

App-based food Go/No-Go training: User engagement and dietary intake in an opportunistic observational study

Matthias Burkard Aulbach^{a,*}, Keegan Knittle^a, Samantha Barbara van Beurden^b, Ari Haukkala^{a,c}, Natalia S. Lawrence^d

^a Faculty of Social Sciences, Department of Social Research, University of Helsinki, Unioninkatu 37, 00014, Helsinki, Finland

^b College of Medicine and Health, University of Exeter, Barrack Road, EX2 5DW, Exeter, United Kingdom

^c Helsinki Collegium for Advanced Studies, University of Helsinki, Fabianinkatu 24, 00014, Helsinki, Finland

^d School of Psychology, College of Life and Environmental Sciences, University of Exeter, EX4 4QG, Exeter, United Kingdom

ARTICLE INFO

Keywords:

Cognitive training
Eating behavior
mHealth
Obesity
Behavior change
Response inhibition

ABSTRACT

Food Go/No-Go training aims to alter implicit food biases by creating associations between perceiving unhealthy foods and withholding a dominant response. Asking participants to repeatedly inhibit an impulse to approach unhealthy foods can decrease unhealthy food intake in laboratory settings. Less is known about how people engage with app-based Go/No-Go training in real-world settings and how this might relate to dietary outcomes. This pragmatic observational study investigated associations between the number of completed app-based food Go/No-Go training trials and changes in food intake (Food Frequency Questionnaire; FFQ) for different healthy and unhealthy food categories from baseline to one-month follow-up. In total, 1234 participants ($m_{\text{BMI}} = 29 \text{ kg/m}^2$, $m_{\text{age}} = 43 \text{ years}$, 69% female) downloaded the FoodT app and completed food-Go/No-Go training at their own discretion (mean number of completed sessions = 10.7, $sd = 10.3$, range: 1–122). In pre-registered analyses, random-intercept linear models predicting intake of different foods, and controlled for baseline consumption, BMI, age, sex, smoking, metabolic syndrome, and dieting status, revealed small, significant associations between the number of completed training trials and reductions in unhealthy food intake ($b = -0.0005$, $CI_{95} = [-0.0007; -0.0003]$) and increases in healthy food intake ($b = 0.0003$, $CI_{95} = [0.0000; 0.0006]$). These relationships varied by food category, and exploratory analyses suggest that more temporally spaced training was associated with greater changes in dietary intake. Taken together, these results imply a positive association between the amount of training completed and beneficial changes in food intake. However, the results of this pragmatic study should be interpreted cautiously, as self-selection biases, motivation and other engagement-related factors that could underlie these associations were not accounted for. Experimental research is needed to rule out these possible confounds and establish causal dose-response relationships between patterns of engagement with food Go/No-Go training and changes in dietary intake.

1. Introduction

Obesity and associated health conditions (e.g., cardiovascular diseases, type 2 diabetes, and certain cancers), have become major concerns worldwide (Global Health Observatory, 2017). The burden of overweight and obesity on healthcare systems and quality of life, calls for effective and efficient interventions. While the underlying reasons leading to gaining excess weight may be multiple and complex, an imbalance between energy intake and expenditure is a key factor (Christiansen & Garby, 2002). Behavioural weight management

interventions focus on changing this imbalance through increasing physical activity and/or reducing calorie intake (Johns et al., 2014).

Researchers, health practitioners, and policy-makers have devised a wide array of interventions aiming to improve diets by generally focussing on reductions in the intake of unhealthy foods (e.g., those high in sugar, fat, and salt) and/or increases in the intake of healthy foods (e.g., fruit and vegetables). Typically, these interventions range from providing information about healthy diets (Hackman & Knowlden, 2014) to placing taxes on food/drink items that are high in added sugars, fat, or are otherwise detrimental to health (Thow et al., 2014). These

* Corresponding author. PL 54 (Unioninkatu 37), 00014, Finland.

E-mail address: Matthias.aulbach@helsinki.fi (M.B. Aulbach).

<https://doi.org/10.1016/j.appet.2021.105315>

Received 9 April 2020; Received in revised form 4 May 2021; Accepted 11 May 2021

Available online 17 May 2021

0195-6663/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

interventions share the assumptions that people (a) make conscious dietary choices, (b) consider their long-term health as the ultimate goal in these decisions, and (c) are able to take into account different pieces of information. However, research has demonstrated that eating behaviour is strongly influenced by unconscious, impulsive psychological processes (Bargh et al., 2012).

Contemporary dual-process models of behaviour regulation, such as the Reflective-Impulsive Model (RIM; Hofmann et al., 2008; Strack & Deutsch, 2004), posit that behaviour is regulated by two interacting, qualitatively distinct systems. While people may take information into account in behavioural decisions, and follow slow, deliberate thinking when they have the time and motivation to do so, they often rely on fast, impulsive processes to make decisions when cognitive resources are low (Hofmann et al., 2008). These impulsive processes rely on associative networks that link, for example, unhealthy food cues with (a) the concept of “tasty”, (b) a motivational orientation to approach and consume those foods, and (c) the activation of behavioural schemata related to eating that trigger motoric responses such as reaching out, grabbing the food, and chewing. The modern food environment with ubiquitous and easily accessible unhealthy food constantly triggers these impulsive processes, making decisions based on nutritional information and long-term health goals less likely. Various ways to support individuals to manage their impulsive processes are available with the aim of facilitating dietary change. This could be done using reflective techniques (e.g., Implementation Intentions) that equip individuals to engage reflective processes to actively override impulsive processes, and novel impulse-focused techniques that alter the impulsive processes that generally lead to unhealthy food choices (van Beurden et al., 2016).

Much research in recent years has led to the development and evaluation of impulse-focused interventions (Sheeran et al., 2016). These interventions commonly target the associations between unhealthy foods and behavioural tendencies to consume them by repeatedly pairing images of unhealthy, yet hard-to-resist foods, with: (a) the withholding of a response, (b) diverting of attention, or (c) enacting avoidance responses (Kakoschke et al., 2017). The food Go/No-Go task in particular has demonstrated the most reliable effects on change in food intake (Allom et al., 2016; Aulbach et al., 2019; Jones et al., 2016). This computerised task presents participants with images of unhealthy foods and control pictures (of healthy foods and/or non-food items). In addition to these pictures, they are presented with a “Go” or a “No-Go” cue (e.g., a green or red border around the picture). Participants are instructed to press a button on the keyboard whenever they see a Go cue and to withhold from pressing the key whenever a No-Go cue appears. The pairing of images and cues is such that unhealthy foods are always paired with No-Go cues and control pictures with Go cues, usually without the participant being aware of this in advance. This repeated pairing of exposure to unhealthy food cues with inaction is thought to disrupt the associative link between the unhealthy, palatable food and the motoric responses and develop an association between those cues and the inhibiting of a motoric response (Chen et al., 2016). However, evidence regarding the underlying mechanisms in Go/No-go training is not conclusive. Some analyses suggest that doing this training task leads to decreased liking of the stimuli that had been consistently paired with cues signalling the inhibition of a response (Aulbach et al., 2019; Chen et al., 2016; Jones et al., 2016). An important aspect of Go/No-Go training is that participants learn the associations between stopping/unhealthy foods and going/healthy foods. Different error rates and reaction times for food and control images are indicators of this learning process (Houben & Jansen, 2015; Jones et al., 2016).

Most studies investigating Go/No-Go training have been conducted in laboratory settings (Aulbach et al., 2019), although a growing number are investigating this task out of the lab, using mobile apps and the internet as task delivery and data collection platforms (Forman et al., 2019; Lawrence et al., 2015; Poppelaars et al., 2018; Stice et al., 2017; Veling et al., 2014). While lab-based experimental studies may have demonstrated the efficacy of the Go/No-Go task in terms of its ability to

change food intake under ideal and controlled circumstances (see above), it is important to move out of the laboratory to assess the effectiveness of the task if used in real-world circumstances, in a way that matters to the individual using the task (Flay, 1986). For example, we do not yet know whether a “minimum dosage” needs to be adhered to in order to attain measurable effects, nor whether frequency or intensity of the training affects the size of the behavioural effects. Most studies apply a certain amount of training and compare trained participants with a control group. Variations in dose are mostly only present *between* studies, not *within* studies (Veling et al., 2013, being one notable exception). To be able to provide evidence-based recommendations or instructions on how to use this training task in daily life, it is crucial that we know how much training is required to achieve meaningful effects and whether users are prepared to do this.

Digital, internet connected devices enable both the delivery of an intervention and measurement of behavioural outcomes in applied real-world contexts (Murray et al., 2016), addressing some of the shortcomings of lab-based studies (see above). Smartphones allow researchers to access large amounts of potential participants from many different social groups (i.e., 60–80% of people in industrialized nations have access to a smartphone; Poushter, 2016, p. 45) and enable regular and flexible access to interventions with relatively little participant burden (Atkinson & Gold, 2002; Schoeppe et al., 2016). Another aspect is that digital interventions are often self-delivered and therefore reliant on the user’s willingness to access and engage with the intervention. Any self-enacted digital intervention therefore requires a realistic assessment of how much potential users might engage with the app (Michie et al., 2017) and such an evaluation is not yet available for food Go/No-Go training.

