

Beyond Inhibitory Control Training: Inactions and Actions Influence Smartphone App Use Through Changes in Explicit Liking

Niklas Johannes

Radboud University and University of Glasgow

Moniek Buijzen

Radboud University and Erasmus University Rotterdam

Harm Veling

Radboud University

Human behavior can be classified into 2 basic categories: execution of responses and withholding responses. This classification is used in go/no-go training, where people respond to some objects and withhold their responses to other objects. Despite its simplicity, there is now substantial evidence that such training is powerful in changing human behavior toward such objects. However, it is poorly understood how simple responses can influence behavior. Contrary to the remarkably tenacious idea that go/no-go training changes behavior by strengthening inhibitory control, we propose that the training changes behavior via changes in explicit liking of objects. In two preregistered experiments, we show that go/no-go training influences explicit liking for smartphone apps (Experiments 1 and 2) and that this liking partially mediates the effect of the training on consequential choices for using these apps 1 day later (Experiment 2). The results highlight the role of evaluations when examining how motor response training influences behavior. This knowledge can inform development of more effective applied motor response training procedures and raises new theoretical questions on the relation between motor responses and affect.

Keywords: motor response training, behavior, affect, smartphone apps, go/no-go


Human behavior toward attractive objects is difficult to change. For instance, modifying food choices or alcohol intake to promote health and well-being is famously difficult (Marteau, Hollands, & Fletcher, 2012). Yet in recent years, there has been accumulating evidence that so-called motor response training where people

respond to some objects and not to others can have profound and long-lasting effects on behavior toward these attractive objects (Aulbach, Knittle, & Haukkala, 2019; Jones et al., 2016; Schonberg et al., 2014; Turton, Bruidegom, Cardi, Hirsch, & Treasure, 2016). Crucially, there is a debate on how such simple nonreinforced motor actions and inactions influence behavior (e.g., Aulbach et al., 2019; Veling, Lawrence, Chen, van Koningsbruggen, & Holland, 2017). Here, we propose that changes in liking play an important role in changing behavior for trained objects.

Despite its variety, human behavior can be divided into two basic categories: execution of responses and inhibition of responses (Guitart-Masip, Duzel, Dolan, & Dayan, 2014). This basic categorization is used in one version of motor response training, called go/no-go (GNG) training. Participants execute simple motor responses to images of some objects when a go cue is presented (go objects) and withhold motor responses to other objects when a no-go cue is presented (no-go objects). When these objects are foods or beverages, participants tend to consume less of no-go objects than go objects and choose go objects over no-go objects for consumption (Allom, Mullan, & Hagger, 2016; Jones et al., 2016; Veling, Chen, Huaiyu, Quandt, & Holland, 2019).

The mechanism underlying GNG effects is poorly understood. Responses in GNG are neither incentivized nor punished; people respond or do not respond to objects without receiving feedback. As a result, training effects cannot be accounted for by Pavlovian, evaluative, or operant conditioning alone, where objects or behaviors are paired with events with clear evaluative connotations (De Houwer, 2007). Instead, GNG was first introduced as a training of

This article was published Online First September 3, 2020.

 Niklas Johannes, Behavioural Science Institute, Radboud University, and Institute of Neuroscience and Psychology, University of Glasgow; Moniek Buijzen, Behavioural Science Institute, Radboud University, and Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam; Harm Veling, Behavioural Science Institute, Radboud University.

This research is part of a programmatic line of research by Harm Veling to understand how go and no-go responses to objects influence object evaluations and behavior. Most of this work to date has focused on changing responses to attractive food items. During collaboration with Niklas Johannes, who examines cognitive and emotional responses to smartphone cues, the question was raised whether and how go/no-go training may influence smartphone app use. This question provided an opportunity to not only examine the generalizability of the training but also examine the mechanism of how simple go and no-go responses change behavior. This work has been presented as a poster at the Annual Conference of the American Psychological Association in Washington, DC in May 2019.

Correspondence concerning this article should be addressed to Niklas Johannes, who is now at the Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, United Kingdom. E-mail: niklas.johannes@oii.ox.ac.uk

inhibitory control (Houben & Jansen, 2011) and is still presented as such (e.g., Allom et al., 2016; Carbine & Larson, 2019; Forman et al., 2019; Jones et al., 2016; Yang et al., 2019). According to this account, repeatedly not responding to attractive objects serves as an exercise to train the brain to become better at executing control in order to resist temptations.

However, there are substantial problems with this account. First, there is no evidence that GNG and other motor response trainings can improve inhibitory control to such an extent that training effects can be observed for behavior (Beauchamp, Kahn, & Berkman, 2016; Enge et al., 2014; Inzlicht & Berkman, 2015). Second, strengthening inhibitory control is assumed to be a long-term process, yet previous research often observes effects of GNG after a single training session (e.g., choices for go objects over no-go objects; Chen, Holland, Quandt, Dijksterhuis, & Veling, 2019). Third, effects of GNG on choosing go over no-go objects have been shown in within-participant designs (Chen et al., 2019). A general effect of GNG on inhibitory control capacity cannot explain within-participant effects because inhibitory capacity improvements should apply to both go and no-go high-value options. Finally, it is doubtful whether executive control can be trained with such a simple training procedure (Veling et al., 2017). In light of these findings and arguments, the current and consistent portrayal of GNG as an inhibitory control training (e.g., Carbine & Larson, 2019; Forman et al., 2019; Yang et al., 2019) is remarkably tenacious. A lack of convincing evidence for alternative explanations may contribute to this status quo.

