998). However, research almost exclusively focuses on their effects on behavior. Many ideas on psychological modes of action are discussed, but empirical findings to back up these ideas are scarce.

The mode of action of prompts and implementation intentions seems to be very similar (Guynn et al., 1998, for reminders; Gollwitzer & Brandstätter, 1997, for implementation intentions). (1) Planning processes are activated by setting up a prompt or implementation intention, which should facilitate implementation of an intended behavior. (2) The association between a situational cue and the intended behavior reminds a person to perform the behavior at the right moment. A prompt is an object explicitly set up to be associated with the behavior; implementation intentions use naturally occurring cues to link to the target behavior. (3) Memory aids (Shapiro & Krishnan, 1999) and implementation intentions (Chasteen, Park, & Schwarz, 2001) remind a person how to perform the behavior. A dynamic model presented by Tobias (2009) on the effects of memory aids found that the more committed a person is to performing a behavior, the more effective are memory aids in reminding that person to perform the behavior. This, in turn, leads to stronger habits, which, when developed, prevent forgetting the behavior. It can be expected that prompts will be particularly effective if combined with interventions that increase commitment. Examples of such BCTs include implementation intentions and public self-commitments.

While the effectiveness of self-commitment on behavior has been well investigated (e.g., Burn & Oskamp, 1986), only speculation exists regarding their mode of action. In general, authors refer to tension states (e.g., dissonance, Festinger, 1957) that emerge if the behavior and/or attitudes are not in line with self-commitment, causing them to be adjusted accordingly. Some authors (e.g., Cialdini, 2001) assume that self-commitment and implementation intention work similarly, with the latter specifying the situational context while the former specifying the behavior. In the case of public self-commitment, further normative effects are expected, due to increased social pressure to conform to the publicly displayed self-commitment and increased visibility of the descriptive norm (e.g., Nyer & Dellande, 2010).

To summarize, the most probable mode of action of the three BCTs seems to be the facilitation of *remembering*. Furthermore, effects on norms and attitudes have to be considered. In the present study, norms are differentiated into *injunctive* and *descriptive* norms (Cialdini, Reno, & Kallgren, 1990). While *injunctive norms* describe perceptions of what ought to be done, *descriptive norms* express people's perceptions of what is done in the social environment. In a similar fashion, we differentiate attitudes into *affective* and *instrumental* attitudes (e.g., Trafimow & Sheeran, 1998). *Affective attitude* refers to how pleasant it is to perform a certain behavior, while *instrumental attitude* refers to the advantages and disadvantages of performing the behavior. Behavioral control also needs to be taken into account (Ajzen, 1991), as this can be an important constraint to behavior performance (for example, in cases wherein people do not have enough time or resources to perform the behavior).

The Present Study

This study presents a methodology that quantifies BCT effects in a way that allows forecasting the psychological and behavioral effects of intervention campaigns. The approach is applied to investigate the effects of prompts, public self-commitments, and implementation intentions, using longitudinal data from a large-scale campaign promoting solar water disinfection (SODIS) in rural Bolivia. The effects on behavior and the psychological constructs of each BCT alone and of some parallel and sequential combinations, as well as success in delivering the BCTs, are quantified.

From an applied research perspective, the principal question is, "How much behavior change can be achieved with certain BCTs or their combinations?" Thus, the study investigates two applied research questions (RQ):

> **RQ1:** How successful was the delivery of the different BCTs to the target population?
> **RQ2:** What are the effects of different BCTs and their combinations on behavior?

From a theoretical perspective, we want to know: "What are the psychological processes that mediate the effects of the BCTs on behavior?" Thus, two different questions are investigated:

> **RQ3:** What are the effects of the different BCTs and their combinations on the investigated psychological constructs?
> **RQ4:** How do the BCTs work; what are the psychological mechanisms of these techniques?

Finally, a fifth question is addressed:

> **RQ5:** How can the changes in behavior in the control group be explained?

This question stems from the fact that the results show a considerable change in behavior, even for households without intervention.

Method

Participants, Procedure, and Interventions

Data were gathered in the northern highlands of Chuquisaca (Bolivia) during a campaign aimed at reducing the high infant mortality rate caused by diarrhea by promoting SODIS and hand washing. SODIS is undertaken by filling transparent plastic bottles with water and exposing them to the sun for 6 hr (or 2 consecutive days if cloudiness exceeds 50%). Sunlight inactivates pathogenic microorganisms due to the radiation in the UV-A spectrum. SODIS significantly reduces levels of bacterial contamination in the laboratory (e.g., Berney, Weilenmann, Simonetti, & Egli, 2006) and under field conditions (e.g., Sommer et al., 1997). A brief overview of microbiological, medical, and psychological research on SODIS is given by McGuigan et al. (2012). Although the application of SODIS is simple, the adoption rate has been rather slow (Tamas & Mosler, 2011). Therefore, a campaign was initiated by the local non-governmental organization (NGO) Fundación SODIS and implemented in collaboration with the Ministry of Health of Chuquisaca and the Department of Health Service (SEDES). These organizations had the required permits, and the study was carried out in accordance with universal ethical principles.[1]

The project started in June 2007, with a baseline evaluation followed by three longitudinal panels in August 2007, November 2007, and March 2008. Intervention waves were placed between these panels, and radio spots promoting SODIS were on the air for the entire campaign. Because the population was made aware of the SODIS promotion between the baseline and the first panel, changes during this period might be overstated. To avoid bias in the estimates of the intervention effects due to social desirability, the changes from the baseline to the first panel were not used in the analyses presented. The BCTs were distributed during information events before the campaign by local health volunteers who were trained in applying these techniques. The BCTs applied are compiled in Table 1.

Due to the high illiteracy rate, surveys were conducted via face-to-face interviews; written consent could not be obtained from survey participants. However, the persons contacted by the interviewers had been clearly informed about the study and that participation was completely voluntary. Eight students from Sucre were recruited and trained to conduct the interviews in an interviewer workshop. Nine villages were selected, mainly based on their accessibility. Due to small community sizes and low density of households within the

Table 1. Behavior-Change Techniques Used in the Campaign Investigated: Names, Descriptions, and Criteria for a Case Being Considered as Having Received the Intervention.

Name	Description	Criterion
Prompt (reminder, external memory aid)	Cuboids made of cardboard of about 15 cm × 15 cm × 30 cm in size. One side prompted to do SODIS ("Put the bottles into the sun"), one side presented the steps of doing SODIS, one side prompted hand washing, and one side had a current calendar. The prompts were printed in color and could be placed on furniture or hung from the ceiling. The health volunteers gave the prompt with an instruction to place it near where water was usually handled.	Prompt well visible when visited by interviewer
Public self-commitment	A plasticized A4-sized poster stating in Spanish "We are committed to drink water treated by the sun," a SODIS-logo and a picture of a promoter shaking hands with a Bolivian woman holding a SODIS bottle in her other hand. The health volunteers asked how many SODIS bottles the household needed to treat all drinking water. The subjects then committed themselves by stating in Spanish: "I will prepare ___ bottles of SODIS water every day." The "contract" was sealed with a handshake. The public self-commitment poster was set up above the outside door of the house.	Poster well visible when visited by interviewer
Implementation intention	A paper sheet, A4 size, containing the sentence in Spanish "Every day after _____ (e.g., getting up, breakfast) I will prepare the SODIS-bottles and put them _____ (e.g., on the roof) where they are lying in the sun the whole day," a SODIS logo, and two pictures, one showing bottle filling; the other, bottles in the sun. Promoters discussed the best time and place for doing SODIS and filled out the sentence on the paper accordingly. The subjects were asked to form the implementation intention by pronouncing the completed phrase.	The household was able to produce the paper sheet; asking

Note. SODIS = solar water disinfection.

villages, random sampling was not feasible, and the interviewers were instructed to interview every possible household. Figure 1 shows the flowchart of study participants and the rather complex design of the study. Not all households that were interviewed were reached by the health volunteers who distributed the BCTs. Households without interventions serve as the control group, ensuring maximal similarity between the control and intervention groups.

Measures

Based on previous research on SODIS promotion (e.g., Heri & Mosler, 2008), a standardized questionnaire was developed, translated by local experts, discussed for identical understanding of items with the interviewers, and pre-tested. Table 2 compiles the items for the constructs analyzed here. For the behavior measure, the interviewees were asked to estimate the quantity of water (in cups) consumed by the household on an average day. Then they estimated the number of cups that are boiled, the number treated with SODIS, and the number consumed raw. Based on this information, the percentage of each water type was calculated and used in the analyses. Previous studies indicate that this measure is a good indicator for observed SODIS behavior and, in particular, that it shows very similar changes due to interventions as indicators based on observation (Mosler et al., 2013).

Each item is interpreted as an approximate metric measure that quantifies the evaluation of one specific aspect of the investigated behaviors. These single evaluations are then aggregated into more abstract constructs by computing scores using specific formulas (see Table 2). Following the taxonomy of Law, Wong, and Mobley (1998), the constructs used here are *aggregate* constructs and not latent constructs. Latent constructs could not be used because for modeling *individual* processes, data must be gathered that can be interpreted for each individual case (i.e., in absolute terms). In contrast, latent constructs can only be interpreted in comparison to other cases (Borsboom, Mellenbergh, & van Heerden, 2003). A further discussion of this approach is presented in the Supporting Information (SI) on page S-2.

The use of aggregate instead of latent constructs allows the estimation of absolute effects. However, a consequence of this approach is that we cannot assume a priori that the items of a construct are correlated. For example, a person who thinks SODIS-treated water is good for health does not necessarily also thinks treating water with SODIS is cheap. Therefore, estimating the reliability of the measures based on internal consistency (e.g., Cronbach's alpha) is not possible. Because we do not know how reliable our data are, we explicitly consider the uncertainty in the forecasts of the intervention effects. The uncertainties can be estimated due to the longitudinal design, but they comprise not only the