The Reflective-Impulsive Model (Strack & Deutsch, 2004) postulates that connections in the associative network of the impulsive system are determined by the frequency and recency of their co-activation. It would therefore predict that more frequent pairing of unhealthy food images with inhibiting responses leads to larger changes in eating behaviour. However, contrary to these predictions, Veling et al. (2013) demonstrated that the number of pairings had no effect on food evaluations and food choice in the laboratory. Moreover, several meta-analyses (Allom et al., 2016; Aulbach et al., 2019; Jones et al., 2016) found no significant effect of amount of training across studies on outcomes. However, as outlined above, these studies mostly consisted of single-session lab-based tasks and are likely to differ from real-world applications. In naturalistic real-world settings, usage of Go/No-Go training is likely to differ from person to person and within persons over time. Collecting data on this variability would therefore enable observational examinations of the associations between the amount of naturally occurring app use and changes in behavioral outcomes. This study has two main inter-related aims: the first aim is to describe usage patterns of the food Go/No-Go training application “FoodT” after public release and advertising, to investigate how many participants follow recommendations, and how much app use can be realistically expected in potential future users without offering any incentives. The second aim is to investigate naturally-occurring associations between the amount of Go/No-Go training completed and changes in the intake of healthy and unhealthy foods. While examining these associations will not establish any causal relationships, as app use was not explicitly manipulated and as any observed statistical relationships could be due to other underlying confounding factors (e.g. motivation), preliminary data on these associations are essential in planning future experimental studies to establish dose-response relationships for food Go/No-Go training.

The foods used in the FoodT app differ from each other with regard to characteristics such as taste (sweet vs salty) or the breadth of the food category in which it might fit (for example, pizza as a category is narrower than cake). These characteristics might also influence the effects of training on food intake. Thus, in addition to examining global associations between app use and dietary intake, we also separately examine the observed associations for each included food category. Furthermore,

as implicit bias change can result from training and as real world users may space out or concentrate their training temporally, we also examine the associations of associative learning and training density with changes in dietary outcomes.

2. Methods

This opportunistic study uses data collected from a large trial comparing the effectiveness of web-based and app-based food go/no-go training. The main analyses on overall intervention effectiveness and more detail on trial procedures can be found in a manuscript in preparation by Lawrence et al., (Lawrence et al., n.d.). The present paper presents secondary analyses collected from participants in the app-based arm of the main trial. Data were collected using the FoodT app, which at the time could be used on any device running an Android operating system. Unless stated otherwise, hypotheses and statistical analyses were pre-registered on the Open Science Framework (<https://osf.io/nhq4y/>) using the template for secondary data analysis by Weston et al. (2019). Any divergence from these analyses resulted from the peer review process. The authors responsible for data analysis (MA, KK, AH) did not have access to the data before uploading the pre-registration to the OSF and were not aware of any patterns in the data. Ethical approval was obtained from The School of Psychology Research Ethics Committee at the University of Exeter.

2.1. Procedure and participants

This pragmatic open trial was advertised through popular and social media. Participants could freely download the FoodT app from the Google Play Store. Thus, all included data was collected through Android devices. Upon download, users could consent to contribute their data for research and all users received access to the same intervention content.

At the outset of the study, participants were asked to indicate their age, sex, height, weight, whether they smoked or not, were trying to lose weight, and whether they had a metabolic condition or not. In addition, they filled out a Food Frequency Questionnaire (FFQ; adapted from Churchill & Jessop, 2011; Lawrence et al., 2015). In this questionnaire, participants indicated how often ("4 or more times a day", "2 or 3 times a day", "Once a day", "5 or 6 times a week", "2 to 4 times a week", "Once a week", "1 to 3 times a month", "Less often or never") over the previous month they typically consumed the food categories that were available in the training (i.e., alcohol, biscuits, white bread, cheese, red and processed meat, pizza, cake, chocolate, crisps, fast food, fizzy drinks, and sweets as unhealthy food groups, and fruit, vegetables, and crispbread as healthy food groups). Due to an initial coding error, no data was collected for ice cream and pastries, and data for crisps are available for only about half of the participants. In the main analyses we used the average score across all unhealthy (or healthy) food categories as an outcome.

After completing baseline measures, participants were encouraged to use the FoodT application once a day for the first week and once a week for the rest of the one-month period (i.e. 10 times in total) but were free to use it as much or as little as they liked. See section 2.2 below for further detail on the Go/No-Go training protocol.

At the end of the study period (i.e. at least 27 days after starting app use), participants were asked to complete a follow-up questionnaire, consisting of the same measures completed at baseline. Participants who did not immediately complete the follow-up measures received a reminder and were able to complete the follow-up questionnaire up to 90 days after completing baseline measures.

Participants were only included in data analyses if they indicated a baseline BMI larger than 18.5 (the common threshold for underweight), their age was between 18 and 100 years, and if they did not change their smoking status or whether they had a metabolic disease from baseline to follow-up. We also excluded participants who did not respond to the

second FFQ within 90 days from starting app use and those who did not fill out any FFQ questions.

2.2. Go/No-Go training

The application delivered Go/No-Go training in sessions with each session consisting of three blocks of 32 trials each. Completing one session (96 trials) takes about 4 min. Participants could take breaks between blocks but during completion of the blocks, images appeared at a set speed.

One trial consisted of an image appearing on a random location on the screen closely followed (100 ms later) by a green (Go-cue) or red (No-Go cue) circle around the image. Each image appeared for 1500 ms with a 500 ms interstimulus interval. Participants were instructed to tap images with a green circle and to refrain from tapping images with red circles. Correct taps were rewarded with a point and incorrect taps (on no-go images) lost a point. Images depicted unhealthy or healthy foods or non-food objects such as flowers or pieces of clothing.

Of the 32 trials in a block, 8 trials paired an unhealthy food with a No-Go cue, 8 paired a healthy food with a Go-cue, 8 paired a control image with a Go cue and 8 paired the same control images with a No-Go cue. Control trials were included to keep the task challenging and to facilitate associative learning with regards to the more consistent pairing of unhealthy food images and No-Go cues. Fig. 1, panel A depicts the structure of the sessions and blocks and panel B depicts examples for each of the possible trial types.

By default, the 8 unhealthy No-Go trials within each block included two pictures of biscuits, one of chocolate cake, two of chocolates, and three of potato crisps (the same exemplars were used throughout the task, based on Lawrence et al., 2015). Users could alternatively choose to personalize the application by selecting up to three of the following unhealthy food categories: alcohol (including beer, wine, and cocktails), biscuits, (white) bread, cake, cheese, chocolate, crisps, fast food, fizzy drinks, ice cream, (red and processed) meat, pastries, pizza, and sweets. If a participant had chosen only one category, then they would receive 8 trials of that category within all 3 blocks (i.e. 24 per session made up of the same 8 unhealthy exemplars presented 3 times each). If a participant had chosen two or three categories, then they would undertake 8 trials of each chosen category within that session, all within the same block. As an example, if a participant had chosen two personalized categories (e.g. alcohol and cheese), one block would contain all 8 alcohol/no-go trials, and another block would contain all 8 cheese-no-go trials. The third block would then contain the 8 default unhealthy No-Go trials. If a participant had chosen three personalized categories (e.g. alcohol, cheese, and fizzy drinks), then one block would contain all 8 alcohol/no-go trials, another block would contain all 8 cheese/no-go trials, and the third block would contain all 8 fizzy drink/no-go trials. Participants were free to choose new personalized categories between sessions.

The filler and healthy food images consisted of three different sets that were always presented together in one block (clothing, flowers, and stationery for fillers and different images of healthy foods). These sets were randomly allocated to blocks. For example, block one could show a set of clothing or flowers or stationery, block two any of the remaining two sets, and block three the remaining set. For healthy foods, one block presented three images of fruit, four images of vegetables, and one of crispbread. Different exemplars of these healthy foods were presented in each block (i.e. 24 different healthy foods in total).

The training data was recorded and sent to the central database on a University of Exeter secure server. Training data consisted of the images displayed, trial type (Go/No-Go), whether the participant showed the correct response, and response time.