If GNG does not strengthen inhibitory control, how can we explain its robust effects on behavior? In line with the long tradition in psychology demonstrating the importance of evaluations for changing behavior (Ajzen, 1991; Sheeran et al., 2016), we propose that evaluations play a crucial role. Indeed, GNG has been shown to decrease explicit liking of no-go objects compared to go objects or untrained objects for a variety of stimuli such as abstract art-like shapes (Clancy, Fiacconi, & Fenske, 2019), food items (Chen, Veling, Dijksterhuis, & Holland, 2016), cigarette cues (Scholten, Granic, Chen, Veling, & Luijten, 2019), and erotic images (Driscoll, de Launay, & Fenske, 2018; Ferrey, Frischen, & Fenske, 2012).

Several theoretical accounts explain how nonreinforced actions and inactions may influence evaluations. During operant evaluative conditioning, valence from a conditioned response transfers to an unrelated stimulus, thereby influencing explicit evaluations of the stimulus (Blask, Frings, & Walther, 2016; De Houwer, 2007; Eder, Krishna, & Van Dessel, 2019). In a previous experiment (Eder et al., 2019), participants first went through a phase of operant conditioning (i.e., using a key was followed by an unpleasant or pleasant picture). This response key was then used to categorize another set of neutral stimuli (i.e., assign group members to groups). The affect associated with the response (i.e., the key) transferred to this new set (i.e., the groups). Importantly, this affect transfer was shown for explicit ratings of the groups. When applied to GNG, some scholars argue that people have learned that responding is associated with positive outcomes and withholding responses with negative outcomes (Clancy et al., 2019; Guitart-Masip et al., 2014; Guitart-Masip et al., 2012), although there is also evidence that effects of avoidance behaviors in motor response training are malleable with context (Mertens, Van Dessel, & De Houwer, 2018). This affect may transfer to stimuli during the

GNG, thereby changing people's explicit evaluations of the stimuli (Hughes, De Houwer, & Perugini, 2016).

According to the inference account (Van Dessel, Hughes, De Houwer, 2018b, 2018a), people make inferences about affect associated with a response ("It doesn't feel good to stop, and I stopped for this object, so I think I don't like this object that much"). This mechanism could also explain the transfer of affect during operant evaluative conditioning (Eder et al., 2019). Third, the devaluation-by-inhibition account assigns evaluations a crucial function for task performance (e.g., Raymond, Fenske, & Westoby, 2005). According to this account, people devalue distracting objects so that they do not interfere with the task. No-go objects may be considered distractors, particularly when no-go objects are attractive, because they interfere with the goal of responding (Quandt, Holland, Chen, & Veling, 2019). Finally, according to the behavior stimulus interaction theory (Chen et al., 2016; Veling, Holland, & van Knippenberg, 2008), inhibition of a response to an attractive object leads to a response conflict, which is accompanied by negative affect that becomes attached to the attractive object.

So far, it is unclear whether changes in evaluations as a result of actions and inactions are strong enough to explain behavior change (Chen et al., 2016). To test changes in stimulus evaluations as a mechanism for behavior change, there is a need for evidence that changes in evaluation mediate the effect of the training on behavior. Such evidence is lacking for several reasons. First, studies demonstrating an effect of GNG on objectively measured, consequential behavior did not investigate the role of stimulus evaluation (Chen et al., 2019). Second, the few studies measuring both a form of stimulus evaluation and behavior did not assess actual, consequential behavior. For instance, previous work reporting mediation relied on self-reported retrospective behavior (Houben, Havermans, Nederkoorn, & Jansen, 2012) or hypothetical choices (Veling, Aarts, & Stroebe, 2013). In addition, because these studies did not employ a priori power analyses, they might well have been underpowered to detect a mediation effect. Another study employed an indirect measurement of behavior (i.e., weight loss), self-reported eating behavior, and explicit liking of food after GNG but did not observe evidence for mediation (Lawrence et al., 2015).

The choice of how to assess stimulus evaluation seems to contribute to a lack of a convincing test. Recent evidence suggests that motor response training may have stronger effects on explicit than on implicit evaluations (Van Dessel et al., 2018a). This could explain generally small effects of motor response training on implicit evaluations (e.g., Aulbach et al., 2019). For example, GNG influenced implicit evaluations of alcoholic beverages, but these implicit evaluations presented only a marginally significant predictor of self-reported alcohol intake (Houben et al., 2012). An approach/avoidance training did influence alcohol-approach tendencies and implicit associations, but neither of these implicit measures mediated the effect of the training on treatment outcome (Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011). Hence, in addition to employing self-reported or hypothetical measures of behavior, previous studies likely did not observe evidence for mediation because measurements of implicit evaluations may not be suitable (and/or reliable; Jones et al., 2016) to test whether changes in evaluation mediate the effect of motor response training on behavior.

Taken together, changes in evaluations as a result of motor response training may play a critical role for behavior, but convincing evidence for this hypothesis is currently lacking. Therefore, the main goal of this study was to move beyond GNG as inhibitory control training and test a more plausible mechanism. We aimed to provide first evidence that motor response training in the form of GNG leads to changes in behavior via changes in explicit object evaluation. We addressed the limitations of previous research and conducted preregistered, well-powered experiments relying on explicit measures of evaluations and consequential behavior. Thereby, we provide a rigorous test of an alternative mechanism to the inhibitory control training account.