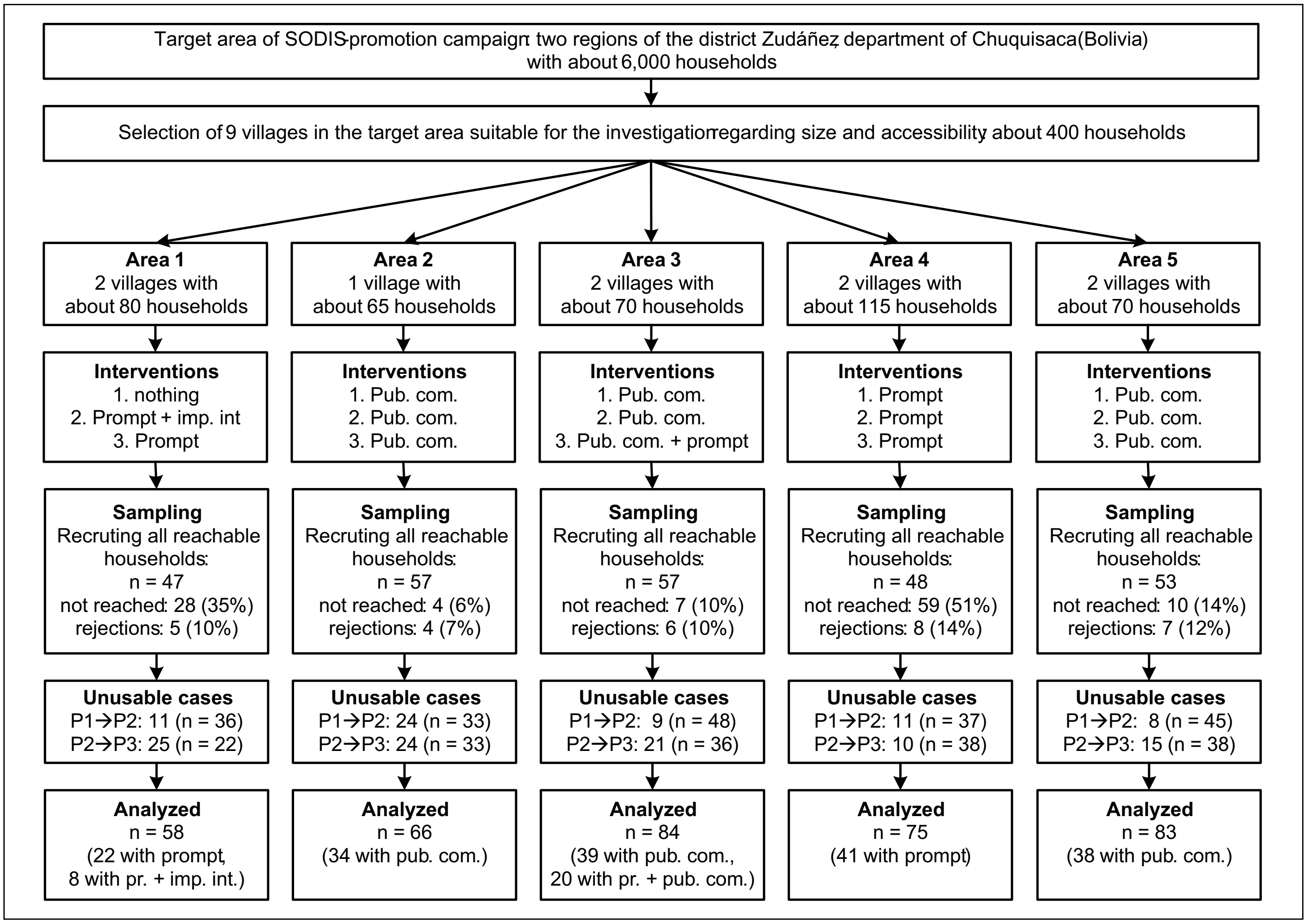

Figure 1. Flowchart of study participants.
Note. Interventions were applied in three subsequent waves (1, 2, and 3). Pub. com. = public self-commitment; pr. = prompt; imp. int. = implementation intention; P = panel. Unusable cases are cases excluded from the analysis due to missing values or combinations of behavior-change techniques too rare to be investigated on their own.

effects of unreliable measurements but also other influences, such as instability of the constructs over time, variations in intervention effects among subjects, or shortcomings of the model. To conclude, we cannot quantify the reliability of our measures, but the fit indicators of the models quantify the effects of all random influences that might impair a forecast and, thus, the interpretation of the data.

Table 2 compiles the formulas for computing the scores of the constructs. These formulas are linear combinations of items in which evaluations in favor of SODIS have a positive value and evaluations in favor of raw water have a negative value. Furthermore, the scores are scaled to the range of [−1, +1] and [−1, 0] respectively. Data were gathered regarding behavior, whether the individual remembered to practice the SODIS technique, affective and instrumental attitudes, and injunctive or descriptive norms. Evaluations of the target behavior (consuming SODIS water) and the competing behavior (consuming raw water) were performed. For attitudes and the injunctive norm, the two evaluations were considered as separate constructs. In the case of behavior and the descriptive norm, the two evaluations were aggregated to one construct each, as they depend on each other: If Behavior A is performed more often, Behavior B has to be performed less often; the extent to which individuals consume raw water qualifies the number of times they consume SODIS-treated water. For remembering and behavior control, an evaluation regarding raw water consumption does not make sense. In the case of behavior control, an alternative for disinfecting water had to be considered: boiling the water. With the exception of the availability of fuel for boiling the water, no evaluation of consuming boiled water showed any effect on the consumption of raw or SODIS water. The evaluation of the behavior of boiling water before consumption is included in the construct of behavior control for SODIS, as the behavior directly reduces the need for SODIS. The item on the difficulty of performing SODIS correctly was entered as an instrumental attitude and not as a behavior control, as the question is not about being able to perform SODIS but

Table 2. Psychological Constructs Used in the Analyses: Names, Questions, and Answering Options of Items, and the Formulas for Calculating the Scores of the Constructs.

Construct	Questions	Answering options	Formula
Behavior	1. How much water do you disinfect with SODIS? 2. How much raw water do you and your family drink?	5: nothing (0) . . . all (1) of what we need	Item 1 – Item 2
Remembering	1. Do you always remember to do SODIS?	5: never (−1) . . . always (0)	Item 1
Behavior control	1. Do you have enough firewood to boil your water? 2. Do you think that preparing your water with SODIS costs a lot or not much time? 3. Do you have enough bottles to prepare SODIS?	5: never (0) . . . always (1) 5: costs no (0) . . . a lot of (1) time 5: never (0) . . . always (1)	(Item 1 – Item 2 + Item 3 – 0.5) / 1.5
Affective attitude RW	1. How good or bad do you think is drinking raw water? 2. What do you think about the taste of raw water? 3. Do you like to drink raw water; how pleasant is it?	9: very bad / unpleasant (−1) . . . very good / pleasant (1)	(Item 1 + Item 2 + Item 3) / 3
Instrumental attitude RW	1. Do you think drinking raw water is good or bad for your health?	9: very bad (−1) . . . very good (1)	Item 1
Affective attitude SW	1. How good or bad do you think is using SODIS? 2. What do you think about the taste of SODIS water? 3. Do you like to do SODIS; how pleasant is it?	9: very bad / unpleasant (−1) . . . very good / pleasant (1)	(Item 1 + Item 2 + Item 3) / 3
Instrumental attitude SW	1. Do you think drinking SODIS water is good or bad for your health? 2. Do you think SODIS costs a lot of money? 3. Do you think it is difficult to do SODIS correctly?	9: very bad (−1) . . . very good (1) 5: it costs no (0) . . . much (1) money 5: not at all (0) . . . very (1) difficult	(Item 1 – Item 2 – Item 3 + 1) / 2
Injunctive norm RW	1. How do other people, who are close to you, think of you when you are consuming raw water?	9: very bad (−1) . . . very good (1)	Item 1
Injunctive norm SW	1. How do other people, who are close to you, think of you when you prepare your water with SODIS?	9: very bad (−1) . . . very good (1)	Item 1
Descriptive norm	1. Do you think or know that other people (friends, neighbors) prepare their water with SODIS? 2. Do you think or know that other people (friends, neighbors) drink raw water?	5: nobody (0) . . . everybody (1)	Item 1 – Item 2

Note. All constructs have a range of [−1, +1] with exception of the construct for remembering [−1, 0]. SODIS = solar water disinfection; RW = raw water; SW = SODIS treated water.

about confidence that the positive effects of SODIS can actually be achieved.

Analyses

Because the methods of this investigation are not widely used in psychology, they are explained in detail in the SI on pages S-3 to S-6. Most importantly, the models consider that the dependent variables are bounded (in [−1, +1] and [−1, 0] respectively), and therefore, not all theoretically possible changes can actually be observed. For example, if a person reports a behavior of 0.75 in a previous panel, the maximal observable intervention effect is 0.25, even if the intervention can have an effect of up to 1.0 in other individuals. As this model is linear between these bounds, it is called the *bounded linear* model.

The data from the three panels were combined into one set of differences by subtracting the values (for all investigated cases) of Panel 1 from the Panel 2 values, and the Panel 2 values from the Panel 3 values. Therefore, most cases are used twice in the analyses; this is permissible because dependencies are neutralized through the use of differences. From a modeling perspective, this means that the effects under investigation are assumed to be time-independent. However, in regression models, two constants are used (one for each time step), as unexplained changes might differ over time. In addition to the difference values, dummy variables for the interventions are used (0 = no intervention, 1 = intervention). Interactions of interventions must also be considered, specifically (1) if two interventions were applied together at the same time and (2) if interventions were applied in the previous time step (for all interactions, see Table 3).

A regression model for change in behavior on intervention variables is used to answer RQ2; regression models for changes in each psychological construct on intervention variables are calculated to answer RQ3; and a regression model for change in behavior on changes of all psychological constructs is estimated, with the exception of perceived behavioral control, to answer RQ4. The latter construct could not be considered in this model because of too many missing values. RQ5, which attempts to explain the behavior change in the control group, is investigated with a bounded linear regression model of behavior on the psychological constructs prior to the interventions. For all regression analyses,

Table 3. Descriptive Statistics of Psychological Constructs Used in the Analyses: Mean, Standard Deviation, and Number of Cases of the Values Before Change and the Changes Panel 2 – Panel 1 and Panel 3 – Panel 2.

	Before change		Change		
Psychological constructs	*M*	*SD*	*M*	*SD*	*n*
Behavior	−0.270	0.544	0.403	0.658	357
Remembering	−0.346	0.273	0.084	0.302	219
Behavioral control	0.257	0.315	0.018	0.391	160
Affective attitude (raw water)	0.291	0.411	−0.210	0.587	362
Affective attitude (SODIS water)	0.509	0.256	0.123	0.291	218
Instrumental attitude (raw water)	−0.167	0.526	−0.129	0.696	363
Instrumental attitude (SODIS water)	0.752	0.208	0.055	0.232	217
Injunctive norm (raw water)	0.038	0.354	−0.061	0.549	365
Injunctive norm (SODIS water)	0.388	0.340	0.036	0.457	218
Descriptive norm	−0.349	0.396	0.196	0.463	213
Dummy variables for interventions			Sum	*n*	DS
Current prompt			91	169	54%
Current public self-commitment			131	233	56%
Current prompt and implementation intention			8	36	22%
Interaction current prompt and public self-commitment			20		
Previous prompt			35		
Previous public self-commitment			58		
Interaction previous and current prompt			22		
Interaction previous and current public self-commitment			48		

Note. For the intervention dummy variables, number of cases that received the respective intervention (sum), number of cases that should have received the intervention (*n*), and delivery success (DS = sum / *n*) are presented. SODIS = solar water disinfection.

confidence intervals were estimated using a bootstrap approach with 1,000 samples. In addition, data was tested for multicollinearity of independent variables, heteroscedasticity, and autocorrelation. Outliers with residuals greater than two standard deviations were eliminated (on average, about 10% of all the cases). Pre-conditions for estimating regression models are met for all the presented results. However, not all parameters could be estimated for some models. The fit of the models to the data is quantified with the following indicators: explained variance (R^2 and adjusted R^2); standard error of the estimate (s of e); mean absolute error (MAE); root mean squared error (RMSE); percentage of cases with absolute residuals smaller than the measurement resolution of 0.25 ($|e| \leq 0.25$; interpreted as adequate forecasts); and the percentage of cases with absolute residuals larger than twice the measurement resolution ($|e| > 0.5$; interpreted as unusable forecasts).

To answer the research questions, forecasts of the intervention effects were calculated based on the parameters estimated in the regression models, and the effect sizes were classified based on the following considerations. The maximal measurable effect is 2 (i.e., a change from −1 to +1; in the case of remembering, the maximal measurable effect is 1). An intervention within an expensive large-scale campaign should lead at least to an effect of one step on a questionnaire scale (i.e., 0.25). Furthermore, because other factors increase the uncertainty of intervention effects in the real world, the factors that can be calculated should be of high probability; the 0.25 minimal effect should be reached with a probability of 97.5%. Intervention effects are evaluated based on the lower or upper limit of the 95% confidence interval, depending on whether the effect is positive or negative. If both limits have the same sign, the weaker effect has to be at least 0.125 to be mentioned. Effects between 0.25 and 0.5 are labeled as small; between 0.5 and 0.75, medium; between 0.75 and 1.0, strong; and greater than 1.0, very strong.