2.3. Statistical analyses

We present descriptive statistics for the patterns of usage including

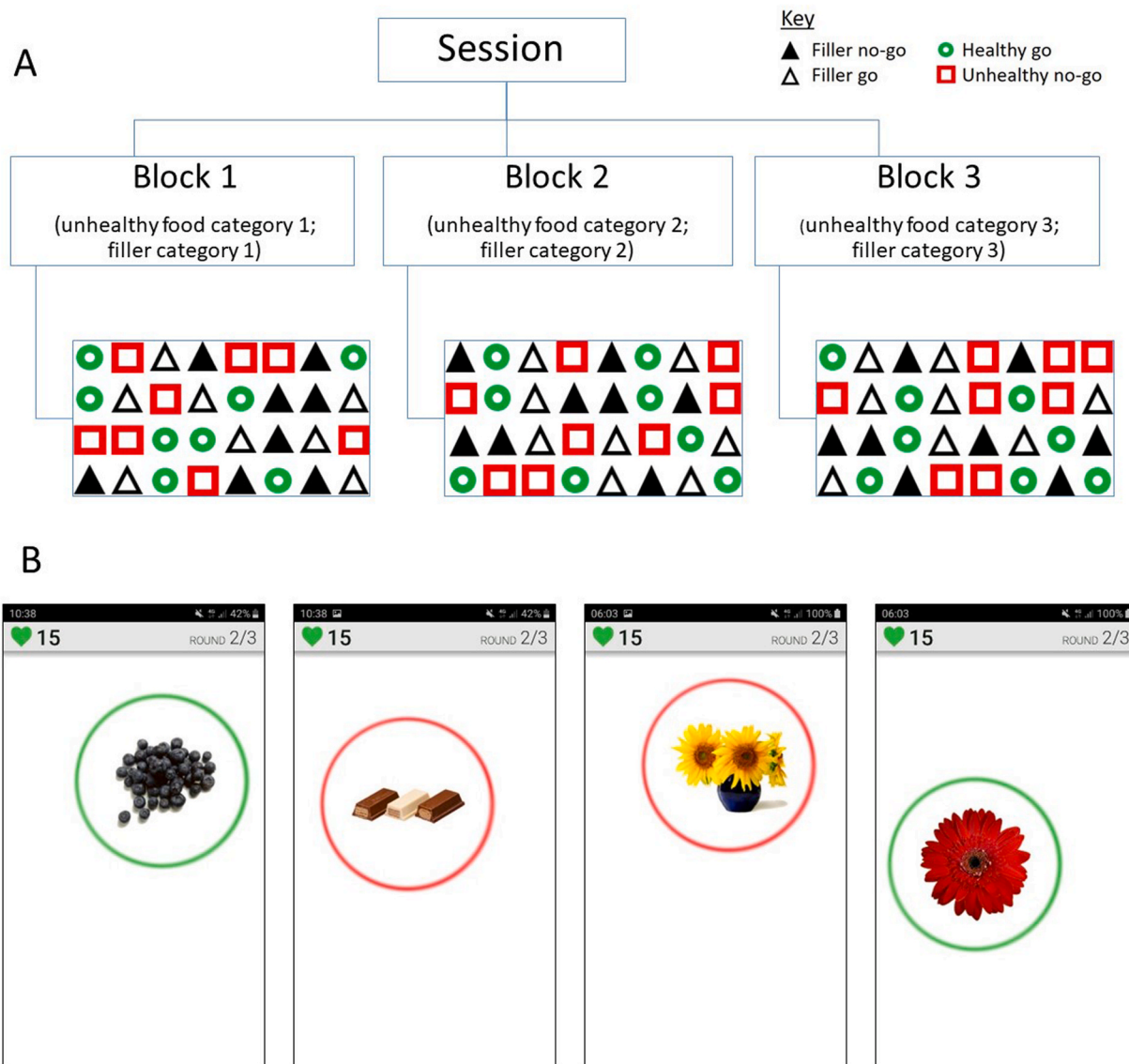


Fig. 1. Task structure. Panel A shows that one session consists of three blocks and each block of 32 trials includes 8 healthy-Go (represented here by green circles), 8 unhealthy No-Go (red rectangles), 8 filler Go (white triangles), and 8 filler No-Go (black triangles). Each block presents images from one chosen food category and one randomly assigned filler category (flowers, stationery, or pieces of clothing). Panel B depicts the four different kinds of trials as they appear in FoodT (from left to right): Healthy-Go, Unhealthy-No-Go, Filler-No-Go, Filler-Go. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

total amount of training conducted, the time until participants stopped using the application, and the distribution of training across food categories.

All models testing the associations between app use and changes in food intake were random intercept linear mixed models with baseline BMI, age, sex, smoking and dieting status, and presence of a metabolic condition entered as covariates. This means that we added all main effects for these variables and the interaction term between timepoint (baseline vs follow-up) and the respective predictor for each hypothesis as the main predictor of interest.¹ The only random effect included in the model was participants' intercept. P-values are based on the Satterthwaite approximation for degrees of freedom which show acceptable levels of error rates (Luke, 2017).

¹ The formula for those models in the lme4 package in R was: intake ~ number of trials*timepoint + BMI + age + sex + smoking status + dieting status + metabolic condition + (1|ID).

All analyses were performed in R software (R Core Team, 2017) using the packages *dplyr* (Wickham et al., 2020), *tidyr* (Wickham & Henry, 2020), *lme4* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017), *openxlsx* (Walker, 2018), *ggplot2* (Wickham, 2016), *stargazer* (Hlavac, 2018), *statsr* (Rundel et al., 2018), *lsr* (Navarro, 2015), *pwr* (Champely, 2020), and *apaTables* (Stanley, 2018).

2.3.1. Hypothesis 1: overall association between app use and consumption changes

To estimate the descriptive association between the amount of training and reductions in unhealthy food intake, we estimated the regression weight of the interaction term between time and the number of unhealthy No-Go trials conducted after control variables are accounted for with the mean score of unhealthy food intake as the outcome. We then ran the same model with healthy food intake as the outcome and the interaction term between time and the number of healthy Go trials as the predictor.

2.3.2. Hypothesis 2: differences between different foods

Hypothesis 2 assumes that the associations from Hypothesis 1 differ by food category. We thus conduct the same analyses as for Hypothesis 1 separately for the different food categories, using the number of trials of the respective categories as the predictor.²

2.3.3. Hypothesis 3: measures of associative learning

To assess the effect of learning associations between No-Go and unhealthy foods, we computed an *error learning index* by subtracting the error rates on unhealthy No-Go trials from the error rates on filler No-Go trials. Someone who shows relatively fewer errors on no-go trials to unhealthy foods than fillers (as a result of learning the food-no-go associations) will show a larger learning index. This index was then used as a predictor in a random intercept linear mixed model predicting unhealthy food intake.

Similarly we calculated a measure of learning the association between healthy foods and showing a go reaction. This *reaction time learning index* consisted of the difference between reaction times to healthy foods (100% predictive of a Go signal) and reaction times to filler trials (not predictive of a Go signal). The index then functioned as the main predictor in the linear mixed model predicting healthy food intake.

2.3.4. Sensitivity analyses

We ran all the analyses described above again for single food categories removing participants who: (1) reported never eating food from an unhealthy food category or chose the highest value for healthy foods at baseline, as these individuals had no potential to reduce/increase their intake, and/or (2) never trained to a given category. Including these participants in the analysis could obscure real intervention effects for individuals who do have room for improvement.

As described above, participants could personalize FoodT by selecting the unhealthy food categories that they wanted to train to. To explore effects of using personalized training, we used a binary variable indicating whether a participant used the personalization feature at least once and conducted the main analyses separately for those who used the personalization feature and those who did not.

To investigate potentially different effects for different user groups, we also analysed the data separately for participants with normal weight (BMI smaller than 25), overweight (BMI between 25 and 30) and obesity (BMI larger than 30) as well as for dieters and non-dieters. We also re-analysed the data removing participants who reported a metabolic condition as they might have substantially different consumption patterns than those without a metabolic condition.

2.3.5. Exploratory (not pre-registered) analyses

The Food Frequency Questionnaire used in this study (adapted from Churchill & Jessop, 2011; Lawrence et al., 2015) is not linear in the sense that the difference between two categories is not equal across the range of the scale. In other words, the meaning of a one-point change in the FFQ score differs depending on where someone starts. A shift in FFQ score from 0 to 1 would represent a change of half a serving per week, whereas a shift from 7 to 8 would represent a change of up to 10.5 servings per week. To ameliorate this issue, we re-coded the FFQ scores into weekly servings as follows (Mikkilä et al., 2015): “4 or more times a day” = 28 servings per week, “2 or 3 times a day” = 17.5 servings per week, “Once a day” = 7 servings per week, “5 or 6 times a week” = 5.5 servings per week, “2 to 4 times a week” = 3 servings per week, “Once a week” = 1 serving per week, “1 to 3 times a month” = 0.5 servings per week, “Less often or never” = 0 servings per week. We conducted all analyses again using this measure of weekly servings instead of raw FFQ scores.

² We also pre-registered a multivariate regression model but found this to not be feasible due to issues of multicollinearity.

To quantify how participants distributed their training over time, we calculated a measure of training density by dividing the number of trials conducted on the most active day by the total number of trials. This measure is high when participants conduct most of their total training on one day and lowest when training is spread out evenly over all training days. We then entered this index as a predictor in the same random intercept model as used above.

Since training effects might wash out over time, we tested the effect of the time lag between the last training session and filling out the FFQ for the second time by using it as a predictor in the same kind of mixed model as in the other analyses.

3. Results

3.1. Sample characteristics and power considerations

1234 participants (857 women, 377 men) contributed sufficient data according to our pre-specified criteria and met our other inclusion criteria (<https://osf.io/nhq4y/>). Using an alpha level of 0.05, this sample provides statistical power of .80 to detect a Cohen's *d* of 0.08 (a small effect) for paired *t*-tests examining changes in weight or food intake. To detect a Cohen's *d* of 0.15 in these comparisons, power is greater than 99%. For regression effects, this sample provides 80% power to detect an f^2 effect size of 0.0064, and over 99% power to detect an f^2 of 0.015. More detailed considerations of statistical power can be found in the supplementary materials (<https://osf.io/yksgp/>).

Table 1 gives an overview of the sample characteristics. The average number of completed sessions was 10.7, in line with the recommendation provided in the app (10 sessions). The number of completed sessions varied widely between participants (range 1–122; *sd* = 10.3; *median* = 8) with strong positive skew in the distribution. See Fig. 2, panel A for the distribution of conducted trials for each food category and panels B–D for the distribution of total sessions and their drop-off over time. Participants who reported they were currently dieting to lose weight did not use the app significantly more than those who did not report dieting (*m* (dieters) = 297, *sd* (dieters) = 297; *m* (non-dieters) = 271, *sd* (non-dieters) = 264, $F(1, 1227) = 2.679, p = .10, d = 0.09$).