A secondary goal of this study was to test whether effects of GNG also can be observed for an unstudied class of objects: smartphone apps. We chose smartphone apps for two reasons. First, previous work on motor response training has mostly focused on objects that are attractive because they yield nonsymbolic, immediate bodily reward. Food, alcohol, cigarettes, or sex have been the primary focus of investigation in the response training literature (Allom et al., 2016; Aulbach et al., 2019; Jones et al., 2016). Smartphone apps represent symbolic reward without immediate bodily consequences. Expanding the effect of GNG to this unexplored category of objects may yield important insights into the generalizability of motor response training (Yarkoni, 2019). Second, whereas smartphone use does not appear detrimental to well-being (Ellis, 2019; Orben, Dienlin, & Przybylski, 2019; Orben & Przybylski, 2019), many users voice concerns about decreased productivity because of smartphone distractions (Johannes, Dora, & Rusz, 2019). Therefore, GNG may serve as an intervention to modify preferences for smartphone apps, presenting a promising tool for those users who would like to reduce their smartphone use.

The Present Research

In Experiment 1, we predicted that GNG causes no-go smartphone apps to be liked less from pretraining to posttraining, compared to both go smartphone apps and smartphone apps not used in GNG (i.e., untrained apps). We did not predict increased liking for go items as the version of GNG we employed here tends to influence evaluations of no-go rather than go objects (Chen et al., 2016). After establishing an effect on evaluations in Experiment 1, we predicted that the evaluations of smartphone apps modified by GNG would fully mediate the effect of the training on consequential choices for using these apps in Experiment 2.

Experiment 1

We preregistered our hypotheses, sampling plan, exclusion criteria, and confirmatory analysis plan and provide access to all data and stimulus materials on the Open Science Framework (<https://osf.io/7ck43/>).

Method

Sample. Power calculations with mixed-effects models can be complicated (Brybaert & Stevens, 2018; DeBruine & Barr, 2019). Therefore, we used the *simr* package (Green & Macleod, 2016) to simulate power based on the data of Experiment 1 of Chen et al.

(2016), which was almost identical to our design. Given the novelty of smartphone stimuli as objects and to properly power our experiment, we calculated power for 75% of the effect size they found. To detect such an effect with 80% power at $\alpha = .05$, we needed to recruit 63 participants. Note that 63 participants also are needed to achieve 95% power for a repeated-measures analysis of variance (ANOVA) assuming $\eta^2 = .117$, which is 75% of the smallest effect size reported in Chen et al. for experiments similar to our design. In order to account for exclusion criteria, we recruited 70 participants. Participants were students from our institute and received credit or €10. We obtained institutional review board approval; all participants gave informed consent.

We had three inclusion criteria: First, we only recruited students between the ages of 18 and 30 years as they are part of the population who display the strongest phone use (CBS, 2018). Second, because the GNG training required at least 30 app icons, participants needed to (a) have 30 apps installed and (b) rate a minimum of 30 app icons during the preevaluation. Third, we only recruited iPhone users. All participants should be familiar with our stimuli; contrary to the Android OS, of which the open source nature allows manufacturers to amend app icons, icons are standardized across iOS devices. This way, we could be certain the app icons participants rated were identical to those they know from their own phones. None of the participants fulfilled our preregistered exclusion criteria of (a) 85% accuracy or lower on the GNG task, where exceeding the response window counted as incorrect, and (b) a mean on the preevaluation of lower than -50 across all conditions. Thus, our final sample was $N = 70$ ($M_{\text{age}} = 22.20$, $SD_{\text{age}} = 2.54$, 55 female).

Design. We employed a 3 (Condition: go vs. no-go vs. untrained) \times 2 (Time: pre vs. post) within design. The dependent variable was app icon evaluation as dependent variable.

Procedure.

Deprivation period. Research on GNG or cued-approach training and food evaluations or preferences usually asks participants to fast before the experiment (Chen et al., 2016; Zoltak, Veling, Chen, & Holland, 2018) to reduce between-participants differences in hunger levels, which may impact food ratings. In addition, fasting will ensure that the food items are on average at least somewhat appealing so that no-go devaluation can occur. We emulated this procedure for smartphone apps: Participants came to the lab 1 hr before the experiment and locked their phones away. This way, we minimized between-participants differences in how recently participants had used their phones right before the experiment started. By minimizing this difference, we reduced the probability that app icon ratings might be strongly influenced by one specific interaction with an app right before the start of the experiment. In addition, an hour of deprivation has shown to increase motivation to use one's phone (Johannes et al., 2019), ensuring that apps were perceived as attractive for the rating task. After handing in their phones, participants were free to go about their day, but we instructed them not to use any of their iPhone applications on a laptop or tablet. After 1 hr, they returned and did the experiment. Participants received their phones back after the study.

Materials. For the GNG training, we needed ratings of 30 app icons. Because we could not be sure which apps participants were familiar with, we presented them with a large selection of preinstalled and popular apps. Specifically, we selected 43 app icons

that are preinstalled on each iPhone with iOS 11; in addition, we selected the top 100 list of free apps on iTunes (<https://www.apple.com/nl/itunes/charts/free-apps/>, April 3, 2018). Icons were presented at the center of the screen against a white background. The task was programmed with Python 2.7 (Python Core Team, 2018), using PsychoPy (Peirce, 2007).