Results

Descriptive and Regression Results

The descriptive statistics of the variables used are compiled in Table 3. The lower part of Table 3 presents intervention counts. Cases with implementation intentions are sparse, and results for this BCT have to be interpreted with caution. The success rate of delivering the different BCTs was limited. Even the most successfully delivered BCTs—prompts and public self-commitments—reached only a little more than half of the targeted population (54% and 56%, respectively). Implementation intentions reached only 22% of the targeted population.

Table 4. Prediction of Intervention-Effects (Changes of Behavior and Psychological Constructs) Adjusted for Control Cases Without Intervention.

Scenario			Behaviour	Remembering	Behavioural control	Affective attitude		Instrumental attitude		Injunctive norm		Descriptive norm
						RW	SODIS	RW	SODIS	RW	SODIS	
Scenarios for first time-step	No intervention (comparison basis)	UL	0.329	0.094	0.328	0.023	0.093	−0.019	0.073	0.065	−0.040	0.465
		M	0.237	0.001	0.198	−0.046	0.016	−0.123	0.038	−0.014	−0.165	0.356
		LL	0.146†	−0.097	0.072	−0.119	−0.056	−0.226	0.002	−0.092	−0.285	0.242†
	Prompt	UL	0.578	0.717	0.168	−0.021	0.343	−0.006	0.744	0.091	0.466	0.538
		M	0.395	0.601	−0.025	−0.170	0.166	−0.264	0.664	−0.044	0.208	0.325
		LL	0.211†	0.494*	−0.200	−0.316	−0.002	−0.544	0.261*	−0.184	−0.031	0.112
	Public self-commitment	UL	0.399	0.987	0.013	0.039	0.290	−0.047	0.057	0.056	0.572	0.176
		M	0.249	0.808	−0.114	−0.066	0.171	−0.233	0.010	−0.047	0.402	0.031
		LL	0.088	0.567**	−0.256	−0.172	0.022	−0.427	−0.039	−0.154	0.230†	−0.124
	Prompt with public self-commitment	UL	1.293	1.031	0.519	−0.338*	0.193	−0.728**	N/A	−0.300*	0.246	0.511
		M	0.986	0.823	0.310	−0.579	0.021	−1.041	N/A	−0.570	0.018	0.252
		LL	0.638**	0.564**	0.083	−0.841	−0.130	−1.336	N/A	−0.813	−0.213	0.002
	Prompt with implementation intention	UL	1.524	0.301	−0.093	0.126	0.134	0.072	N/A	0.022	0.266	0.464
		M	0.883	0.121	−0.535	−0.034	0.001	−0.297	N/A	−0.208	−0.204	0.038
		LL	0.489*	−0.084	−1.023	−0.234	−0.127	−0.881	N/A	−0.468	−0.556	−0.429
Scenarios for second time-step	No intervention (comparison basis)	UL	0.705	0.113	0.467	−0.110	0.257	−0.372*	0.147	0.142	0.026	0.478
		M	0.525	−0.030	0.259	−0.247	0.148	−0.589	0.086	−0.016	−0.178	0.289
		LL	0.317*	−0.163	0.060	−0.389	0.035	−0.845	0.025	−0.176	−0.387	0.087
	Nothing after prompt	UL	0.277	0.863	0.040	0.574	0.314	0.652	0.708	0.556	0.931	0.647
		M	−0.121	0.678	−0.274	0.208	0.092	0.124	0.557	0.279	0.562	0.322
		LL	−0.523	0.474*	−0.618	−0.081	−0.113	−0.390	0.144	0.017	0.178†	−0.022
	Nothing after public self-commitment	UL	0.295	1.114	−0.190†	0.122	0.173	0.610	0.057	0.649	0.559	0.454
		M	−0.104	0.880	−0.449	−0.137	0.023	0.162	0.010	0.331	0.232	0.069
		LL	−0.558	0.588**	−0.725	−0.421	−0.136	−0.381	−0.039	0.033	−0.110	−0.279
	Repeated prompt	UL	1.355	1.503	0.413	−0.131†	0.366	−0.071	1.765	0.217	0.640	0.895
		M	0.981	1.170	0.097	−0.416	0.141	−0.734	1.626	−0.071	0.193	0.553
		LL	0.597**	0.777***	−0.227	−0.696	−0.058	−1.438	1.216***	−0.349	−0.232	0.228†
	Repeated public self-commitment	UL	1.397	1.162	−0.170†	0.091	0.405	0.537	0.092	0.252	0.925	0.458
		M	1.081	0.919	−0.408	−0.169	0.217	0.151	0.004	0.014	0.576	0.184
		LL	0.686**	0.618**	−0.645	−0.424	0.022	−0.259	−0.082	−0.217	0.290*	−0.087
	Prompt with public self-commitment after public self-commitment	UL	2.316	1.170	0.262	−0.422*	0.242	−0.257*	N/A	−0.227†	0.514	0.698
		M	1.818	0.933	0.016	−0.681	0.067	−0.658	N/A	−0.508	0.193	0.405
		LL	1.362***	0.641**	−0.249	−0.943	−0.111	−0.976	N/A	−0.784	−0.136	0.140†
Fit	R^2		61%	59%	29%	38%	28%	34%	63%	12%	16%	39%
	Adjusted R^2		60%	57%	23%	36%	24%	32%	61%	9%	12%	36%
	MAE		0.347	0.144	0.220	0.300	0.171	0.371	0.094	0.297	0.319	0.312
	RMSE		0.409	0.178	0.269	0.361	0.202	0.453	0.108	0.408	0.401	0.369
	$\lvert e \rvert \leq 0.25$		46%	94%	62%	47%	78%	48%	100%	53%	52%	44%
	$\lvert e \rvert > 0.50$		26%	0%	1%	18%	0%	26%	0%	28%	28%	20%
	n		305	193	142	304	186	313	153	332	211	200

Note. Effects that with p = 2.5 % are stronger than ±0.125 are marked with †, ±0.25 with *, ± 0.5 with **, and ± 0.75 with ***. RW = raw water; SODIS = solar water disinfection; LL / UL = lower / upper limit of the 95% confidence intervals; N/A = not estimated due to numerical problems; MAE = Mean absolute error; RMSE = Root mean square error.

The results of the bounded linear regressions for estimating the intervention effects are presented in Table S-1 of the SI. The fit indicators for each model are compiled at the bottom of Table 4. While the overall fit of the models is acceptable, the fit was only found to be good for remembering, behavior control, and both SODIS attitudes MAE, ≤0.25; more than 50% of cases with $\lvert e \rvert \leq 0.25$ and less than 20% cases with $\lvert e \rvert > 0.5$.

Predictions of Intervention Effects

Table 4 shows predicted intervention effects together with their 95% confidence intervals adjusted for the changes that occurred without interventions (control group). Unadjusted estimates can be found in Table S-2 of the SI. The results in Table 4 can be used directly for planning a campaign. For example, it is expected that prompts increase the behavior

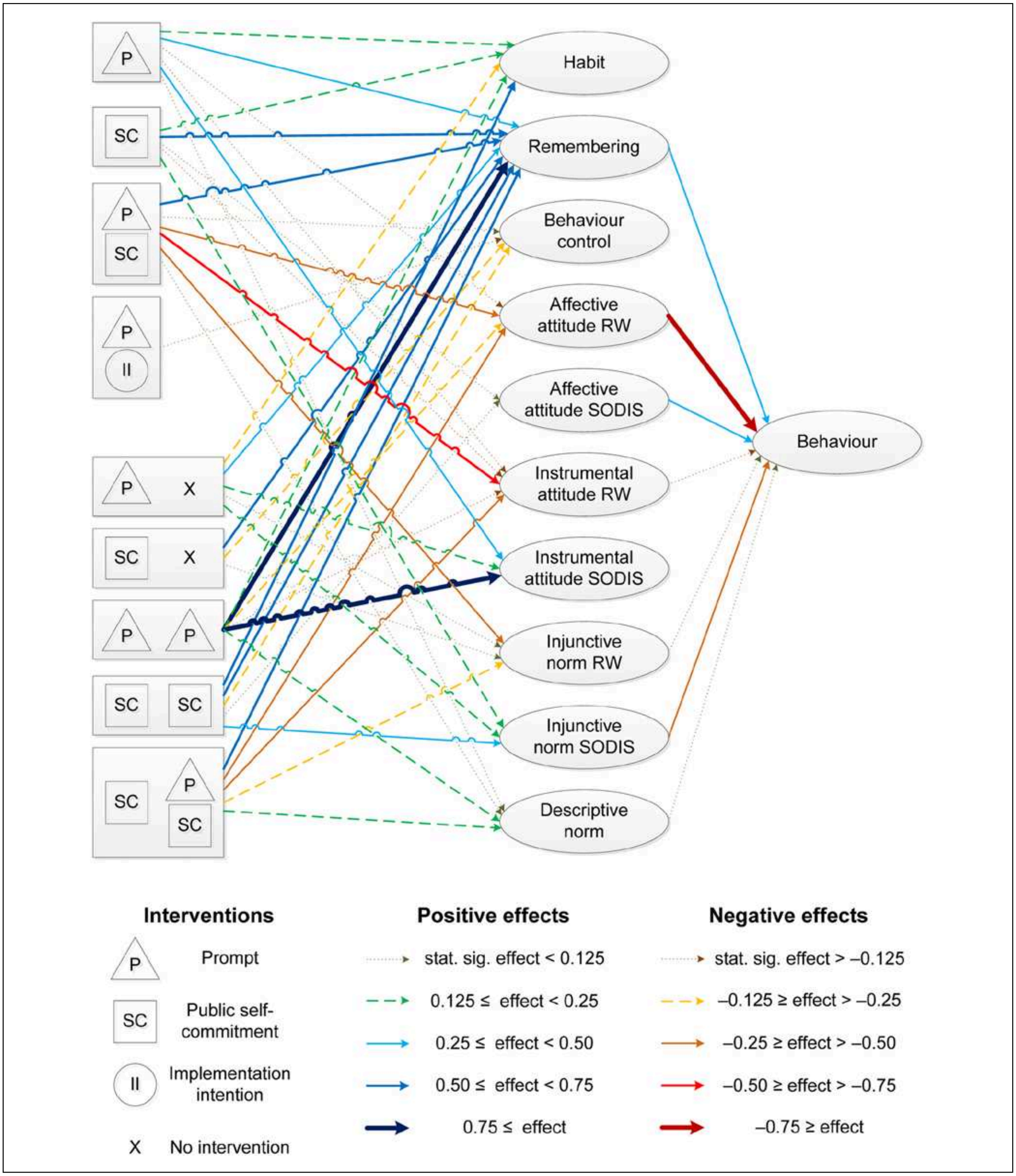

Figure 2. Overview on effects of interventions and changes of constructs.

0.395; with 97.5% probability, the effect will be 0.211 or larger. Answering RQ2 and RQ3, the main results from Table 4 can be summarized as follows. Figure 2 gives an overview of the effects compiled in Table 4.

Effects on behavior change. Applied once and alone, neither prompts (LL = 0.211) nor public self-commitments (LL = 0.088) show relevant effects. However, if combined, the effect of these BCTs is more than double the

Table 5. Results of the Bounded-Linear Regression of Changes in Behavior on Changes in Psychological Constructs.