On average, participants reduced their mean unhealthy intake score on the FFQ by 0.35 points ($t(1191) = 12.91, p < .0001, d = 0.37$) and their body weight by 556 g ($t(746) = 5.55, p < .0001, d = 0.20$). The average reduction of intake for the unhealthy food categories ranged from 0.12 (pizza) to 0.59 (sweets) points and the average increase of fruit and vegetable intake was 0.24 and 0.18 points, respectively. More information about overall intervention effects are reported in the main paper (Lawrence et al., n.d.). Fig. 3 shows FFQ scores by food category and divided by dieting status. Dieting status did not have a substantial

Table 1
Summary descriptive statistics for the sample.

Variable	Mean/ Percentage	SD	Min	Max
Age	43.03	13.98	18	92
BMI baseline	28.98	6.15	18.55	63.69
BMI follow-up	28.62	6	14.99	63.29
Weight baseline	81.82	19.05	44	159
Weight follow-up	81.29	18.54	44	160
No-Go error rate unhealthy foods	0.01	0.02	0	0.19
No-Go error rate fillers	0.02	0.03	0	0.29
Go-reaction time healthy foods	745.77	86.5	563.59	1128.92
Go-reaction time fillers	762.75	87.57	557.41	1152.01
Percent female	70.02	–	–	–
Percent dieters	46.14	–	–	–
Percent with metabolic condition	14.48	–	–	–
Percent smokers	7.66	–	–	–
Percent used personalization	50	–	–	–

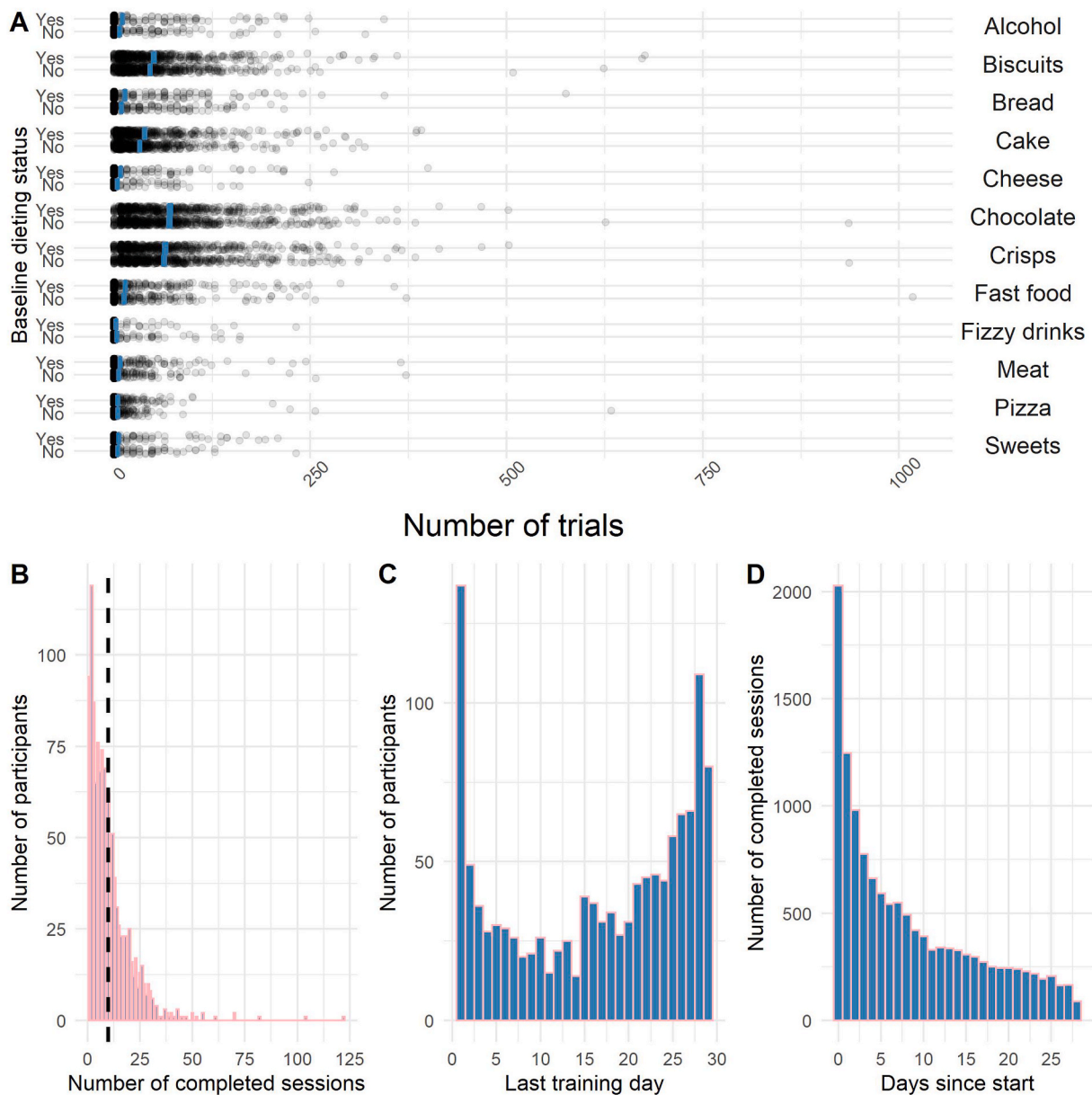


Fig. 2. Visualizations of app usage. Panel A shows the number of conducted trials for each unhealthy food category with grey dots representing single participants and the blue bars representing mean values, separated by baseline dieting status. Panel B shows the distribution of the total amount of training sessions conducted by participants with the dashed line indicating the amount of training participants were encouraged to conduct (10 sessions). Panel C shows the distribution of the last active day: it is obvious that many participants stopped using the app after the first day but there is another peak (towards the end of the study period). Panel D shows the total number of completed sessions by day since starting use and shows a sharp decrease in the first few days. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

influence on changes in food intake across food categories ($t(1185) = -0.48, p = .63, d = 0.03$).

The below reports the results from the main analyses related to our hypotheses and the exploratory and sensitivity analyses outlined in the methods section. Further visualizations of and information about the data are available in the supplementary materials (<https://osf.io/yksgp/>).

3.2. Hypothesis 1: healthy and unhealthy food intake

The random-intercept model for unhealthy food intake delivered a regression coefficient for the critical interaction between timepoint and the number of trials of $b = -0.0005, CI_{95} = [-0.0007; -0.0003], t(876.07) = -4.28, p < .001$. This shows that in this sample, performing

one trial was associated with a 0.0005 drop in the mean FFQ score across unhealthy foods. Fig. 4 illustrates these results.

Extrapolating this point estimate, a one-point decrease on the mean score of unhealthy food intake (i.e., across all unhealthy foods whether trained or not) would be associated with completing 2090 trials (equal to roughly 87 training sessions). Considering there were 12 unhealthy food categories in the application, a one-point change in one category leads to a change of 0.083 points in the overall mean unhealthy score. Based on the regression weight, attaining such a change would be associated with performing roughly 174 trials (roughly 7 sessions). Note however, that this calculation depends on how many food categories were actually filled out and the average participant responded to 11 of the twelve categories.

The analogous model for healthy food intake indicated a regression

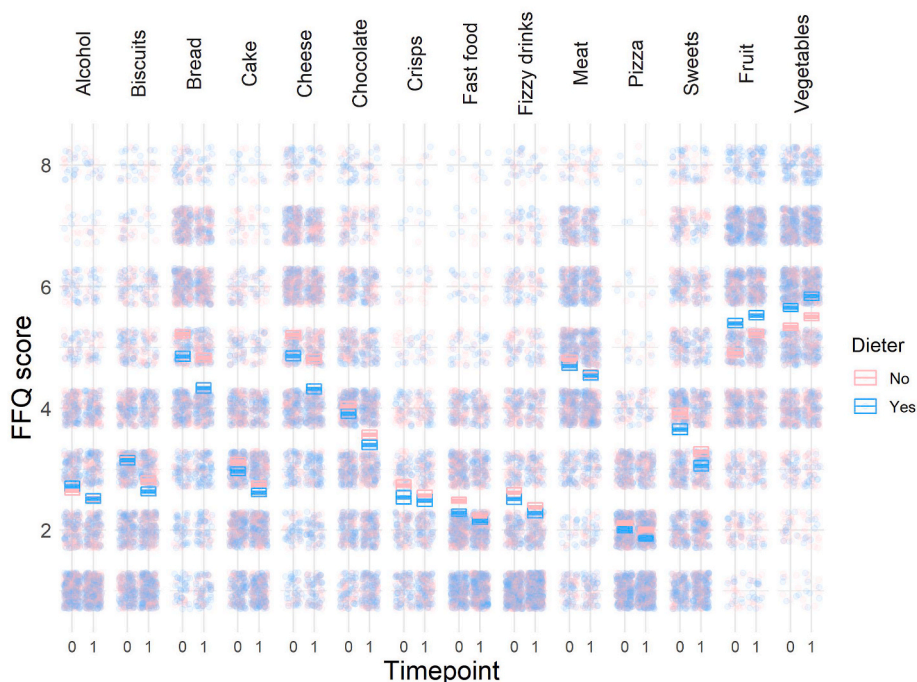


Fig. 3. FFQ scores with means and standard errors by food category and timepoint. Blue dots represent dieters and pink dots non-dieters. Timepoint: 0 = baseline and 1 = follow-up. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