Pretraining evaluations. Participants were instructed that the first part was about what makes an app icon look attractive to users. Thus, they rated the 143 app icons on the question “How attractive does this app icon look to you?” on a visual analogue scale, ranging from -100 (*not at all*) to 100 (*extremely*). We chose this question wording because it was closest to the validated question used in previous studies on rating food pictures (e.g., Chen et al., 2016). Item order was randomized. If participants did not know an app, defined as not being familiar or never having used the app, they could press the “D” button to skip the rating. All 143 apps received at least one rating, and participants were familiar with a large number of apps ($M = 65.81$, $SD = 23.78$). The highest-rated app (rated by at least seven participants, 10% of the sample) was WhatsApp ($M = 61.43$, $SD = 42.89$); the lowest-rated app was Egg ($M = -35.44$, $SD = 39.43$).

Condition assignment. After the prerating, the experimental Python program rank-ordered the app icons from highest to lowest. It then repeatedly assigned the three conditions (go, no-go, untrained) from highest- to lowest-rated app icon for the 30 highest-rated apps. To minimize prerating differences between conditions, item assignment was mirrored across the 30 apps starting from the top-ranked icon (e.g., go, no-go, untrained, untrained, no-go, go, etc.). This order was counterbalanced across participants. There were 10 app icons per condition.

Go/no-go training. After the pretraining evaluations, participants were informed that they would do an attention task. We told them that we were interested in how well people can focus their attention while looking at different apps. During the GNG training, each trial began with a single app icon in the middle of the screen. After 100 ms, participants heard one of two tones via headphones for 300 ms. The tones were at the 1,000 Hz and 400 Hz frequencies and served as the go or no-go cue. Which tone served as which cue was fully crossed with the condition assignment. For go cues, participants had to press the “B” key as quickly as possible. For no-go cues, they were instructed to not press any key. To rule out that participants were affected by reduced exposure time if the app icon disappeared after the “B” press, go and no-go icons both stayed on-screen for 1,000 ms. Intertrial interval was random in steps of 100 ms for each trial and ranged between 1,000 ms and 1,500 ms.

Participants first received a practice block of 20 trials. Icons for the practice trials were taken from the bottom of the list that rank-ordered all 143 app icons, from which the program also selected the 30 highest-rated icons for the training. Thus, icons used for practice trials either received no or very low ratings in the pretraining evaluations. These icons were then randomly assigned to the go or no-go condition. During practice trials, participants received error feedback. After the practice block, participants were given the opportunity to practice again. If they chose to proceed, participants received 160 total experimental trials. The 20 icons were presented eight times; presentation order was random. After each 40 trials, participants could take a short break and received

progress feedback. There was no more error feedback during the experimental block.

Posttraining evaluations. After the GNG training, participants did the same rating task again, but this time only for the 30 selected icons. They were instructed to rate how attractive each app icon was as if it were the first time they saw it.

Results

We used R (Version 3.5.0; R Core Team, 2018) for all analyses. There was strong evidence that our condition assignment was successful in creating conditions that were matched on pretraining ratings; a Bayesian repeated-measures ANOVA with the `anovaBF` command (BayesFactor package, Version 0.9.12–2; Morey & Rouder, 2015) with the default priors of the function indicated a Bayes factor (BF) of 173 in favor of a model with no differences between the three conditions. Evaluations of the icons decreased from pretraining ($M = 53.15$, $SD = 21.47$) to posttraining ($M = 32.50$, $SD = 19.72$). This is in line with previous work and generally interpreted as regression to the mean (Chen et al., 2016). Similar to previous work, all participants were highly accurate during the GNG task ($M = 98.4\%$, $SD = 1.6\%$). The mean reaction time (RT) on correct experimental go trials was 399 ms ($SD = 56$ ms).

Confirmatory analyses. Following our preregistered analysis plan, we calculated difference scores between posttraining evaluations and pretraining evaluations (post minus pre), where lower scores indicate stronger devaluation ($M = -20.65$, $SD = 19.00$). We tested the effect of training on the difference score with a linear mixed-effects model using the `lmer` function (`lme4` package, Version 1.1–17; Bates, Mächler, Bolker, & Walker, 2015). Following recommendations on best practices for mixed-effects models to avoid inflated Type I error, we employed a maximal random-effects structure (Barr, Levy, Scheepers, & Tily, 2013). Because we had two grouping factors, we modeled two random intercepts, one for participant and one for icon, in order to account for the nested nature of the data. In addition, we modeled condition (i.e., go vs. no-go vs. untrained) as a fixed effect and as a random slope varying across participants and icons.

The model converged without warnings. In line with our pre-registration, to obtain p values, we computed bootstrapped likelihood ratio tests using the mixed function (`afex` package, Version 0.20–2; Singmann, Bolker, Westfall, & Aust, 2018), which in turn calls the function `PBmodcomp` (`pbkrtest` package, Version 0.4–7; Halekoh & Højsgaard, 2014). All tests relied on 10,000 bootstraps. The effect of condition on the difference score was significant, $PBtest = 16.93$, $p < .001$. To obtain an approximation of the effect size, we squared the correlation between observed and fitted values, $R^2 = .38$. Further, we called the `r.squaredGLMM` function to obtain Pseudo R^2 for mixed models (Barton, 2018), which yielded an estimate of .009 for the variance explained by fixed factors and .34 for the variance explained by both fixed and random factors.