Psychological constructs	*B*	*p*	95% CI LL	95% CI UL
Affective attitude RW	−0.844	<.001	−0.978	−0.734
Affective attitude SW	0.608	<.001	0.334	0.849
Instrumental attitude RW	−0.107	.021	−0.200	−0.008
Instrumental attitude SW	N/A	N/A	N/A	N/A
Injunctive norm RW	0.229	<.001	0.089	0.357
Injunctive norm SW	−0.436	<.001	−0.573	−0.300
Descriptive norm	0.257	<.001	0.119	0.405
Remembering	0.680	<.001	0.422	0.929
Constant time 1	0.160	<.001	0.057	0.258
Constant time 2	−0.026	.399	−0.099	0.054

Note. n = 172, R^2 = 86%, adjusted R^2 = 86%, standard error of the estimate = 0.279, MAE = 0.238, RMSE = 0.281, $|e| \leq 0.25$ = 67.8%, $|e| > 0.50$ = 4.7%. The parameters for the instrumental attitude of SODIS-treated water could not be estimated due to numerical problems. B = nominal value of estimated parameters; CI = confidence interval; LL/UL = lower/upper limit of the 95% CI; RW = raw water; SW = SODIS treated water; N/A = not estimated due to numerical problems.

sum of their single effects (LL = 0.638) and shows a statistically significant difference when compared with the effect of prompts and public self-commitments alone. Similar effects are found for the combination of prompts and implementation intentions (LL = 0.489) and repeating the same BCTs (LL = 0.597 and LL = 0.686, respectively). Repeating and combining prompts and public self-commitments has the strongest effect (LL = 1.362, showing a statistically significant difference in comparison with all the other effects). The confidence intervals (CIs) for differences between effects are compiled in Table S-3 of the SI.

Effects on psychological constructs. Prompts (LL = 0.494) and public self-commitments (LL = 0.567) have a weak to medium effect on remembering, and prompts also have a weak effect on the instrumental attitude toward SODIS (LL = 0.261). Furthermore, public self-commitments have a non-relevant influence on the injunctive norm of SODIS (LL = 0.230). If combined, the two BCTs show the same effect on remembering (LL = 0.564), but weak to medium effects for all evaluations of raw water (LL = −0.338 on affective attitude; LL = −0.728 on instrumental attitude; LL = −0.300 on injunctive norm). The combination of prompts with implementation intentions shows no relevant effect on psychological variables. This might be due to the small number of cases that received this treatment. If the BCTs are repeated, the patterns of the effects (i.e., their influence relative to each other) remain the same, although the overall effects become stronger.

Explaining Behavior Change With Psychological Constructs

Changes due to interventions. Table 5 compiles the results of the bounded linear regression analysis of behavior change on the changes of the psychological constructs. The estimate for the parameter of the instrumental attitude toward SODIS-treated water turned out to be unstable (i.e., the value changed completely when the model specification was altered slightly, such as by adding or removing a variable). This problem might be due to the many cases with high values for this variable that already existed before the intervention. The potential intervention effect had to be estimated with a few cases that still had a low value before the intervention. Because of this, we did not use the instrumental SODIS attitude in the analyses, even though this construct might be a strong mediator for intervention effects in many cases. The model fits the data well (adj. R^2 = 86%, MAE = 0.238, 68% of the data points have an $|e| \leq 0.25$, and only 5% an $|e| > 0.5$). This indicates a high reliability of the measurements and an adequate model specification, including considering all the important determinants of behavior change.

All psychological constructs have statistically significant effects on behavior change, even though the effect of the instrumental attitude toward raw water is minimal. Surprisingly, the injunctive norms are negatively related to the respective behavior. Because the injunctive norm for SODIS is not correlated with behavior (r = 0.031, p = 0.652) and the injunctive norm for raw water consumption is negatively correlated (as expected) with behavior (r = −0.250, p < 0.001), this result can be explained in relation to the strong effect of the affective attitudes and remembering. Two interpretations are possible: (1) Persons who felt social pressure to perform SODIS did not respond as strongly to the change in attitudes and remembering as did persons with less perceived social pressure; and (2) what seems more probable is that those with more perceived social pressure reported a stronger increase in attitude and remembering that is not reflected in the behavior change. Thus, the effects of the injunctive norms might correct the effects of social desirability on attitudes and remembering. The constant also has a weak influence in the first time step (B = 0.16). Therefore, some systematic change in behavior is not explained by the psychological variables.

Changes without intervention. For the first time step, no relevant effects were found in the control group, even with the barely relevant effect on descriptive norms (LL = 0.242). In the second time step, weak effects on behavior (LL = 0.317) and the instrumental attitude toward raw water (LL = −0.372) were observed in the control group. The remaining effects of single interventions were found only for remembering (LL = 0.474 for prompts; LL = 0.588 for public self-commitments). The behavior returned to the same value as for the control

Table 6. Results of the Bounded-Linear Regression of Behavior on the Psychological Constructs Before the Change for Cases Without Intervention.

Psychological constructs	B	p	95% CI LL	95% CI UL
Affective attitude RW	0.785	<.001	0.458	1.189
Affective attitude SW	0.619	.115	−0.077	1.380
Instrumental attitude RW	−0.085	.462	−0.412	0.187
Instrumental attitude SW	1.352	.004	0.618	1.909
Injunctive norm RW	−0.035	.891	−0.491	0.468
Injunctive norm SW	0.006	.854	−0.362	0.405
Descriptive norm	−0.088	.713	−0.494	0.396
Remembering	−0.555	.066	−1.123	0.015
Constant time 1	−1.046	.010	−1.600	−0.415
Constant time 2	−0.945	.006	−1.316	−0.481

Note. $n = 51$, $R^2 = 59\%$, adj. $R^2 = 49\%$, MAE = 0.305, RMSE = 0.375, $|e| \leq 0.25 = 45\%$, $|e| > 0.50 = 22\%$. B = nominal value of estimated parameters; CI = confidence interval; LL/UL = lower/upper limit of the 95% CI; RW = raw water; SW = SODIS-treated water.

group ($M = -0.121$ for prompts, $M = -0.104$ for public self-commitments).

To explain the aforementioned behavioral changes in the control group, a regression of behavior change on the psychological variables before the interventions was computed, as shown in Table 6. Due to the small number of cases, most estimates of the parameters are not statistically significant. The attitudes toward SODIS are strong predictors of behavioral change without interventions ($B = 1.35$ for instrumental and $B = 0.62$, n.s., for affective attitude). Surprisingly, however, the affective attitude toward raw water has a strong *positive* effect on behavior change ($B = 0.79$). This might be related to the fact that the taste of SODIS water is very close to the taste of raw water, and taste is an important component of the affective attitude. Thus, persons who prefer the taste of raw water to that of boiled water evaluated SODIS-treated water more positively. The only relevant (but barely statistically significant) negative effect is on remembering ($B = -0.56$), indicating that a lack of fully developed habits might be the principal barrier for people with positive attitudes to routinely perform SODIS.

Discussion

We applied a methodology that quantifies BCT effects in a way that allows forecasting the psychological and behavioral effects of intervention campaigns. The analyses led to a number of interesting substantial results. Those results are discussed next, after which the methodology itself is discussed.

Discussion of the Substantial Results

To investigate the potential impact of prompts, public self-commitments, and implementation intentions on behavior and on a number of psychological constructs, bounded linear regression models were fitted to data gathered during a large-scale behavior change campaign in rural Bolivia. Based on these models, the effects of the different BCTs were estimated. The results allow all of the RQs to be answered.

Results regarding the success of delivering the different BCTs (RQ1) and effects on behavior change (RQ2) are summarized from an application-oriented perspective. Prompts, on their own, had the strongest effect on behavior (expected change, EC = 0.395 of a maximal possible effect of 2.0), and they were the easiest to distribute (about 54% of the target population received the prompt). The same delivery success rate (56%), but a lower impact on behavior change (EC = 0.249), was observed for public self-commitments. Implementation intentions failed with respect to delivery success rates (22%); however, they relevantly increased the effect of prompts (EC = 0.883). A similar effect was found for the combination of prompts and public self-commitments (EC = 0.986). Thus, combining BCTs increases the effects on behavior change to levels higher than the sum of the two single BCTs. The same holds true for repeating the same technique over time (EC due to repeated prompts = 0.918, and due to repeated public self-commitments = 1.081). The strongest effect found in this study was the combination of prompts and public self-commitments after initial public self-commitments (EC = 1.818). Neither prompts nor public self-commitments showed long-term effects on the behavior if applied only once. No data are available on the long-term effects of combinations and repetitions of BCTs.

These results have two main implications for theory development and application. First, the limited success of delivering the BCTs illustrates the importance of determining whether interventions actually reach the target population. For theory development, the effects of successfully applied interventions need to be quantified independently from the success in delivering them, as completely different processes determine the two success rates. Regarding application, BCTs should be designed in a way that they appear attractive and are easy to understand and apply. The colorful and practical prompts could be handed out with little instruction and were well received by the people, which in turn motivated the health volunteers. The implementation intentions turned out to be too difficult for the local health volunteers to understand, too time-consuming to be applied, and not very attractive to the target population. Thus, within a given amount of time, the health volunteers were not able to apply as many of these BCTs as the others and encountered more refusals; thus, the two factors might have reduced the motivation of the volunteers, leading to an even lower delivery success rate. The critical role of promoter characteristics, particularly their level of commitment, was recognized by Meierhofer and Landolt (2009). Therefore, it is important to work with BCTs that keep the motivation of the promoters high.

Second, the effects of BCTs depend on other interventions applied before or at the same time. This is not surprising, but it is almost never scientifically investigated. Mosler et al. (2013) applied a number of BCTs consecutively, but the interactions between BCTs were not considered in the statistical models. The conclusion that can be derived from our results in terms of theory development is that intervention effects need to be investigated as individual processes to understand how they interact. For application purposes (i.e., campaign planning), it can be concluded that the investigated BCTs should be combined and repeated to increase their effectiveness.

Results regarding the effects of the BCTs on psychological constructs (RQ3) and the psychological mechanism behind the BCT effects (RQ4) can be summarized together from a theoretical perspective: Prompts affect the instrumental attitude toward SODIS, while public self-commitments affect the injunctive norm of SODIS. Combining prompts and public self-commitments leads to effects on raw-water-related attitudes and norms. Thus, the often assumed effects on attitudes and norms were confirmed. However, the strongest effects were found on remembering, which is absent in most behavior change theories. Only Tobias (2009) considered this construct a central driver of behavior change dynamics, and Mosler (2012) included it in his conceptual model. Moreover, no relevant effects were found on the descriptive norm. A possible reason may be the scattered layout of the settlements in the target region, which makes it difficult to know what neighbors are doing.

Considering the relationships between behavior change and changes in the psychological variables, it turns out that only change in remembering reflected a positive mediating effect of the BCTs on behavior change. It must be mentioned, however, that the effect of the change in instrumental attitude toward SODIS-treated water and the effect of perceived behavioral control on behavior change could not be estimated, due to numerical problems and too few cases, respectively. In particular, instrumental attitude might be an important mediator for the prompt effect in cases where this construct is not very high before the interventions. Roughly, the BCTs changed remembering at least by about 0.5 (lower limit of the 95% CI), while the lower limit of the CI for the effect of a change in remembering on behavior change is 0.4. Thus, with a probability of more than 97.5%, the investigated BCTs changed the behavior by at least $0.5 \times 0.4 = 0.2$ due to a mediation of the change in remembering. The expected mediation effect of remembering is about 0.5 for single interventions and 0.7 for repeated interventions on behavior.