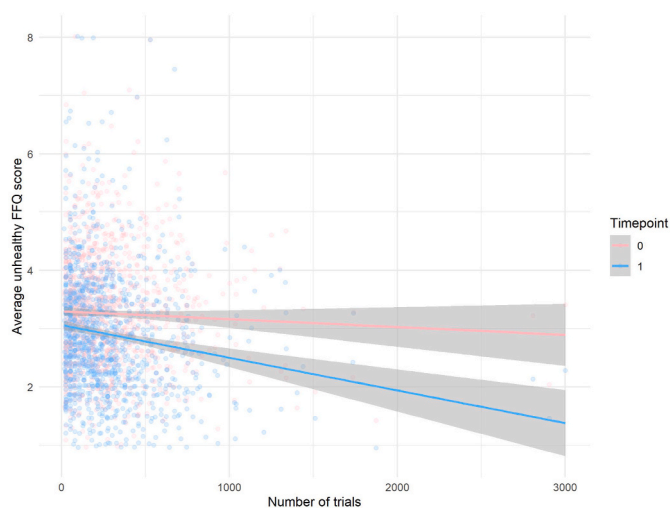


Fig. 4. Illustration of the association observed between amount of no-go training and change in unhealthy food intake. The x-axis represents the total amount of conducted training trials, the y-axis the mean unhealthy FFQ score. Pink represents data at baseline, blue at follow-up. The difference in slopes of the regression lines at the two timepoints indicates that the amount of training is unrelated to unhealthy food intake at baseline but relates negatively to intake at follow-up. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

weight of $b = 0.0003$, $CI_{95} = [0.0000; 0.0006]$, $t(859.05) = 2.22$, $p = .027$ for the interaction between timepoint and the number of trials completed. This indicates that participants who performed more training increased their intake of healthy foods more than those who performed less training.

3.3. Hypothesis 2: differences between food types

Regression weights for the crucial interaction term between

timepoint and number of completed trials ranged from 0.001 [0.000; 0.002] for fruit to -0.007 [-0.012 ; -0.003] for pizza, indicating that the association between training completion and change in food intake is stronger for some food categories than for others. Fig. 5 shows the regression weights for the interaction term between timepoint and number of trials for the different food categories. The numerical values can be found in Supplementary Table 1.

The regression weights for the main effect of interest from the models predicting weekly servings are reported under Model 2 in Fig. 5. The only categories that showed a regression coefficient with a confidence interval excluding 0 were alcohol, fizzy drinks, pizza, and vegetables.

Model 3 and Model 4 in Fig. 5 (and Supplementary Table 1) show the results of the sensitivity analysis outlined above and only included participants who had room for improvement and trained a given category at least once. The regression weight in Model 3 (using FFQ scores as an outcome) is only significantly different from zero for chocolate, fast food, pizza, and vegetables and in Model 4 (using weekly servings as an outcome) only for vegetables.

3.4. Hypothesis 3: measures of associative learning

Average task performance in terms of No-Go error rates and reaction times is displayed in Table 1. Filler-Go reaction times were significantly slower than healthy food Go reaction times ($t(1233) = 30.02$, $p < .0001$, $d = 0.85$) and error rates to filler No-Go trials were higher than to unhealthy food No-Go trials ($t(1233) = -13.73$, $p < .0001$, $d = 0.39$). These results indicate that participants on average did learn the associations as intended.

The model testing the error learning index ($m = 0.009$; $sd = 0.02$) delivered a regression weight of $b = 1.33$, $CI_{95} = [-1.28; 3.95]$, $t(899.7) = 1.00$, $p = .32$. This indicates that the strength of learning the association between unhealthy foods and inhibiting a response, as measured by this index, is not related to unhealthy food intake reduction.

Testing the effect of the reaction time learning index ($m = 17.4$; $sd = 19.9$), we found a regression weight of $b = -0.001$, $CI_{95} = [-0.004; 0.002]$, $t(859.3) = -0.87$, $p = .38$. This suggests that learning associations between healthy foods and responding is not related to changes in

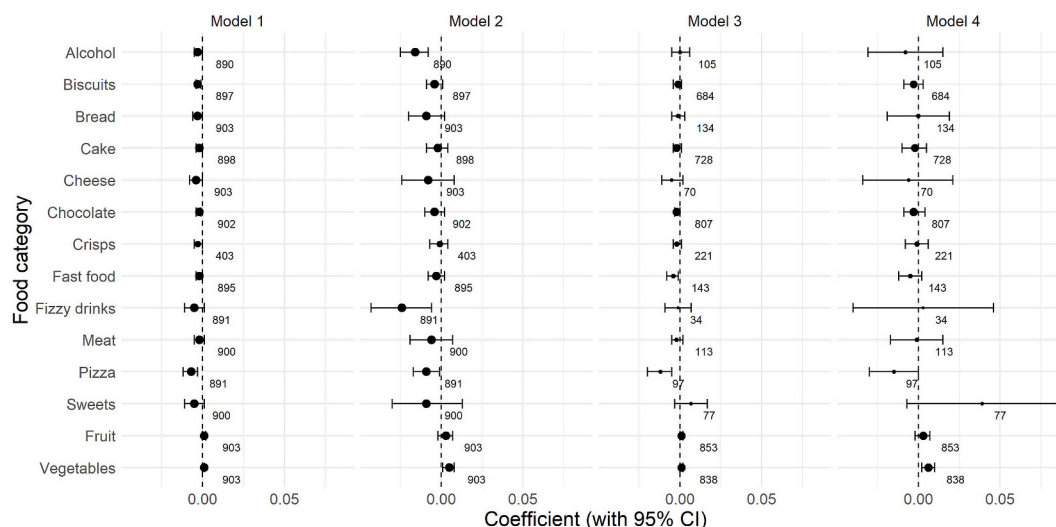


Fig. 5. Illustration of the regression weights with 95% confidence interval bounds for the *timepoint* by *number of trials* interaction term. The size of the points represents the number of participants used in each model which is also indicated by the number below the points. *Model 1* used FFQ scores as an outcome and data from all participants. *Model 2* used weekly servings as an outcome as re-coded from the FFQ scores and also data from all participants. *Model 3* and *Model 4* only use data from participants who (1) trained the respective food category at least once and (2) did not indicate the minimum (for unhealthy foods) or maximum (for healthy foods) at baseline. *Model 3* uses the FFQ scores as an outcome and *Model 4* uses weekly servings as the outcome. All models were controlled for age, sex, baseline BMI, presence of a metabolic condition, and diet and smoking status. The differences in the regression weights between models result from differences in the distribution of the dependent variable (*Model 1* and *Model 3* vs *Model 2* and *Model 4*) and in included participants (*Model 1* and *Model 2* vs *Model 3* and *Model 4*).

healthy food intake.

3.5. Sensitivity analyses

3.5.1. Personalization feature

Participants who used the personalization feature at least once ($n = 617$) conducted more training on average ($m = 1285$ trials, equalling roughly 13 sessions) than those who never used that feature ($n = 617$; $m = 729$ trials, 8 sessions), $F(1, 1232) = 108.76$, $p < .0001$, $d = 0.59$). Regression weights for the interaction between timepoint and training were highly similar for those who personalized the application ($b = -0.00046$, $CI_{95} = [-0.00095; -0.00018]$, $t(450.7) = -3.23$, $p = .001$) and those who did not personalize the application ($b = -0.00057$, $CI_{95} = [-0.00095; -0.00018]$, $t(422.5) = -2.87$, $p = .004$). These largely overlapping confidence intervals indicate that the association between training amount and food intake was similar for both groups.

3.5.2. BMI, dieting status, and metabolic condition

While regression weights for the timepoint \times amount of training interaction were highly similar for the three BMI groups, the respective estimates were only significant for overweight ($n = 333$; $b = -0.0004$, $CI_{95} = [-0.0008; -0.0001]$, $t(312.45) = -2.50$, $p = .013$) and obese participants ($n = 334$; $b = -0.0005$, $CI_{95} = [-0.0009; -0.0002]$, $t(319.45) = -3.10$, $p = .002$) but not for normal-weight participants ($n = 248$; $b = -0.0003$, $CI_{95} = [-0.0008; 0.0002]$, $t(239.69) = -1.11$, $p = .27$).

Dieters and non-dieters showed highly similar associations between training completion and dietary intake change (For dieters: $n = 567$, $b = -0.0005$, $CI_{95} = [-0.0008; -0.0002]$, $t(395.63) = -3.25$, $p = .001$; For non-dieters: $n = 662$, $b = -0.0004$, $CI_{95} = [-0.0008; -0.0001]$, $t(478.75) = -2.78$, $p = .006$).

Removing participants with a metabolic condition ($n = 178$) did not substantially alter the overall results ($b = -0.0004$, $CI_{95} = [-0.0007; -0.0002]$, $t(770.42) = -3.72$, $p = .0002$).

3.6. Exploratory analyses

3.6.1. Training density

Training density (the share of total training conducted on the most

active day) was inversely associated with changes in unhealthy food intake ($b = 0.496$, $CI_{95} = [0.278; 0.714]$, $t(884.92) = 4.46$, $p < .0001$), with more concentrated training associated with smaller decreases in intake. The association between training density and dietary intake remained significant after adding the amount of completed training as an additional factor to the model, ($b = 0.335$, $CI_{95} = [0.076; 0.594]$, $t(882.58) = 2.54$, $p = .01$).