Following our pre-registration, we conducted follow-up models to investigate the pairwise comparisons; see Figure 1. As predicted, no-go stimuli ($M = -25.26$, $SD = 22.09$) had a significantly lower difference score than go stimuli ($M = -16.90$, $SD = 18.50$), $PBtest = 13.95$, $p < .001$. No-go stimuli also displayed a significantly lower difference score than untrained stimuli

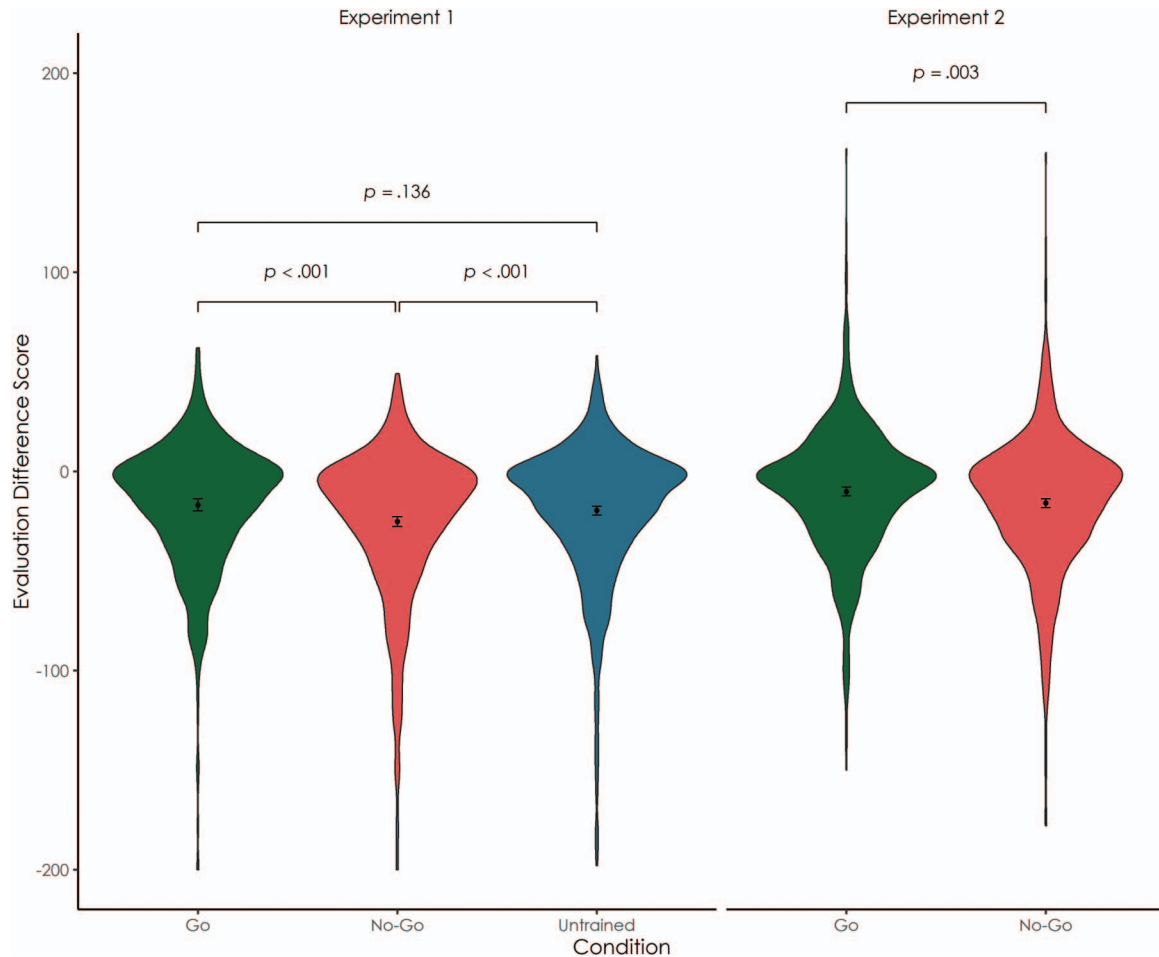


Figure 1. Violin plots of the evaluations for both experiments. Black dots in the violins represent the mean; bars of these points represent the 95% confidence interval of the within-subjects standard error (Morey, 2008). The difference score was calculated by subtracting pretraining evaluations from posttraining evaluations. See the online article for the color version of this figure.

($M = -19.78$, $SD = 22.20$), $PBtest = 12.35$, $p < .001$. This specific pattern of results is known as the no-go devaluation effect.

Exploratory analyses. The main effect of condition also aligned with a Bayesian repeated-measures ANOVA with the standard Cauchy prior, with condition as within-subjects predictor and difference scores as outcome, which displayed very convincing evidence (Lee & Wagenmakers, 2013) in favor of the model with condition as predictor compared to a null model, $BF = 10,360$. Follow-up paired t tests with the standard Cauchy prior displayed strong evidence in favor of a difference in difference scores between no-go and untrained stimuli ($BF_{10} = 45.39$) and between no-go and go ($BF_{10} = 139$).