It might be seen as surprising that prompts and public self-commitments show such similar effects, as they are often categorized as completely different behavior change techniques. For example, Michie et al. (2013) categorized prompts within the cluster "associations," and behavioral contracts (what we call self-commitment) in the cluster "goals and planning" (together with action planning, what we called implementation intentions). However, the effects of BCTs are determined not only by the form of the BCT but also by the problem it solves. Here, the problem was forgetting to put the bottles in the sun at least 6 hr before the water was needed. As it seems that both techniques could solve the problem, and because this was the only critical factor that hindered the performance of the behavior, both techniques show similar effects. One might wonder how self-commitment can prevent forgetting. One explanation is that the commitment sign worked as a prompt; another is that the self-commitment intervention increased the importance of the behavior, and, thus, the persons in charge put more effort into not forgetting to put the bottles in the sun (e.g., by setting up self-made reminders or associating specific situations with performing the behavior—something that also occurred in the control group in the experiment by Gollwitzer & Brandstätter, 1997).

In addition, strong relations between the affective attitudes and behavior change and weaker ones between changes in the norms and behavior change were observed, but no systematic effects of the intervention on these constructs were detected. It might be that the relationships between the changes in the psychological constructs and behavior change reflect only a self-report bias (i.e., over-reporting of the change in attitudes by people who felt social pressure to use SODIS) or that the effects of the interventions on the psychological constructs were not detected due to insufficient measurement or modeling.

Another modeling concept used for planning behavior change campaigns can be discussed as well: stage models (e.g., the Transtheoretical Model of Change by Prochaska & DiClemente, 1983; the Health Action Process Approach by Schwarzer, 2008). Such models have been successfully applied to the use of SODIS (e.g., Kraemer & Mosler, 2011; Mosler & Kraemer, 2012). However, our model explains that the data comprising all levels of SODIS use almost perfectly without considering stages of change. The reason for this is that the bounded value ranges of the variables have been considered in the model. As in stage models, changing some constructs might have no effect for certain persons. However, this is not because these persons are in stages wherein the changes of the constructs have no effect, but because the constructs have values close to their bounds, and, thus, nothing can be won by changing these constructs. In addition, cases without interventions (i.e., in the control group) showed changes in behavior at the second time step (EC = 0.525) and changes in instrumental attitude toward raw water (EC = −0.589). Such changes in the control group were also observed in other SODIS promotion campaigns (e.g., Mosler et al., 2013), but they were never analyzed. According to our results, persons who demonstrate a greater increase in SODIS use have not only more positive attitudes toward SODIS but also a more positive affective attitude toward consuming raw water. As mentioned before, this might be due to the

Table 7. Limitations of Previous Studies and How These Issues Were Solved in This Study.

Limitations of previous studies	How these issues were solved in this study
Setting is artificial: laboratory or field experiment instead of a real-world campaign	For this study, data were gathered during an actual campaign.
Investigation limited to test for statistically significant differences	The changes are quantified using meaningful (psycho-)metric scales.
Only differences between an intervention group and a control group are investigated, or a combination of various techniques is investigated as one intervention without differentiating the effects of the techniques applied.	A number of explicitly defined combinations of different intervention techniques are investigated considering the effects of each technique on its own and in combination with other techniques or repetitions of the same technique.
Only behavioral outcomes are measured.	Behavioral outcomes and a number of psychological constructs are measured.
Considering only cross-sectional data	Longitudinal data were used for this study.
Data is only investigated in form of linear covariance structures.	A case-based approach with bounded-linear models was used considering the limited range of the constructs.
The delivery of the behavior-change techniques is artificial or assumed to be perfect.	The behavior-change techniques were delivered within a real-world campaign and the delivery success investigated.

similarity of the taste of SODIS-treated water to that of raw water. More generally, it is important to note that a positive attitude toward a competing behavior does not necessarily impede the target behavior; in fact, it may even promote it. Therefore, theories of behavior change should consider the interaction of competing behaviors.

There is another interesting aspect related to these results: The attitudes are not related to the behavior itself but to the *change* in the behavior. Thus, persons who changed their behavior without intervention treated less water with SODIS at the beginning than at the end of the campaign. If these persons evaluate SODIS so positively, why did they not use SODIS right from the beginning? Because remembering is negatively related to behavior change, it can be concluded that these persons often forgot to apply SODIS at the right moment. However, as Tobias (2009) demonstrated, if they were still applying SODIS at least once in a while, habits could have developed that prevented forgetting after some months, even without intervention.

Discussion of the Method

The methodology used in this study to quantify intervention effects consists of a number of key elements. Most of the methods used are common in engineering and the natural sciences but not as common in psychology. Therefore, the approach is summarized in terms of the design of the campaign, data gathering, and data analysis.

Campaign. To investigate intervention effects, data should be gathered during actual campaigns. Ideally, these are large-scale campaigns, but smaller pilot campaigns can be more practical for trying out a number of BCTs. In these campaigns, different BCTs should be applied on their own, in combination, and in sequence to different target groups. Furthermore, data gathering must be designed in a manner that allows investigation of possible problems with the delivery of the BCTs.

Data gathering. To obtain usable estimates of intervention effects, longitudinal and (approximate) metric data must be gathered. In psychology, this can be achieved by asking the interviewees for very simple and specific evaluations of the behaviors of interest. Based on these data, scores for more complex and abstract constructs can be computed.

Analysis. Intervention effects should be quantified in absolute terms and not just by demonstrating the statistical significance of an effect compared with a control group. This can be accomplished with regression analyses. However, the following points must be considered. First, bounded linear models are necessary if the depended variable is bounded, as in the case of metric data on psychological constructs. Second, to determine the effects of BCTs, predictions of the effects must be calculated based on the parameter estimates. Third, the uncertainty of these predictions must be estimated. Bootstrapping can be a useful approach, because it considers interdependencies among uncertainties of parameter estimates.

This approach quantifies intervention effects in a form that can be used for the development of behavior change campaigns such as health-promotion interventions. Thus, the shortcoming of many studies criticized in the introduction (i.e., that they only show *that* an intervention is effective but not *how much* and *what type* of effect can be expected) is overcome. Furthermore, this methodology supports the development of process theories of behavior change and solves all of the limitations of previous studies mentioned in the introduction. Table 7 summarizes these issues and how they were solved using the above approach.

A particular strength of the present study is that it uses data gathered from a "real-world" campaign, thereby ensuring high external validity. However, this comes at the price of not having data of the highest quality. Therefore, for this specific study, a number of shortcomings have to be considered before generalizing the results.

First, the forecasts of the intervention effects are rather rough. For most constructs, adequate forecasts (i.e., with $|e| \leq 0.25$) could only be achieved for about half of the sample, and in many cases, about one-quarter of the forecasts were unusable (i.e., $|e| > 0.5$). The problem is also reflected in the poor explained variance (adj. R^2), which is often below 50%, and for some models, is even below 20%. This needs to be considered, particularly in the case of further analyzing the forecasts. A second problem is the small number of cases for some interventions. In particular, few cases received combinations of BCTs, and results regarding these effects should be interpreted with caution. Third, the cases were not completely randomly assigned to the treatment conditions. The households could not select the BCTs, but they could reject them. However, no statistically significant differences in behavior or psychological constructs before the interventions were found between households with and without intervention.[2]

Finally, we could not implement a mediation analysis, even though they are commonly performed in psychological studies (e.g., Baron & Kenny, 1986; Hayes, 2009; MacKinnon, Fairchild, & Fritz, 2007). We did calculate all models (i.e., the direct effects of the interventions on the behavior, the effects of the interventions on the psychological constructs, and the effects of the psychological constructs on the behavior), but we did not correct for direct effect in the regression of the behavior on the psychological constructs. This was not possible due to the high explicative power of the models: Considering the direct effect in the regression of the psychological constructs would have led to strong multicollinearity. However, our approach allows an absolute estimate of the mediation effect under consideration on uncertainty in the sample, measurement, and model, which might be even superior to the traditional approach, which only tests for statistical significance. Nevertheless, when comparing the results with other mediation analyses, this difference has to be considered.

Conclusions

For any science, the ability to forecast the effects of the application of a technique in the real world is a critical step in the application as well as development of a knowledge domain. Regarding application, science should help us foresee the consequences of our actions; regarding theory development, deviations from expected and observed consequences are the most valuable basis for improving models. A prerequisite for this step is to apply adequate methods of analysis to adequate data. In this article, we presented a methodology that allows quantification of the effects of BCTs in a form that allows forecasting of the effects of real-world campaigns. Furthermore, we provided first insights into how three BCTs worked in a large-scale health promotion intervention. However, these results are only a first step, and further research is needed to actually understand how BCTs work. Besides quantifying the effects of other campaigns to see how far the results of this case study can be generalized and what other effects the investigated BCTs can have, new theories of behavior change are needed. Such theories need to focus on psychological *processes* triggered by BCTs instead of only listing possible determinants and stages of behavior change. Investigating individual processes during large-scale campaigns requires different approaches to data gathering and analysis, as presented herein. Knowing what effects can be expected from different BCTs could help in the design of better behavior change campaigns, and, thus, mitigate many urgent problems faced by humanity.

Acknowledgments

The authors acknowledge the support of the UBS Optimus Foundation, Switzerland. They thank the SODIS Foundation (Fundación SODIS; Cochabamba, Bolivia), the Ministry of Health of Chuquisaca, and the Departmental Health Service (SEDES; Sucre, Bolivia); Carmen Camargo for local project management; and Chantal Bourloud, Esther Bisig, and Jasmin Meyer for conducting the field work. Further thanks go to Hans-Joachim Mosler (Eawag), who initiated the project, and Thomas Buchsbaum (Eawag), who helped develop the non-linear regression analyses. In addition, both provided valuable comments that improved this manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

Notes

1. Because the researchers only gathered and analyzed the data and the campaign was implemented by official organizations, and because no medical or physical act was involved in this data gathering, under Swiss regulations, ethics approval was not required.
2. The only exception is the injunctive SODIS norm, where the mean of cases with intervention (0.30, $n = 76$) is significantly smaller than the mean for cases without intervention (0.43, $n = 61$, $p = 0.019$).

References

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, *50*, 179-211.

Armitage, C. J. (2007). Effects of an implementation intention-based intervention on fruit consumption. *Psychology & Health*, *22*, 917-928.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Penalty and Social Psychology*, *51*, 1173-1182.

Berney, M., Weilenmann, H. U., Simonetti, A., & Egli, T. (2006). Efficacy of solar disinfection of Escherichia coli, Shigella flexneri, Salmonella Typhimurium and Vibrio cholerae. *Journal of Applied Microbiology*, *101*, 828-836.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203-219.

Burn, S. M., & Oskamp, S. (1986). Increasing community recycling with persuasive communication and public commitment. *Journal of Applied Social Psychology*, *16*, 29-41.

Cairncross, S., Shordt, K., Zacharia, S., & Govindan, B. K. (2005). What causes sustainable changes in hygiene behaviour? A cross-sectional study from Kerala, India. *Social Science & Medicine*, *61*, 2212-2220.

Chasteen, A. L., Park, D. C., & Schwarz, N. (2001). Implementation intentions and facilitation of prospective memory. *Psychological Science*, *12*, 457-461.

Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Glenview, IL: Scott, Foresman.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*, 1015-1026.

Cox, C. D., Cox, B. S., & Cox, D. J. (2005). Long-term benefits of prompts to use safety belts among drivers exiting senior communities. *Journal of Applied Behaviour Analysis*, *38*, 533-536.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*, 493-503.

Gollwitzer, P. M., & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology*, *73*, 186-199.

Guynn, M. J., McDaniel, M. A., & Einstein, G. O. (1998). Prospective memory: When reminders fail. *Memory & Cognition*, *26*, 287-298.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*, 408-420.

Heri, S., & Mosler, H.-J. (2008). Factors affecting the diffusion of solar water disinfection: A field study in Bolivia. *Health Education & Behaviour*, *35*, 541-560.