3.6.2. Time lag between end of training and follow-up

The length of delay between the last training day and filling out the FFQ for the second time was positively associated with unhealthy food intake ($b = 0.012$, $CI_{95} = [0.008; 0.016]$, $t(885.79) = 5.84$, $p < .0001$). Adding the number of completed training trials to the model resulted in regression weights of $b = 0.010$, $CI_{95} = [0.005; 0.015]$, $t(881.80) = 4.13$, $p < .0001$ for the delay by time interaction, and $b = -0.0002$, $CI_{95} = [-0.0004; 0.0001]$, $t(871.87) = -1.28$, $p = .20$ for the number of unhealthy No-Go trials by time interaction. This indicates that longer time lag is associated with smaller decreases in unhealthy food intake at follow-up independently of the amount of training.

4. Discussion

The current study presents data from a pragmatic open trial in which 1234 participants conducted app-based food Go/No-Go training at their own individual intensity. Usage frequency varied widely, but a substantial share of participants (42%) conducted at least the ten recommended sessions of training. Participants who used the app more reduced their intake of unhealthy foods more and showed a stronger increase in healthy food intake over the course of a month. Contrary to our predictions, indices of learning unhealthy-No-Go and healthy-Go associations were not predictive of changes in food intake of either category.

In terms of user engagement, 42% of participants in this study followed instructions to conduct at least 10 training sessions (960 trials, based on Lawrence et al., 2015 who delivered 864 trials in four six-block sessions) over the one-month period without receiving any reward for doing so. We do not know how people would respond to more demanding or intense instructions and how much training users would

then be willing to conduct. Critically, a major issue in this study was fast drop-off in app usage (see Fig. 2), a pattern commonly observed not only in studies on m-health but in app usage more generally (Eysenbach, 2005; Perro, 2016). This limits what we can learn from studies such as ours since the amount of app use is always confounded with other factors, presumably mainly motivation. Keeping user engagement high seems particularly important since we found that the delay between the last training session and the second food intake measurement was predictive of changes in food intake, which implies that sustained training is important. Combined with qualitative evidence about perceived reductions of training effects and studies showing diminished effects after a delay (Chen et al., 2019) this indicates that effects of the training wear off over time. Evidence from other studies is more mixed: self-reported weight loss six months after several sessions of Go/No-Go training was larger than after one month (Lawrence et al., 2015) and Schonberg et al. (2014) reported sustained effects on attention two months after a single session of cue-approach training.

In our sample, the average unhealthy FFQ score moved from 3.26 to 2.90 ($M = -0.35$). Reducing intake of one of the twelve assessed unhealthy foods by one point leads to a reduction in the average overall unhealthy score of 0.083 points. Therefore, participants on average reduced intake by about four points distributed across the twelve unhealthy food scales. Lawrence et al. (2015) found a similar decrease in FFQ scores ($M = -0.37$) averaged across four trained foods over a one-month period, in the active training group. In the current sample, participants who achieved a comparable reduction in FFQ scores conducted roughly one training session per day over a one-month period. While this seems like a lot, sessions take less than 5 min to complete and training can be accessed at any time. Moreover, Forman et al. (2019) reported good adherence (89% of prescribed sessions were completed) to 42 daily 10-min training sessions delivered on home computers. Since training on mobile devices is even more convenient, high adherence rates seem realistic. While the observed relationship between amount of training and change in food intake is interesting, it is important to highlight the limitations of the study and data. As an important caveat, the current data were obtained in a single-arm pragmatic study. Therefore, several other factors could contribute to the observed effects. For example, one might argue that the relationship between amount of use and reduced intake could be due to the strength of belief in the intervention's effectiveness causing placebo effects for those using it more, participants' personality traits such as conscientiousness, or socioeconomic variables. Relatedly, we could only use data from users who were motivated to provide it, adding another layer of self-selection. Moreover, motivation for dietary change could potentially be a key confounding variable. Motivation is likely to affect intervention use but also affects dietary change directly (or through other pathways), outside of any potential intervention effects. However, it is interesting to note that in our study participants who indicated they were currently dieting did not use the app more. Additionally, the relation between the amount of app use and changes in food intake did not significantly differ between dieters and non-dieters. Also the positive effect of distributing training over time (regardless of total use) speaks against a pure effect of motivation. However, due to the above constraints, the current findings are simply a first observation that the amount of training is associated with changes in food consumption. As discussed below, further research that randomises matched groups of participants to receive high vs. low amounts of active training is required to build on these results.

In addition to the relation between total training and changes in food intake, our exploratory analyses suggest that spacing out training over time may be more beneficial than concentrating it on one day. Bakkour et al. (2018) found a similar trend, demonstrating in a controlled trial that training on two different days, instead of massed on one day, had a stronger and longer lasting influence on participants' snack choice. Our current results are in line with and expand on this and the large body of learning research about the beneficial effects of spaced learning sessions (Carpenter et al., 2012) by demonstrating a similar effect in a real-world

context. The effect of spaced training might be due to the fact that participants learn associations in different contexts when spacing out training, facilitating the later activation of those associative networks in a variety of contexts (Strack & Deutsch, 2004). However, this hypothesis is speculative and would require further investigation. Another potential explanation (also described above) is that learned associations decay over time and need constant renewing, as implied by the recency principle in the RIM: more recently-activated associations are more likely to be activated again under similar conditions (Strack & Deutsch, 2004).

We had expected that participants who learn associations between No-Go and unhealthy foods better would find it easier to inhibit their impulses towards unhealthy foods in real-life situations and thus consume less of them (Jones et al., 2016) but our data did not support that hypothesis. Similarly, our index of learning the association between Go and healthy foods was not predictive of changes in healthy food intake. Both of these results are most likely due to the crudeness of the learning indices we calculated: as they are calculated across all trials, they do not take into account temporal trends. Earlier research has shown that these associations are learned very early during the performance of the task (Lawrence et al., 2015). Also, accuracy was overall very high (0.6% No-Go errors for unhealthy foods, 1.5% for No-Go filler trials) which undermined the sensitivity of the index due to a lack of variation. Future studies should therefore aim to measure learning of these associations in a more thorough way, for example by including 'catch' trials on which participants have to respond to unhealthy foods. Those who have learned a strong No-Go association should show greater slowing on these catch trials (Best et al., 2016; Veling & Aarts, 2011). These should, however, be rare as they can undermine training effects, as outlined below (Jones et al., 2016).

The results from the analyses on separate food categories indicated substantial variety in the usage and training effects. However, the pattern of results depended greatly on the specific outcome and analysis strategy. Specifically, analysing weekly servings instead of FFQ scores rendered most of the effects non-significant. This results from the data transformation on the raw FFQ data that takes into account the unequal differences between points on the FFQ scale (where a one point change can equal from half a serving per week to over ten servings per week). Removing participants from the analyses based on their baseline consumption and whether they trained to a given category rendered some results non-significant, most likely due to a loss of statistical power. It is also possible that the significant effects in the analyses on the whole sample were partly driven by those participants who did not train to a given category and showed small changes in intake as those would contribute to a larger relation between predictor and outcome. Removing them would therefore reduce the size of the effect. It is also important to note that conducting multiple tests on the same data set inflates type I error rates and the results thus need to be interpreted with caution. That being said, our main goal was to provide a description of the relationship between amount of training and changes in food intake rather than testing these values against a null hypothesis.

The sensitivity of the results to the precise analysis strategy raises the question of how to best analyse data sets from applications that allow personalization and thus contain different amounts of information for different aspects in the application (in our case different food categories) and for different groups of users (for example based on the degree to which users personalize the app). As Fig. 2 shows, the frequency with which categories were chosen differed substantially between categories. When interpreting the varying effects of different food categories it seems possible that some categories were more affected by "generalised" effects of app use (e.g., healthy vs unhealthy): sweets, for example, showed the largest decrease of all categories but a non-significant relation between intake and the amount of sweets-related no-go training. This could indicate that sweets are easily identified as a food category to be avoided and the change in intake could therefore be independent of the amount of sweets-specific training: effects of training to other sweet foods might have generalised to sweets as they might be

perceived as an archetype of sweet unhealthy foods. It is also important to take into account consumption at baseline as this determines the potential for change. For example, pizza was consumed rarely at baseline (average FFQ score 2.06) and there was thus little room for improvement whereas sweets were, on average, consumed more regularly (mean FFQ score 3.8). More generally, the observation that results differ across food categories emphasizes the need for future research to be specific about trained food categories instead of using broad categories such as “healthy” vs “unhealthy”.

The personalization feature was added in response to earlier user feedback and was supposed to increase engagement with the app (Druce et al., 2019) and optimise overall intervention effects due to more targeted training of hard-to-resist foods. While participants who personalized the app did use it substantially more, we did not find differences in the association between app use and intake between participants who chose to personalize training images and those who did not. Possibly, the selection of default foods (crisps, biscuits, chocolate and cake) matched well with many users’ “problem foods” so the default training was already targeted. Indeed, the default foods were selected on the basis of being the most frequently consumed unhealthy snack foods in a large community sample (Lawrence et al., 2015). Alternatively, this hints at the possibility of effects independent of actual intervention content and/or wide generalisations to the category of “unhealthy foods” (Serfas et al., 2017). In either case, personalization did increase engagement and, as reported elsewhere, participants commented positively on the personalization feature and have asked for additional personalization for the healthy food categories.