Our design typically leads to devaluation of no-go apps but not to increased evaluation of go apps (Chen et al., 2016). In line with previous research, although go stimuli had higher difference scores than untrained stimuli, this difference was not statistically significant, $PBtest = 2.12$, $p = .136$. A Bayesian paired t test on difference scores showed that the data provided anecdotal (Lee & Wagenmakers, 2013) evidence for a lack of a difference between these conditions, $BF_{01} = 2.59$.

Furthermore, it is possible that the negative affect people experience is due to making errors during the GNG. As such, errors could explain the devaluation effect we found. To rule this possibility out, we ran the same model again, excluding all items from the evaluation tasks on which participants made at least one error during the GNG. Excluding these icons did not explain the effect as all model parameters remained stable; that is, no-go apps still had a lower difference score than go and untrained apps ($p < .001$ for main effect).

Finally, as an additional robustness check, we also tested whether our results were robust to excluding possible influential participants or icons. Specifically, we removed three participants and three icons because they stood out on plots visualizing Cook's distance and DFBETAs. All tests were robust to the exclusion of these cases, and all parameters remained virtually unchanged (all significant differences remained at $p < .001$).

Experiment 2

For Experiment 2, we had three aims. First, we aimed to test whether the effect found in Experiment 1 was robust and would

replicate. To that end, we predicted again that no-go apps would decrease more strongly from pretraining to posttraining evaluations compared to go apps. Second, we aimed to demonstrate the effect of the training on choices. Namely, we predicted that the probability of choosing go over no-go items for actual use would be significantly higher than 50%. Third, we aimed to test the mediation mechanism. When participants must choose between a go and a no-go app, we expected that they would make their decision based on the difference in evaluations between those two apps. Therefore, we predicted that evaluations would fully mediate the effect of the training on choices. In addition, after testing our predictions, we aimed to explore a complementary account of the influence of the training on choices. Repeatedly not responding to apps during GNG training may create an association between these apps and not responding (Best, Lawrence, Logan, McLaren, & Verbruggen, 2016; Verbruggen & Logan, 2008), which also may impact choices. In the exploratory section before the general discussion, we present tests of this so-called stimulus-stop account. Preregistration of this experiment, specifying our hypotheses, sampling plan, exclusion criteria, and analysis plan, as well as all data, analysis scripts, and materials can be found on the Open Science Framework project of this article (<https://osf.io/7ck43/>).

Method

Sample. According to our power simulation for Experiment 1, we needed 63 participants to reliably detect an effect of the training on evaluations with 80% power. To detect the effect of the training on choice, we followed recent power simulations by Chen et al. (2019). They showed that 60 participants are needed for 80% power to detect an effect of at least the meta-analytic effect size of $d = 0.50$ (Allom et al., 2016). Just like in Experiment 1, we aimed to be conservative with our sample to account for the novelty of our stimuli. Therefore, we collected a roughly 25% larger sample than would be required according to the simulations; that is, we preregistered to recruit 80 participants ($M_{\text{age}} = 22.33$, $SD_{\text{age}} = 2.20$, 57 female). Inclusion criteria were the same as in Experiment 1, except that this time participants had to have 35 rather than 30 apps installed and needed to rate at least 32 apps during the first rating task. No one fulfilled our preregistered exclusion criteria of (a) 85% accuracy or lower on the GNG task, where exceeding the response window counted as incorrect; (b) a mean on the preevaluation of lower than -50 across all conditions; or (c) choosing one side during the choice task 90% of the time or higher. We obtained institutional review board approval; all participants gave informed consent.

Design. We employed a 2 (Condition: go vs. no-go) \times 2 (Time: pre vs. post) within design with evaluation as dependent variable. The second dependent variable was choice, with the same condition factor, but only one measurement.

Procedure.

Day 1. The procedure on the first day was identical to Experiment 1, except for the following changes. First, we updated the top 100 free apps on iTunes (October 2018). Second, during the pretraining evaluations, participants were instructed to only rate apps they had installed on their phones and to skip rating those apps they had not. The change in instructions was evident in two instances: Not all apps received a rating this time (130 compared to all 143 in Experiment 1), and participants rated less apps on

average ($M = 45.45$, $SD = 8.08$). The highest-rated app (rated by at least eight participants, 10% of sample) was Netflix ($M = 57.67$, $SD = 35.89$); the lowest-rated app was Watch ($M = -29.33$, $SD = 44.99$). Third, we omitted the untrained condition because (a) we already established that no-go items were rated as significantly lower than both untrained and go items and (b) choices may be influenced by familiarity with the icons and untrained items receive less visual exposure than the go and no-go icons. Consequently, the Python program again rank-ordered the app icons from highest to lowest. It then repeatedly assigned the two conditions (go, no-go) from highest to lowest app icon for the 32 highest-rated apps. This condition assignment was counterbalanced.

Finally, in Experiment 1, icons were presented completely randomly across all trials during the GNG task, which can lead to several icons of the same condition occurring in succession. For Experiment 2, the 32 experimental items were presented twice per block, over four blocks, and the order was randomized within each block. Thus, there was a total of 256 experimental trials.

Day 2.