Hill, C., Abraham, C., & Wright, D. B. (2007). Can theory-based messages in combination with cognitive prompts promote exercise in classroom settings? *Social Science & Medicine*, *65*, 1049-1058.

Intons-Peterson, M. J., & Fournier, J. (1986). External and internal memory aids: When and how often do we use them? *Journal of Experimental Psychology: General*, *115*, 267-280.

Kiesler, C. A. (1971). The psychology of commitment: Experiments linking behaviour to belief. New York, NY: Academic Press.

Kraemer, S. M., & Mosler, H.-J. (2010). Persuasion factors influencing the decision to use sustainable household water treatment. *International Journal of Environmental Health Research*, *20*, 61-79.

Kraemer, S. M., & Mosler, H.-J. (2011). Factors from the Transtheoretical Model differentiating between Solar Water Disinfection (SODIS) user groups. *Journal of Health Psychology*, *16*, 126-136.

Law, K. S., Wong, C.-S., & Mobley, W. H. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review*, *23*, 741-755.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593-614.

McGuigan, K. G., Conroy, R. M., Mosler, H.-J., du Preez, M., Ubomba-Jaswa, E., & Fernandez-Ibañez, P. (2012). Solar water disinfection (SODIS): A review from bench-top to roof-top. *Journal of Hazardous Materials*, *235–236*, 29-46.

Meierhofer, R., & Landolt, G. (2009). Factors supporting the sustained use of solar water disinfection: Experiences from a global promotion and dissemination programme. *Desalination*, *248*, 144-151.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., . . . Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, *46*, 81-95.

Michie, S., Rothman, A. J., & Sheeran, P. (2007). Current issues and new direction in Psychology and Health: Advancing the science of behavior change. *Psychology & Health*, *22*, 249-253.

Mosler, H. J. (2012). A systematic approach to behavior change interventions for the water and sanitation sector in developing countries: A conceptual model, a review, and a guideline. *International Journal of Environmental Health Research*, *22*, 431-449.

Mosler, H.-J., & Kraemer, S. M. (2012). Which psychological factors change when habitual water treatment practices alter? *Journal of Public Health*, *20*, 71-79.

Mosler, H.-J., Kraemer, S. M., & Johnston, R. B. (2013). Achieving long-term use of solar water disinfection in Zimbabwe. *Public Health*, *127*, 92-98.

Nyer, P. U., & Dellande, S. (2010). Public commitment as a motivator for weight loss. *Psychology & Marketing*, *27*, 1-12.

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, *51*, 390-395.

Prüss-Üstün, A., Bos, R., Gore, F., & Bartram, J. (2008). Safer water, better health: Costs, benefits and sustainability of interventions to protect and promote health. Geneva, Switzerland: WHO.

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology: An International Review*, *57*, 1-29.

Shapiro, S., & Krishnan, H. S. (1999). Consumer memory for intentions: A prospective memory perspective. *Journal of Experimental Psychology: Applied*, *5*, 169-189.

Sheeran, P., Aarts, H., Custers, R., Rivis, A., Webb, T. L., & Cooke, R. (2005). The goal-dependent automaticity of drinking habits. *British Journal of Social Psychology*, *44*, 47-63.

Sommer, B., Mariño, A., Solarte, Y., Salas, M. L., Dierolf, C., Valiente, C., . . . Wegelin, M. (1997). SODIS-an emerging water treatment process. *Journal of Water Supply Research and Technology-Aqua*, *46*, 127-137.

Tamas, A., & Mosler, H.-J. (2011). Why do people stop treating contaminated drinking water with solar water disinfection (SODIS)? *Health Education & Behavior*, *38*, 357-366.

Tobias, R. (2009). Changing behaviour by memory aids: A social-psychological model of prospective memory and habit development tested with dynamic field data. *Psychological Review*, *116*, 408-438.

Trafimow, D., & Sheeran, P. (1998). Some tests of the distinction between cognitive and affective beliefs. *Journal of Experimental Social Psychology*, *34*, 378-397.

Webb, T. L., Ononaiye, M. S. P., Sheeran, P., Reidy, J. G., & Lavda, A. (2010). Using implementation intentions to overcome the effects of social anxiety on attention and appraisals of performance. *Personality and Social Psychology Bulletin*, *36*, 612-627.

Author Biographies

Robert Tobias, PhD in Psychology and MSc in Civil Engineering, is specialized in theory- and evidence-based modeling of behavior-change campaigns using statistical and computer-simulation methods. His current research interests focus on heuristic information processing and communication, and effects of memory and habits on behavior performance. He is currently employed in a tenure-track position at the Swiss Federal Institute of Aquatic Science and Technology (Eawag).

Alexandra C. Huber, PhD, was a doctoral candidate at the Swiss Federal Institute of Aquatic Science and Technology (Eawag) working on behavior change of health behaviors in developing countries. Currently she is working in Ethiopia for Addis Ababa University as a lecturer in Health Psychology.

Andrea Tamas received her PhD from the Swiss Federal Institute of Aquatic Science and Technology with a specialization in behavior change theory applied to health issues and intervention campaign design. She is now a member of the School of Health Professions at the Zurich University of Applied Sciences.

SUPPORTING INFORMATION (SI)

A methodology for quantifying effects and psychological functioning of behaviour-change techniques

Robert Tobias*, Alexandra C. Huber, and Andrea Tamas

Department Environmental Social Sciences (ESS)

Eawag, Swiss Federal Institute of Aquatic Science and Technology,

Dübendorf, Switzerland

*Corresponding author:

Eawag, ESS, Überlandstrasse 133, P.O. Box 611, 8600 Dübendorf, Switzerland

Phone: +41 (0)58 765 5432, E-Mail: robert.tobias@eawag.ch

Number of text pages: 5

Number of figures: 0

Number of tables: 3

Total number of pages SI: 10

Further Explanations of Methods

Operationalisation of constructs

In health psychology, constructs are mostly operationalized as latent variables. The score of a latent variable is estimated based on a number of highly correlated manifest indicators. Latent variables are usually unbounded and allow fine-graded ordering of individual cases within a sample. This makes them perfectly suited for covariance-based analyses on the level of the population or the overall sample. However, latent constructs exist only on the population level (Borsboom, Mellenbergh, & van Heerden, 2003) and cannot be used for investigating individual processes such as effects of psychological interventions. For estimating the impact of such interventions, data must be gathered that can be interpreted for each individual case (i.e., in absolute terms) and not only in comparison to other cases. Or in other words: Approximately metric data must be gathered on manifest constructs that reflect individual processes.

Without going into the details of measurement theory, it can be summarized that gathering approximately metric data of psychological constructs has its limitations. First, the constructs need to be very simple and specific self-evaluations (one example for such a self-evaluation might be a question on how SODIS water tastes to individuals). Complex constructs or constructs whose values have only a meaning relative to values of other individuals (e.g. intelligence) cannot be measured this way. Second, the interpretation as metric of the scores for such specific self-evaluations is a rough approximation. The unit of these variables is given by the difference between the lowest possible evaluation and the neutral value and/or the difference between the neutral value and the highest possible evaluation. However, the relation between the numeric differences between scores and the psychological meaning of these differences might not be completely linear. For example, a change from 0 to 0.2 might have less effect than a change from 0.8 to 1.0 on a scale from –1 to +1. Anyway, until more information is available on this relation, the simplest model will be used, i.e. a linear function. Finally, the bounded value range of such variables needs to be considered. Since the value bounds are not just artefacts of a limited sample or scale but theoretically required and meaningful, floor and ceiling effects need to be dealt with when using such variables in statistical models (see the sub-section on bounded-linear regression analyses below).

Considering the above limitations, a number of simple and specific self-evaluations have been measured. These variables were not directly used in the regression models, but aggregated into more complex and abstract constructs. A number of reasons argue in favour of this aggregation. (1) The constructs do have a different theoretical meaning and are more generalisable than the specific evaluations. Instrumental attitude, for example, is defined as the sum of all advantages and disadvantages of a behaviour. This construct can be used in many studies. However, the specific evaluations considered as advantages and disadvantages in a given case might differ. (2) For statistical modelling, a low (manageable) number of parameters needs to be considered. Each construct represents the essential information of many specific evaluations. Thus, using aggregations helps avoid problems with computation. (3) Depending on the function used for computing the score of a construct, measurement errors are partially compensated by the aggregation. This is the case for the linear combinations used in this study.

One might ask why we did not directly measure the abstract constructs. For the instrumental attitude, a question like “Is it worthwhile doing the behaviour?” would have captured this construct. However, such an operationalisation does not provide information about the advantages and disadvantages the interviewee considered. Such information is valuable for campaign planning. Furthermore, since each interviewee in each setting might consider other advantages and disadvantages, the reliability of such an item might be low. The closer we come to the individual psychological process leading to the score of a construct, the better the data works for modelling processes of change due to interventions.

As a final point, the consideration of measurement quality in the models requires discussion. In the standard approach using latent constructs, reliability is expressed as internal consistency (e.g., greatest lower bound, Cronbach's alpha): the higher the intercorrelations of the items used for operationalising a construct, the better. However, internal consistency has no meaning at all for aggregate constructs. Using again the example of instrumental attitude, evaluations of whether SODIS has positive or negative effects on health or concerning whether SODIS is expensive or inexpensive may be completely independent. However, since we assume a metric scale for the constructs, reliability can be expressed for each case by the difference of two equal measurements. The difference of values of a construct in two panels is partly due to systematic changes represented by a model of change, partly due to actual changes in value not captured by the model of change, and partly due to a lack of reliability of the measurement. Thus, residuals represent insufficiencies of the model of change on the one hand, and lack of reliability on the other hand. In psychology, no simple way exists to separate these two sources of uncertainty. (Note that, also in the case of low internal consistency of items of a latent variable, it remains unknown whether the problem is due to items measuring the same attribute imprecisely, or different attributes.) So it seems best to document the total uncertainty (e.g. the standard error of the estimate) and let the user of the results decide what imperfection is more problematic. Since in most cases it will be the imperfection of the model of change, it is on the safe side to assume perfect measurement.

Bounded-linear regression analyses

This investigation uses bounded-linear regression models that will be explained here in more detail. One precondition for using linear regression models is that the dependent variable is unbounded, i.e., its range is $(-\infty, +\infty)$. Often, even with bounded dependent variables (e.g., variables that cannot be smaller than 0), usable results can be achieved using linear regression models. For the case presented, however, using a linear regression model would lead to biases. Due to the bounded range of the dependent variable, the observable effect of an intervention can be smaller than its maximal effect. If, for example, the potential effect of an intervention is 0.6, this effect can only be observed for cases that have a value before applying the intervention of 0.4 or less, since the behaviour is bounded in [-1, +1]. Thus, the more cases have a value larger than 0.4 before the intervention the more the effect of the intervention is underestimated. Therefore, the regression model must consider the range an observation of the intervention effect can maximally have. If the observed effect of an intervention is equal to the maximal observable effect, the actual intervention effect can be of this size *or larger*.