It is vital for further development of applications such as FoodT (but also e-health more generally) to increase users’ engagement with the intervention (Perski et al., 2017). In FoodT, this could be done by adding different ways of personalization, rewards for continued playing, or increasing the Go/No-Go task’s difficulty. Ways to make the task more challenging include: (a) adjusting the speed of the task to users’ abilities, similar to the staircase procedure in Stop-Signal Tasks (Verbruggen & Logan, 2008), (b) introducing secondary tasks such as keeping count of certain images (Simmons et al., 2005), (c) including a variety of tasks with similar hypothesized mechanisms of action like a stop-signal task or attentional bias modification (Stice et al., 2017), or (d) using slightly more complex Go/No-Go rules. For example, Veling et al. (2011) gave participants a task with two steps: first, they saw a food or control image alongside a Go or No/Go-cue; second, they saw another cue that indicated whether they were to act on the Go/No-Go cue or not. This way, the task might remain more challenging and engaging. It is important to stress that increased difficulty should not lead to substantially higher error rates, as this might undermine the learning of consistent food-No-Go associations and lead to diminished effects (Jones et al., 2016). When adjusting task speed to participants’ abilities, researchers should make sure that error rates remain low, for example by decreasing the reaction time window on Go-trials as these should not be relevant for learning the critical unhealthy-No-Go associations but may actually boost learning of approach responses to healthy foods (Schonberg et al., 2014).

4.1. Strengths and limitations

We would like to explicitly address a range of strengths and limitations of this study. The major strengths of this study include its realistic setting, the richness of the collected data, its sample size, and the diversity of the sample in terms of BMI and age. To date, only few studies have shown effects of food Go/No-Go training outside of laboratories (Forman et al., 2019; Lawrence et al., 2015; Poppelaars et al., 2018; Stice et al., 2017; Veling et al., 2014) and even fewer have explored effects of smartphone-delivered training (Blackburne et al., 2016; van Beurden et al., 2019). The relative ease of delivery and data collection of such interventions warrants further investigations into how food Go/No-Go training changes eating behaviour in users’ everyday lives.

Through smartphones, it is also more feasible to attain large and diverse samples such as the one in this study. Only with these large and diverse samples is it possible to examine which subgroups of the population benefit the most from the intervention.

The first and most serious limitation is the uncontrolled design of the study which did not assign participants to different amounts of exposure but instead relied on “naturally occurring” variations of training intensity. As pointed out above, this prohibits claims about causal effects of the training amount on changes in food intake. Additionally, it cannot rule out the possibility that the observed associations between app use and changes in food intake are artefacts of unobserved variables. However, the main goal of this pragmatic open trial was to disseminate an active training task to the general public to gather data on acceptability and real-life app use.

Secondly, the majority of participants performed relatively little Go/No-Go training. A trial that actively assigns participants to different amounts of training could avoid this problem. Ideally, such a trial would deliver training of different intensity to different groups and have them report on their food intake throughout the trial period whereas a control group would only track food intake. Future researchers could use the current report for guidance on the choice of dosage: a high dosage group, for example, should conduct training several times a day. If possible, the trial could also include different patterns of training administration (e.g. distributing 30 sessions once per day vs conducting all during one week, etc.).

Thirdly, the data on weight and food intake rely entirely on self-report and might therefore be subject to imprecision and biases. Participants might give poor estimates of their food intake and/or estimate their intake according to what they perceive to be in line with the study’s aim. Specifically, the Food Frequency Questionnaire, whilst simple and low in participant burden, is not an ideal measure of food intake as the difference between the categories is different at different parts of the scale. While we tried to ameliorate this problem by “translating” it into servings per week, these analyses were inconclusive and future studies should aim to use better measures of food intake such as food diaries which could be imported from specialised food-tracking apps or incorporated into a training app. Regarding the potential issue of social desirability, we want to point out that the assessments of weight and food intake were roughly one month apart and not accessible to participants afterwards which makes it unlikely that participants remembered their initial responses and replied in a socially desirable way on the second occasion. It is also important to note that studies using objective weight measures showed similar effects.

Fourthly, the application did not measure any indicator of possible mediators of the behavioural effects and thus does not allow any insights into mechanisms of action. The measures of associative learning in this study were probably too crude and not sensitive enough to detect effects and future studies should aim to assess probable mechanisms including food liking (Chen et al., 2016), implicit biases towards unhealthy foods (Houben & Jansen, 2015; Kakoschke et al., 2017) or automatic motor responses (Best et al., 2016; Veling & Aarts, 2011).

4.2. Conclusions

Using data from 1234 participants who conducted food Go/No-Go training using the FoodT mobile application for a one-month period, this study demonstrates that over 40% of users adhered to the 10 recommended sessions and that participants who used the app more reported larger reductions of unhealthy food intake and larger increases in healthy food intake. Our analyses suggest that spacing training out over time is more beneficial than concentrating it. Future controlled trials should aim to confirm these observational findings to determine optimal training schedules for potential users.

Author contributions

Matthias Aulbach: planning and execution of data analysis, writing the manuscript, creating tables and figures.

Keegan Knittle: planning of data analysis, major contributions to structure and content of manuscript.

Samantha van Beurden: planning of data analysis, contributions to structure and content of manuscript.

Ari Haukkala: planning of data analysis, contributions to structure and content of manuscript.

Natalia Lawrence: data collection, planning of data analysis, major contributions to structure and content of manuscript.

Ethical statement

The research reported in this manuscript has been approved by The School of Psychology Research Ethics Committee at the University of Exeter.

Funding

This research was supported by grants to Matthias Aulbach from the Finnish Cultural Foundation, the Signe and Ane Gyllenberg Foundation, the Alfred Kordelin Foundation, and the University of Helsinki's Faculty of Social Sciences.

Declaration of competing interest

None.

“Acknowledgments

Our dear colleague Ari Haukkala has passed away before publication of this article. We dedicate this article to his memory.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.appet.2021.105315>.