Choice task construction. On the second day, participants received a choice task. We constructed choice trials from the 32 highest-rated apps on the pretraining evaluations from the first day, which also were used during the GNG task. We created two kinds of choice pairs. First, experimental choice pairs consisted of choices between go and no-go icons that were matched on preratings. These pairs allowed us to test whether using go apps would be preferred over using no-go apps. Second, for the purpose of validating both the evaluation task as well as the choice task, we also included filler choice pairs of two apps that differed in value but were both of the same condition (i.e., both go or both no-go). If the ratings of the app icons and choices were meaningful to participants, participants should prefer higher-rated apps over lower-rated apps on these filler trials. Furthermore, these filler trials allowed us to conduct a causal test of mediation. Mediation mechanisms without manipulating the mediator can be prone to bias (Bullock, Green, & Ha, 2010). In our case, the filler trials served as a manipulation of evaluations because we manipulated value of apps (high vs. low). If this manipulation had an effect on choice, we could be confident that any effect of the difference in evaluations between no-go and go on choices would indeed reflect a causal effect.

From the rank-ordered list of 32 apps, going from highest to lowest, we divided apps into ranks in order to construct experimental and filler choice pairs: (a) eight apps for experimental choice trials high in value, (b) four apps for filler choice trials high in value, (c) eight apps for experimental choice trials medium in value, (d) four apps for filler choice trials low in value, and (e) eight apps for experimental choice trials low in value (see Figure 2). Within each of the three experimental ranks of eight icons (high, medium, low), go icons were always paired with no-go icons; thus, there were 16 possible combinations of go and no-go icons in each rank (4 go icons \times 4 no-go icons). We obtained those combinations for each of the three experimental ranks separately to keep the value difference between go and no-go icons for each experimental choice pair relatively low. Thus, there was a total of 48 unique GNG pairs (i.e., 16 pairs each for high, medium, and low experimental rank). These 48 unique pairs were then presented twice in the choice task, where the position of the icons (left or

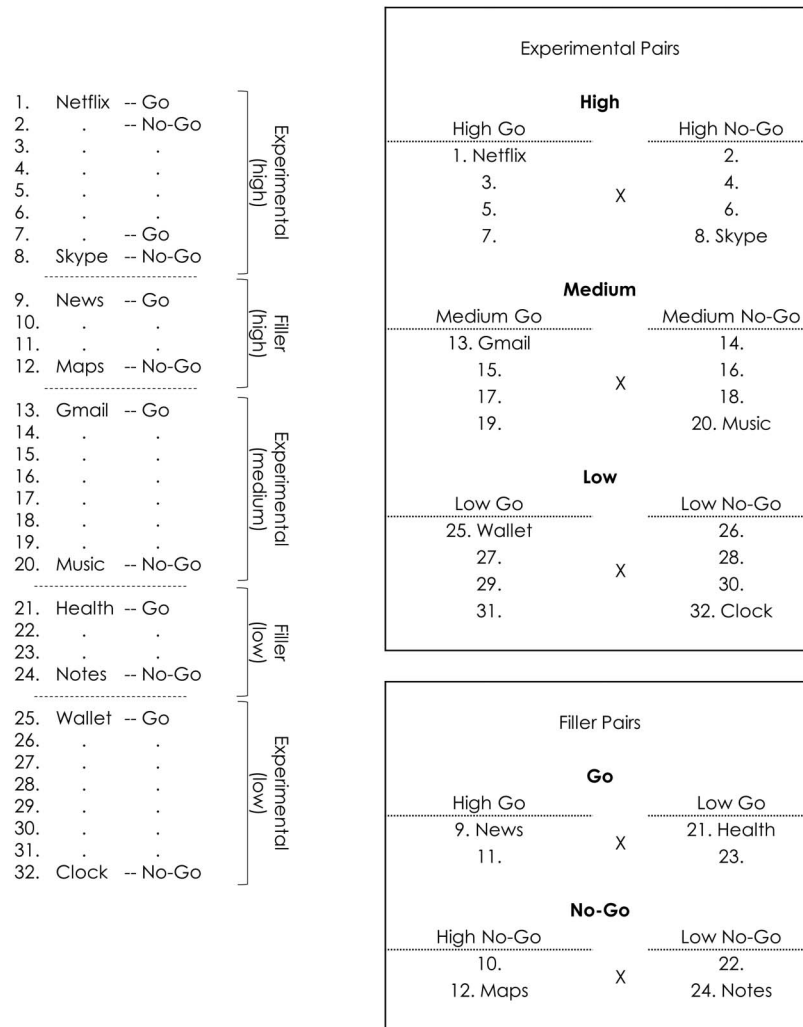


Figure 2. An illustration of how choice trials were constructed. On the left is an exemplary list of the 32 highest-rated apps, ordered from highest pretraining evaluation to lowest. The assignment of the condition (go vs. no-go) was counterbalanced across participants. The 32 highest apps were then divided into different ranks, three experimental (high, low, medium) and two filler ranks (high, low). The two boxes on the right visualize how choice trials were constructed from these ranks.

right) was counterbalanced, resulting in a total of 96 experimental choice trials per participant.

Between the two filler ranks, go icons of the high-value rank were always paired with go icons of the low-value rank, and no-go icons of the high-value rank were always paired with no-go icons of the low-value rank. There were two go icons and two no-go icons in each filler rank, thus allowing for eight unique possible combinations of choice pairs within the same condition, but with different values: $(2 \text{ high value go icons} \times 2 \text{ low value go icons}) + (2 \text{ high value no-go icons} \times 2 \text{ low value no-go icons})$. Just like with experimental choice trials, we counterbalanced the position on the choice task, thus leading up to 16 choice trials. Last, to increase power for the effect of value on choice for filler trials and to be consistent with the number of choice trials employed in previous research (Schonberg et al., 2014), we doubled the number of filler trials once more. That is, there were 32 filler choice trials in total per participant.