In this investigation, we used the following regression model:

$$\Delta\hat{y} = \mathrm{Max}(-1 - y_{t-1}, \mathrm{Min}(1 - y_{t-1}, b_1 x_1 + b_2 x_2 + \dots + b_n x_n + c_1 t_1 + c_2 t_2 + \varepsilon)) \quad (\mathbf{1})$$

$\Delta\hat{y}$ is the expected change of a variable (behaviour or psychological construct) between two panels (i.e., in one time step) due to the interventions or the changes of the mediating constructs. y_{t-1} is the value of this variable in the previous panel (i.e., Panel 1 for the first and Panel 2 for the second time step). x_1 to x_n are dummy variables for the presence of an intervention (e.g., if a prompt was present in a time step, the variables for prompts in the current time step are set to 1; otherwise, they are set to 0). b_1 to b_n are the parameters that quantify the effect of the different interventions on y. c_1 and c_2 are constants expressing systematic changes of y that cannot be explained by the interventions. Here, two constants are used, one for each time step. All changes of the first time step have $t_1 = 1$ and $t_2 = 0$, and all changes of the second time step have $t_1 = 0$ and $t_2 = 1$. Finally, ε is the error term. The formula states that if the sum $b_1 x_1 + \dots + b_n x_n + c_1 t_1 + c_2 t_2 + \varepsilon$ is larger than $-1 - y_{t-1}$ or smaller than $1 - y_{t-1}$ then $\Delta\hat{y}$ is equal to this sum. Otherwise, instead of the sum, $-1 - y_{t-1}$ and $1 - y_{t-1}$ is used respectively.

This model directly represents a linear relationship to a bounded dependent variable. Unfortunately, estimating the parameters for this model is not trivial. This will be explained in the next sub-section.

Estimating the parameter values

Most statistical packages use gradient search algorithms for estimating the parameters of regression models. These algorithms start from an arbitrary combination of parameter values, change these values a little, and then check whether the changes improved the model fit. Based on this result, the next trial of parameter values is determined and then tested. If changing the parameter values does not improve the model fit, these values are returned as a result. Such algorithms are very efficient for many applications, particularly for linear models, but not suitable for the model used here. The problem is that large parts of the solution space are completely flat. Thus, even if the parameter values are far from the optimum, changing them locally does not improve the fit, and a gradient search algorithm erroneously assumes that it converges on the optimum value. The result is completely arbitrary and cannot be interpreted seriously.

The model used in this investigation requires a global search algorithm that probes the entire solution space. We used a Differential Evolution approach (Storn & Price, 1997). One disadvantage of global search algorithms is that their results are less accurate. Because of this, we used the result of the global search algorithm as a starting point for an additional local gradient search. This combination of algorithms is computationally expensive but leads to accurate parameter estimates.

Forecasting potential intervention effects

The parameter estimates of the regression models cannot be directly interpreted in terms of intervention effects. What is needed for campaign planning and this investigation are predictions of the intervention effects based on the estimated regression models. Such predictions are done by inserting the values of the intervention variables that define the scenario under investigation into the estimated regression models. For forecasting changes of empirical cases, Equation 1 is applied separately for each time step. For estimating the *potential* effect of certain interventions, the boundary conditions can be omitted and the effects of various time steps can be added within the same sum. This is explained for the example of the effect on behaviour change of the strongest intervention: a combination of prompts and public self-commitment after a public self-commitment in the previous wave of interventions.

The parameter estimates are presented in Table S-1 of this document. Entering the parameter estimates in the regression equation leads to the following formula for predicting intervention effects:

$$\begin{aligned}\Delta\hat{y} =\ & 0.237 * \text{Time-Step 1} \\ & + 0.289 * \text{Time-Step 2} \\ & + 0.395 * \text{Current prompt} \\ & + 0.249 * \text{Current public self-commitment} \\ & + 0.342 * \text{Interact. prompt w. pub. com.} \\ & + 0.488 * \text{Prompt with imp. int.} \\ & - 0.516 * \text{Prompt in previous step} \\ & - 0.353 * \text{Pub. com. in previous step} \\ & + 0.706 * \text{Interact. curr. \& prev. prompt} \\ & + 0.936 * \text{Interact. curr. \& prev. pub. com.}\end{aligned} \qquad (\mathbf{2})$$

Next, we have to define the scenario we want to investigate. For the intervention described above, the intervention variables have to be set to the following values:

Time-Step 1 = 1
Time-Step 2 = 1
Current prompt = 1
Current public self-commitment = 2
Interact. prompt w. pub. com. = 1
Prompt with imp. int. = 0
Prompt in previous step = 0
Pub. com. in previous step = 1
Interact. curr. & prev. prompt = 0
Interact. curr. & prev. pub. com. = 1

The variable *current public self-commitment* is set to 2, since this effect is active in the first and second time step. Entering these values into Equation 2 leads to the following formula:

$$\Delta\hat{y} = 0.237 + 0.289 + 0.395 + 2 * 0.249 + 0.342 - 0.353 + 0.936 = 2.344 \qquad (\mathbf{3})$$

Thus, the potential effect of this intervention is 2.344[1], as can be seen in Table S-2. In Table 5 of the paper, the effects are adjusted for control cases without intervention. Thus, Equation 3 is calculated without the constants leading to the value 1.818. Note, that the results in the tables are slightly different. The reason for this is that a further step is necessary to get the final results: Estimating the uncertainty of the predictions. This is explained in the next sub-section.

Bootstrapping confidence intervals

Usually, the confidence intervals (or standard errors) of the parameter estimates are calculated based on the assumption of normally distributed errors. This approach is computationally very efficient but requires the errors to be actually normally distributed. For all models, even though the residuals of the regression models were almost normally distributed, the distribution of the errors of the forecasted intervention effects could not be assumed to be normal. Furthermore, for forecasting the intervention effects, dependencies between the parameter estimates need to be considered. Otherwise, the uncertainty of the predictions is overestimated.

For this study, a computationally very expensive but very general approach for estimating uncertainties was used: bootstrapping. This approach does not require any assumption on distributions and implicitly considers dependencies of the parameter estimates when estimating the uncertainty of forecasts. The only precondition for this method is that the sample is representative for the population – an assumption of all methods of inferential statistics. Under this condition, bootstrapping simulates the process of sampling from the population based on the sample. From the sample, cases are randomly selected and put back for the next random selection. This is repeated the same number of times as the sample has cases. Each case represents a particular combination of values. Combinations that are more frequent in the sample have a larger probability of being selected as is the case when sampling from the population. However, if the sample has a large variance, the bootstrap samples drawn might vary considerably. Therefore, this sampling procedure is repeated many times – in this study 1000 times – and the statistics of interested are computed with each bootstrap sample. All values of the statistics computed are saved and after repeating the procedure the preset number of times (here 1000), the mean of these values represents the nominal or expected value of the parameter estimate, the standard deviation is the standard error, and the 2.5 and 97.5 percentile the lower and upper limit of the 95% confidence interval. Thus, the variance in the bootstrap samples determines the uncertainty of the statistics.

This approach was used to compute the uncertainties of all statistics presented in the paper. When applying this approach to regression analyses, some practical issues have to be considered. Firstly, there is a risk of drawing the bootstrap sample in a way that some variables do not have any variance. This might happen when in a drawn sample a certain intervention has the same status (either active or inactive) for all cases in that sample. Secondly, the bootstrap sample might have too great a degree of multicollinearity between the independent variables of the regression. In the present study, such problematic samples were simply omitted and a new bootstrap sample was drawn. Thus, to reach the required number of values for the statistics (i.e., 1000), more bootstrap samples were drawn (for the intervention effects, an average of 1087, with a maximum of 1334). For some models, certain parameters could not be estimated due to numerical problems (e.g. multicollinearity issues). Since a computationally expensive method for estimating the parameter values of the non-linear regression models was used, and the bootstrap approach is itself computationally very expensive, the estimation of each regression model took about half an hour on a current desktop computer. The analyses were computed in Wolfram Mathematica 7.

[1] Note that the maximal possible effect is 2.0. Thus, for cases receiving this combination of interventions, the constants (which are estimated over the entire sample) lead to an over-estimation of the changes. Without the constants, the estimated potential effect is actually smaller than 2.0.

Additional Results

Results of the regression analyses

Tables S-1 compiles the results of the regression analyses, based on which the intervention effects were predicted. Besides the parameter estimates and some confidence statistics (particularly the bootstrapped confidence intervals), the number of cases with the value 1 in the different intervention dummy-variables is presented. Critically low numbers are found for implementation intentions (counts of 1 between 4 and 8). Thus, the parameter estimates and particularly the confidence intervals for this effect might not be estimated adequately. Based on these estimates, forecasts of the intervention effects were calculated as explained earlier.

Prediction of intervention effects

Table 4 in the paper presents the prediction of intervention effects adjusted for control cases without intervention. Table S-2 compiles the same results including the changes over time without interventions. Note that these changes are different for Time Step 1 and 2. Here the changes in Time Step 1 are used for predictions of one time step and the first time step for predictions of two time steps. The constant for Time Step 2 was used for the second time step of predictions over two time steps.

Table S-3 compares the effect of the different interventions on behaviour. This table provides information about the significance of differences in effects observed in Table 4 and S-2 respectively. The same information for the other variables can be obtained on request from the corresponding author.

References

Borsboom, D., Mellenbergh, G. J., & van Heerden J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.

Storn, R., & Price, K.V. (1997) Differential Evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11*, 341–359.

Table S-1. Results of bounded-linear regressions on the change of the respective variable. B = nominal value of estimated parameters; LL / UL = lower / upper limit of the 95% confidence interval; p = error probability; Freq. 1 = frequency of value 1 for the variable of the respective parameter. RW = raw water; N/A = not estimated due to numerical problems. s of e = standard error of the estimate; MAE = mean absolute error, RMSE = root mean squared error, $|e|\leq0.25$ / $|e|>0.5$ = percentage of absolute errors smaller than 0.25 / greater than 0.5.

Parameter for intervention variable		**Behaviour**	**Remem-bering**	**Behaviour control**	**Affective attitude RW**	**Affective attitude SODIS**	**Instrumental attitude RW**	**Instrumental attitude SODIS**	**Injunctive norm RW**	**Injunctive norm SODIS**	**Descr. Norm**
Constant for time step 1	**UL**	0.329	0.094	0.328	0.023	0.093	-0.019	0.073	0.065	-0.040	0.465
	B	0.237	0.001	0.198	-0.046	0.016	-0.123	0.038	-0.014	-0.165	0.356
	LL	0.146	-0.097	0.072	-0.119	-0.056	-0.226	0.002	-0.092	-0.285	0.242
	p	<0.001	0.689	0.001	0.208	0.679	0.009	0.062	0.736	0.009	<0.001
	Freq. 1	167	69	50	170	66	174	60	181	75	73
Constant for time step 2	**UL**	0.418	0.054	0.168	-0.105	0.192	-0.306	0.086	0.101	0.120	0.043
	B	0.289	-0.031	0.062	-0.201	0.132	-0.466	0.048	-0.002	-0.013	-0.067
	LL	0.153	-0.094	-0.037	-0.305	0.062	-0.660	0.010	-0.112	-0.153	-0.186
	p	<0.001	0.266	0.215	<0.001	<0.001	<0.001	0.011	0.964	0.829	0.274
	Freq. 1	138	124	92	134	120	139	93	151	136	127
Current prompt	**UL**	0.578	0.717	0.168	-0.021	0.343	-0.006	0.744	0.091	0.466	0.538
	B	0.395	0.601	-0.025	-0.170	0.166	-0.264	0.664	-0.044	0.208	0.325
	LL	0.211	0.494	-0.200	-0.316	-0.002	-0.544	0.261	-0.184	-0.031	0.112
	p	<0.001	<0.001	0.771	0.026	0.009	0.012	<0.001	0.572	0.052	0.001
	Freq. 1	79	62	41	77	53	78	32	84	59	62
Current public self-commitment	**UL**	0.399	0.987	0.013	0.039	0.290	-0.047	0.057	0.056	0.572	0.176
	B	0.249	0.808	-0.114	-0.066	0.171	-0.233	0.010	-0.047	0.402	0.031
	LL	0.088	0.567	-0.256	-0.172	0.022	-0.427	-0.039	-0.154	0.230	-0.124
	p	<0.001	<0.001	0.093	0.219	<0.001	0.003	0.692	0.434	<0.001	0.683
	Freq. 1	113	80	66	110	80	112	59	114	91	83
Prompt with public self-commitment	**UL**	0.723	-0.436	0.716	-0.049	-0.092	-0.175	N/A	-0.210	-0.272	0.190
	B	0.342	-0.587	0.449	-0.343	-0.316	-0.544	N/A	-0.479	-0.591	-0.104
	LL	-0.043	-0.755	0.164	-0.629	-0.536	-0.897	N/A	-0.754	-0.917	-0.394
	p	0.133	<0.001	0.001	0.007	<0.001	0.010	N/A	0.001	<0.001	0.454
	Freq. 1	20	18	12	20	16	16	N/A	16	17	17