References

- Allom, V., Mullan, B., & Hagger, M. S. (2016). Does inhibitory control training improve health behaviour? A meta-analysis. *Health Psychology Review*, *10*(2), 168–186. <https://doi.org/10.1080/17437199.2015.1051078>
- Atkinson, N. L., & Gold, R. S. (2002). The promise and challenge of eHealth interventions. *American Journal of Health Behavior*, *26*(6), 494–503. <https://doi.org/10.5993/ajhb.26.6.10>
- Aulbach, M. B., Knittle, K., & Haukkala, A. (2019). Implicit process interventions in eating behaviour: A meta-analysis examining mediators and moderators. *Health Psychology Review*, *13*(2), 179–208. <https://doi.org/10.1080/17437199.2019.1571933>
- Bakkour, A., Botvinik-Nezer, R., Cohen, N., Hover, A. M., Poldrack, R. A., & Schonberg, T. (2018). Spacing of cue-approach training leads to better maintenance of behavioral change. *PLoS One*, *13*(7). <https://doi.org/10.1371/journal.pone.0201580>
- Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences*, *16*(12), 593–605. <https://doi.org/10.1016/j.tics.2012.10.002>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Best, M., Lawrence, N. S., Logan, G. D., McLaren, I. P. L., & Verbruggen, F. (2016). Should I stop or should I go? The role of associations and expectancies. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(1), 115–137. <https://doi.org/10.1037/xhp0000116>
- van Beurden, S. B., Greaves, C. J., Smith, J. R., & Abraham, C. (2016). Techniques for modifying impulsive processes associated with unhealthy eating: A systematic review. *Health Psychology*, *35*(8), 793–806. <https://doi.org/10.1037/hea0000337>
- van Beurden, S. B., Smith, J. R., Lawrence, N. S., Abraham, C., & Greaves, C. J. (2019). Feasibility randomized controlled trial of ImpulsePal: Smartphone app-based weight management intervention to reduce impulsive eating in overweight Adults. *JMIR Formative Research*, *3*(2), e11586. <https://doi.org/10.2196/11586>
- Blackburne, T., Rodriguez, A., & Johnstone, S. J. (2016). A serious game to increase healthy food consumption in overweight or obese adults: Randomized controlled trial. *JMIR Serious Games*, *4*(2), e10. <https://doi.org/10.2196/games.5708>
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378. <https://doi.org/10.1007/s10648-012-9205-z>
- Champely, S. (2020). pwr: Basic functions for power analysis. <https://CRAN.R-project.org/package=pwr>.
- Chen, Z., Holland, R. W., Quandt, J., Dijksterhuis, A., & Veling, H. (2019). When mere action versus inaction leads to robust preference change. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspa0000158>
- Chen, Z., Veling, H., Dijksterhuis, A., & Holland, R. W. (2016). How does not responding to appetitive stimuli cause devaluation: Evaluative conditioning or response inhibition? *Journal of Experimental Psychology: General*, *145*(12), 1687–1701. <https://doi.org/10.1037/xge0000236>
- Christiansen, E., & Garby, L. (2002). Prediction of body weight changes caused by changes in energy balance. *European Journal of Clinical Investigation*, *32*(11), 826–830. <https://doi.org/10.1046/j.1365-2362.2002.01036.x>
- Churchill, S., & Jessop, D. C. (2011). Reflective and non-reflective antecedents of health-related behaviour: Exploring the relative contributions of impulsivity and implicit self-control to the prediction of dietary behaviour. *British Journal of Health Psychology*, *16*(2), 257–272. <https://doi.org/10.1348/135910710X498688>
- Druce, K. L., Dixon, W. G., & McBeth, J. (2019). Maximizing engagement in mobile health studies: Lessons learned and future directions. *Rheumatic Disease Clinics*, *45*(2), 159–172. <https://doi.org/10.1016/j.rdc.2019.01.004>
- Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research*, *7*(1), e11. <https://doi.org/10.2196/jmir.7.1.e11>
- Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, *15*(5), 451–474. [https://doi.org/10.1016/0091-7435\(86\)90024-1](https://doi.org/10.1016/0091-7435(86)90024-1)
- Forman, E. M., Manasse, S. M., Dallal, D. H., Crochiere, R. J., Loyka, C. M., Butryn, M. L., Juarascio, A. S., & Houben, K. (2019). Computerized neurocognitive training for improving dietary health and facilitating weight loss. *Journal of Behavioral Medicine*. <https://doi.org/10.1007/s10865-019-00024-5>
- Global Health Observatory. (2017). WHO | Overweight and obesity. WHO. <http://www.who.int/gho/ncd/riskfactors/overweight/en/>.
- Hackman, C. L., & Knowlton, A. P. (2014). Theory of reasoned action and theory of planned behavior-based dietary interventions in adolescents and young adults: A systematic review. *Adolescent Health, Medicine and Therapeutics*, *5*, 101–114. <https://doi.org/10.2147/AHMT.S56207>
- Hlavac, M. (2018). Well-formatted Regression and summary statistics tables. Central European labour studies institute (CELSI). [stargazer https://CRAN.R-project.org/package=stargazer](https://CRAN.R-project.org/package=stargazer).
- Hofmann, W., Friese, M., & Wiers, R. W. (2008). Impulsive versus reflective influences on health behavior: A theoretical framework and empirical review. *Health Psychology Review*, *2*(2), 111–137. <https://doi.org/10.1080/17437190802617668>
- Houben, K., & Jansen, A. (2015). Chocolate equals stop: Chocolate-specific inhibition training reduces chocolate intake and go associations with chocolate. *Appetite*, *87*, 318–323. <https://doi.org/10.1016/j.appet.2015.01.005>
- Johns, D. J., Hartmann-Boyce, J., Jebb, S. A., & Aveyard, P. (2014). Diet or exercise interventions vs combined behavioral weight management programs: A systematic review and meta-analysis of direct comparisons. *Journal of the Academy of Nutrition and Dietetics*, *114*(10), 1557–1568. <https://doi.org/10.1016/j.jand.2014.07.005>
- Jones, A., Di Lemma, L. C. G. G., Robinson, E., Christiansen, P., Nolan, S., Tudur-Smith, C., & Field, M. (2016). Inhibitory control training for appetitive behaviour change: A meta-analytic investigation of mechanisms of action and moderators of effectiveness. *Appetite*, *97*, 16–28. <https://doi.org/10.1016/j.appet.2015.11.013>
- Kakoschke, N., Kemps, E., & Tiggemann, M. (2017). The effect of combined avoidance and control training on implicit food evaluation and choice. *Journal of Behavior Therapy and Experimental Psychiatry*, *55*, 99–105. <https://doi.org/10.1016/j.jbtep.2017.01.002>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lawrence, N. S., O'Sullivan, J., Parslow, D., Javaid, M., Adams, R. C., Chambers, C. D. C. D., Kos, K., & Verbruggen, F. (2015). Training response inhibition to food is associated with weight loss and reduced energy intake. *Appetite*, *95*, 17–28. <https://doi.org/10.1016/j.appet.2015.06.009>
- Lawrence, N. S., van Beurden, S. B., Javaid, M., & Mostazir, M. M. (n.d.). Mass dissemination of computer- and smartphone-delivered food response inhibition training to reduce unhealthy snacking and weight in the general public. In Preparation.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>
- Michie, S., Yardley, L., West, R., Patrick, K., & Greaves, F. (2017). Developing and evaluating digital interventions to promote behavior change in health and health care: Recommendations resulting from an international workshop. *Journal of Medical Internet Research*, *19*(6), e232. <https://doi.org/10.2196/jmir.7126>
- Mikkilä, V., Vepsäläinen, H., Saloheimo, T., Gonzalez, S. A., Meisel, J. D., Hu, G., Champagne, C. M., Chaput, J.-P., Church, T. S., Katzmarzyk, P. T., Kuriyan, R., Kurpad, A., Lambert, E. V., Maher, C., Maia, J., Matsudo, V., Olds, T., Onywera, V., Sarmiento, O. L., & Fogelholm, M. (2015). An international comparison of dietary patterns in 9–11-year-old children. *International Journal of Obesity Supplements*, *5*(S2), S17–S21. <https://doi.org/10.1038/ijosup.2015.14>

- Murray, E., Hekler, E. B., Andersson, G., Collins, L. M., Doherty, A., Hollis, C., Rivera, D. E., West, R., & Wyatt, J. C. (2016). Evaluating digital health interventions: Key questions and approaches. *American Journal of Preventive Medicine*, 51(5), 843–851. <https://doi.org/10.1016/j.amepre.2016.06.008>
- Navarro, D. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners*. University of Adelaide Version 0.5. <http://ua.edu.au/ccs/teaching/lsr>.
- Perro, J. (2016). Mobile apps: What's A good retention rate?. , March 29) <http://info.localytics.com/blog/mobile-apps-whats-a-good-retention-rate>.
- Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2), 254–267. <https://doi.org/10.1007/s13142-016-0453-1>
- Poppelars, A., Scholten, H., Granic, I., Veling, H., Johnson-Glenberg, M. C., & Luijten, M. (2018). When winning is losing: A randomized controlled trial testing a video game to train food-specific inhibitory control. *Appetite*, 129, 143–154. <https://doi.org/10.1016/j.appet.2018.06.039>
- Poushter, J. (2016). *Smartphone ownership and internet usage continues to climb in emerging economies* (Pew Research Center).
- R Core Team. (2017). *R: A language and environment for statistical computing*. R foundation for statistical computing. Vienna: Austria. <https://www.r-project.org/>.
- Rundel, C., Cetinkaya-Rundel, M., Clyde, M., & Banks, D. (2018). statsr: Companion package for statistics with R. <https://CRAN.R-project.org/package=statsr>.
- Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N. A., Williams, S. L., Duncan, M. J., & Vandelanotte, C. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 13(1), 127. <https://doi.org/10.1186/s12966-016-0454-y>
- Schonberg, T., Bakkour, A., Hover, A. M., Mumford, J. A., Nagar, L., Perez, J., & Poldrack, R. A. (2014). Changing value through cued approach: An automatic mechanism of behavior change. *Nature Neuroscience*, 17(4), 625–630. <https://doi.org/10.1038/nn.3673>
- Serfas, B. G., Florack, A., Büttner, O. B., & Voegeding, T. (2017). What does it take for sour grapes to remain sour? Persistent effects of behavioral inhibition in go/no-go tasks on the evaluation of appetitive stimuli. *Motivation Science*, 3(1), 1–18. <https://doi.org/10.1037/mot0000051>
- Sheeran, P., Bosch, J. A., Crombez, G., Hall, P. A., Harris, J. L., Papies, E. K., & Wiers, R. W. (2016). Implicit processes in health psychology: Diversity and promise. *Health Psychology*, 35(8), 761–766. <https://doi.org/10.1037/hea0000409>
- Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, 15(10), 1602–1608. <https://doi.org/10.1093/cercor/bhi038>
- Stanley, D. (2018). Create American psychological association (APA) style tables. apaTables <https://CRAN.R-project.org/package=apaTables>.
- Stice, E., Yokum, S., Veling, H., Kemps, E., & Lawrence, N. S. (2017). Pilot test of a novel food response and attention training treatment for obesity: Brain imaging data suggest actions shape valuation. *Behaviour Research and Therapy*, 94, 60–70. <https://doi.org/10.1016/j.brat.2017.04.007>
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 8(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1
- Thow, A. M., Downs, S., & Jan, S. (2014). A systematic review of the effectiveness of food taxes and subsidies to improve diets: Understanding the recent evidence. *Nutrition Reviews*, 72(9), 551–565. <https://doi.org/10.1111/nure.12123>
- Veling, H., & Aarts, H. (2011). Changing impulsive determinants of unhealthy behaviours towards rewarding objects. Taylor & Francis Group <https://doi.org/10.1080/17437199.2010.539970>. Vol. 5.
- Veling, H., Aarts, H., & Papies, E. K. (2011). Using stop signals to inhibit chronic dieters' responses toward palatable foods. *Behaviour Research and Therapy*, 49(11), 771–780. <https://doi.org/10.1016/j.brat.2011.08.005>
- Veling, H., Aarts, H., & Stroebe, W. (2013). Stop signals decrease choices for palatable foods through decreased food evaluation. *Frontiers in Psychology*, 4(NOV), 1–7. <https://doi.org/10.3389/fpsyg.2013.00875>
- Veling, H., van Koningsbruggen, G. M. G. M., Aarts, H., & Stroebe, W. (2014). Targeting impulsive processes of eating behavior via the internet. Effects on body weight. *Appetite*, 78(March), 102–109. <https://doi.org/10.1016/j.appet.2014.03.014>
- Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, 12(11), 418–424. <https://doi.org/10.1016/J.TICS.2008.07.005>
- Walker, A. (2018). openxlsx: Read, write and edit XLSX files. <https://CRAN.R-project.org/package=openxlsx>.
- Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science*, 2(3), 214–227. <https://doi.org/10.1177/2515245919848684>
- Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, H., François, R., Henry, L., & Müller, K. (2020). Dplyr: A grammar of data manipulation. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, H., & Henry, L. (2020). tidy: Tidy messy data. <https://CRAN.R-project.org/package=tidy>.