Prompt with implementation intention	**UL**	1.140	-0.295	-0.082	0.337	0.035	0.387	N/A	0.085	0.112	0.173
	B	0.488	-0.480	-0.510	0.135	-0.165	-0.033	N/A	-0.165	-0.412	-0.288
	LL	0.064	-0.706	-0.981	-0.092	-0.379	-0.594	N/A	-0.447	-0.815	-0.804
	p	0.016	<0.001	0.003	0.399	0.143	0.991	N/A	0.321	0.069	0.138
	Freq. 1	7	5	4	7	5	8	N/A	8	4	5
Prompt in previous time step	**UL**	-0.199	0.200	-0.040	0.714	0.039	0.823	-0.002	0.545	0.600	0.200
	B	-0.516	0.076	-0.248	0.378	-0.074	0.387	-0.107	0.323	0.354	-0.003
	LL	-0.860	-0.068	-0.500	0.141	-0.191	-0.030	-0.201	0.116	0.084	-0.207
	p	<0.001	0.219	0.026	0.004	0.270	0.011	0.006	0.013	0.032	0.983
	Freq. 1	29	30	22	29	26	32	28	34	29	30
Public self-commitment in previous time step	**UL**	-0.007	0.190	-0.164	0.130	-0.076	0.789	N/A	0.653	0.064	0.348
	B	-0.353	0.071	-0.335	-0.071	-0.148	0.394	N/A	0.379	-0.169	0.038
	LL	-0.760	-0.045	-0.550	-0.314	-0.218	-0.105	N/A	0.115	-0.463	-0.258
	p	0.035	0.387	0.045	0.581	0.128	0.015	N/A	0.006	0.339	0.827
	Freq. 1	50	41	36	49	42	50	N/A	47	47	44
Interaction current and previous prompt	**UL**	1.176	0.237	0.763	-0.092	0.082	0.121	0.815	0.008	-0.083	0.238
	B	0.706	-0.110	0.396	-0.455	-0.117	-0.593	0.405	-0.307	-0.576	-0.095
	LL	0.235	-0.477	0.045	-0.835	-0.329	-1.372	0.274	-0.637	-0.990	-0.427
	p	0.001	0.491	0.018	0.006	0.299	0.025	0.028	0.071	0.006	0.571
	Freq. 1	19	19	14	19	16	20	18	22	18	20
Interaction current and previous public self-commitment	**UL**	1.459	-0.506	0.395	0.345	0.182	0.772	0.045	0.043	0.285	0.434
	B	0.936	-0.769	0.155	0.034	0.023	0.222	-0.016	-0.270	-0.058	0.083
	LL	0.449	-0.980	-0.075	-0.268	-0.114	-0.298	-0.083	-0.588	-0.355	-0.268
	p	<0.001	<0.001	0.375	0.818	0.855	0.285	0.631	0.090	0.770	0.648
	Freq. 1	43	35	33	39	37	40	24	37	41	38
	R^2	61%	59%	29%	38%	28%	34%	63%	12%	16%	39%
	adj. R^2	60%	57%	23%	36%	24%	32%	61%	9%	12%	36%
	s of e	0.406	0.161	0.269	0.361	0.202	0.448	0.105	0.409	0.401	0.370
	MAE	0.347	0.144	0.220	0.300	0.171	0.371	0.094	0.297	0.319	0.312
	RMSE	0.409	0.178	0.269	0.361	0.202	0.453	0.108	0.408	0.401	0.369
	$\|e\| \leq 0.25$	46%	94%	62%	47%	78%	48%	100%	53%	52%	44%
	$\|e\| > 0.50$	26%	0%	1%	18%	0%	26%	0%	28%	28%	20%
	n	305	193	142	304	186	313	153	332	211	200

Table S-2. Prediction of intervention-effects compared to 0 (i.e., not adjusted for control cases without intervention). LL / UL = lower / upper limit of the 95% confidence intervals; RW = raw water; N/A = not estimated due to numerical problems.

	Scenario		Behaviour	Rememb.	Feasibility	Affective attitude RW	Affective attitude SODIS	Instrumental attitude RW	Instrumental attitude SODIS	Injunctive norm RW	Injunctive norm SODIS	Descr. Norm
Scenarios for first time-step	No intervention (comparison basis)	UL	0.329	0.094	0.328	0.023	0.093	-0.019	0.073	0.065	-0.040	0.465
		Mean	0.237	0.001	0.198	-0.046	0.016	-0.123	0.038	-0.014	-0.165	0.356
		LL	0.146	-0.097	0.072	-0.119	-0.056	-0.226	0.002	-0.092	-0.285	0.242
	Prompt	UL	0.810	0.727	0.348	-0.085	0.388	-0.129	0.770	0.070	0.300	0.888
		Mean	0.632	0.603	0.172	-0.216	0.182	-0.387	0.702	-0.057	0.043	0.681
		LL	0.472	0.449	-0.010	-0.355	0.014	-0.648	0.302	-0.181	-0.167	0.482
	Public self-commitment	UL	0.619	0.993	0.180	-0.022	0.302	-0.171	0.088	0.021	0.368	0.505
		Mean	0.485	0.810	0.084	-0.112	0.187	-0.356	0.048	-0.061	0.237	0.387
		LL	0.347	0.578	-0.005	-0.199	0.061	-0.529	0.006	-0.147	0.084	0.259
	Prompt with pub. self-commitment	UL	1.527	1.043	0.709	-0.379	0.207	-0.850	N/A	-0.332	0.089	0.867
		Mean	1.222	0.824	0.507	-0.625	0.037	-1.164	N/A	-0.583	-0.147	0.608
		LL	0.869	0.562	0.288	-0.888	-0.121	-1.471	N/A	-0.821	-0.390	0.358
	Prompt with im-plementation intention	UL	1.750	0.253	0.067	0.083	0.125	-0.071	N/A	0.000	0.083	0.833
		Mean	1.119	0.123	-0.337	-0.080	0.017	-0.420	N/A	-0.222	-0.369	0.393
		LL	0.750	-0.083	-0.833	-0.286	-0.083	-1.000	N/A	-0.469	-0.688	0.000
Scenarios for second time-step	No intervention (comparison basis)	UL	0.705	0.113	0.467	-0.110	0.257	-0.372	0.147	0.142	0.026	0.478
		Mean	0.525	-0.030	0.259	-0.247	0.148	-0.589	0.086	-0.016	-0.178	0.289
		LL	0.317	-0.163	0.060	-0.389	0.035	-0.845	0.025	-0.176	-0.387	0.087
	Nothing after Prompt	UL	0.741	0.813	0.234	0.285	0.446	-0.054	0.782	0.490	0.730	0.877
		Mean	0.405	0.648	-0.014	-0.038	0.239	-0.466	0.643	0.263	0.384	0.611
		LL	0.043	0.469	-0.292	-0.290	0.056	-0.956	0.225	0.034	0.068	0.333
	Nothing after public self-commitment	UL	0.777	1.051	-0.023	-0.172	0.283	-0.022	0.141	0.594	0.298	0.663
		Mean	0.421	0.850	-0.189	-0.384	0.171	-0.428	0.096	0.316	0.054	0.358
		LL	0.011	0.606	-0.409	-0.625	0.047	-0.909	0.050	0.043	-0.261	0.058
	Repeated Prompt	UL	1.837	1.395	0.626	-0.430	0.498	-0.748	1.831	0.164	0.432	1.103
		Mean	1.506	1.140	0.357	-0.663	0.289	-1.323	1.713	-0.087	0.015	0.841
		LL	1.170	0.757	0.103	-0.910	0.105	-2.025	1.304	-0.330	-0.340	0.571
	Repeated public self-commitment	UL	1.861	1.082	-0.003	-0.205	0.528	-0.140	0.152	0.182	0.692	0.657
		Mean	1.606	0.889	-0.148	-0.415	0.365	-0.439	0.090	-0.001	0.398	0.473
		LL	1.245	0.645	-0.280	-0.626	0.200	-0.801	0.027	-0.185	0.167	0.279
	Prompt with pub. com. after pub. com.	UL	2.745	1.108	0.469	-0.709	0.361	-0.955	N/A	-0.293	0.258	0.890
		Mean	2.343	0.904	0.275	-0.928	0.215	-1.247	N/A	-0.524	0.015	0.694
		LL	1.957	0.648	0.087	-1.154	0.085	-1.457	N/A	-0.729	-0.230	0.494

Table S-3. Prediction of differences of various intervention effects on behaviour change. The interventions in the rows are compared to the interventions in the columns. Thus, if the values are positive, the intervention to the left has a stronger effect than the intervention mentioned above. LL / UL = lower / upper limit of the 95% confidence interval.

		Estimate	1.	2.	3.	4.	5.	6.	7.
Scenarios for 1st time-step	**1. Prompt**	**UL**	0.000	0.358	-0.226	-0.066	-0.280	-0.320	-0.984
		Mean	0.000	0.146	-0.591	-0.488	-0.586	-0.686	-1.423
		LL	0.000	-0.053	-0.937	-1.141	-0.901	-1.014	-1.893
	2. Public self-commitment	**UL**	0.050	0.000	-0.395	-0.241	-0.368	-0.475	-1.159
		Mean	-0.146	0.000	-0.737	-0.634	-0.732	-0.832	-1.569
		LL	-0.364	0.000	-1.068	-1.292	-1.117	-1.087	-2.037
	3. Prompt with public self-commitment	**UL**	0.935	1.067	0.000	0.595	0.474	0.393	-0.475
		Mean	0.591	0.737	0.000	0.103	0.005	-0.095	-0.832
		LL	0.224	0.391	0.000	-0.587	-0.466	-0.526	-1.087
	4. Prompt with implementation intention	**UL**	1.140	1.276	0.585	0.000	0.604	0.466	-0.198
		Mean	0.488	0.634	-0.103	0.000	-0.098	-0.198	-0.935
		LL	0.064	0.238	-0.596	0.000	-0.643	-0.686	-1.546
Scenarios for 2nd time-step	**5. Prompt after Prompt**	**UL**	0.901	1.114	0.462	0.642	0.000	0.389	-0.329
		Mean	0.586	0.732	-0.005	0.098	0.000	-0.100	-0.837
		LL	0.269	0.362	-0.476	-0.605	0.000	-0.506	-1.407
	6. Public self-commitment after public self-commitment	**UL**	1.011	1.087	0.521	0.678	0.505	0.000	-0.395
		Mean	0.686	0.832	0.095	0.198	0.100	0.000	-0.737
		LL	0.319	0.472	-0.394	-0.466	-0.391	0.000	-1.068
	7. Prompt with public self-commitment after public self-commitment	**UL**	1.892	2.035	1.087	1.538	1.397	1.067	0.000
		Mean	1.423	1.569	0.832	0.935	0.837	0.737	0.000
		LL	0.981	1.153	0.472	0.196	0.321	0.391	0.000