Experimental blocks served to test the effect of condition, whereas filler blocks served as a check whether participants would indeed choose high over low value in the majority of cases. The total 128 choice trials (96 experimental + 32 filler) were presented in two blocks without a break. Each block contained half of the experimental and half of the filler trials, counterbalancing the position of each app icon. Before the experimental block, participants practiced the choice task with 16 choices between icons that were not used in the training.

Choice task procedure. Before doing the choice task, participants again locked their phones away for 1 hr. Apart from the reasons described earlier, this time the deprivation served an additional purpose. That is, participants made consequential choices during the choice task about which app they could use for a short while during the experimental session. Previous research showed that an hour of deprivation makes participants motivated to use their phones (Johannes et al., 2019), which made our choice

task more relevant to them. When the choice task is employed in food research, participants learn that the program will pick a random trial in the end; whatever participants choose on this trial is the food they receive (Krajbich, Armel, & Rangel, 2010). This way, choices are consequential for participants. Hence, analogous to food research, we instructed participants that the program would randomly pick a trial at the end of the choice task; whatever app participants chose on that trial was the app they were allowed to use for 3 min before we locked their phones away again for half an hour. This second deprivation phase was intended to make the choices meaningful for participants. Without that second deprivation phase, it would not have mattered which trial the program picked as participants could have just used whichever app they felt like after the experiment ended. Thus, participants locked their phones away for an additional half an hour after using the app they chose for 3 min on the trial randomly selected by the program.

After the first hour of deprivation, participants came back to the lab and did the choice task. On each trial, participants chose between two apps that were presented side by side by pressing the “U” or “I” keys. Participants had to make that choice within 1,500 ms. If they chose an app within the response window, their choice was confirmed by a yellow frame surrounding the app for 500 ms. If they did not make the choice in time, they were presented with feedback (“Choose faster!”), and the choice trial was presented again at the end of the block. Participants rarely exceeded the response window (1.22%). Intertrial interval varied randomly between 1,000 ms and 2,000 ms in steps of 100 ms.

Results

Effects on evaluation. There was no evidence for a difference in ratings between go and no-go items at the pretraining evaluations ($BF_{01} = 23.02$). Once more, we observed regression to the mean from pretraining ($M = 36.72$, $SD = 22.64$) to posttraining evaluations ($M = 23.66$, $SD = 21.24$). Accuracy was high ($M = 98.5\%$, $SD = 1.3\%$), and mean RT on correct experimental go trials was 471 ms ($SD = 54$ ms).

Confirmatory. Similar to Experiment 1, we preregistered to test the effect of the training on the difference score between pretraining and posttraining ($M = -13.07$, $SD = 12.28$); see Figure 1. Again, we employed a linear mixed-effects model with a maximal random-effects structure. The initial model failed to converge, most likely because there were 18 apps that received only one rating. Estimating a random slope for the difference between the go and no-go condition for each app requires the app to have a rating for each condition. With only one evaluation, the model cannot estimate a random slope. Consequently, it was necessary to group those apps into an “other” category; the model then treated the 18 evaluations as coming from the same group (i.e., app). Afterward, the model converged without warnings. In line with our preregistration, we followed recommendations by Luke (2017) and obtained p values with Satterthwaite approximation for degrees of freedom. The difference between go items ($M = -10.17$, $SD = 14.07$) and no-go items ($M = -15.97$, $SD = 15.34$) was significant, $F(1, 66.49) = 9.67$, $p = .003$. Squaring the correlation between observed and fitted values yielded $R^2 = .22$. Pseudo R^2 for the variance explained by the fixed factors was .008 and was .17 for the variance explained by both fixed and random factors.

Exploratory. The main effect of condition aligned with a Bayesian repeated-measures ANOVA, $BF_{10} = 4,472$. Again, excluding icons on which participants committed an error did not influence the model ($p = .004$). In addition, we excluded two participants who had a large influence on the estimates, as indicated by Cook’s distance and DFBETAs, and ran the same mixed-effects model again. The effect was robust to excluding outliers ($p = .008$). Last, we ran the same model again without those 18 apps that we grouped into the “other” category to ensure the effect was not driven by those cases. The effect was robust to the exclusion of this category ($p = .003$).

Effects on choice. Participants chose go items over no-go items on 54.8% of experimental trials ($SD = 12.3\%$). This percentage was similar across the three value ranks: 53.5% for choices high in value, 56.0% for choices medium in value, and 55.0% for choices low in value. Note that choosing go items over no-go items is equivalent to a choice against no-go items. We use this wording to stay consistent with the wording used in previous work.

Confirmatory. To test whether the overall percentage of choosing go items over no-go items was different from 50%, we ran a generalized mixed-effects intercept-only model with a random intercept per participant. Please note that we deviated from the preregistration here: We preregistered to obtain the p value with Satterthwaite approximation for degrees of freedom, which is not possible for a generalized intercept-only model. Instead, we report the p value based on Wald’s test statistic. Because Wald’s test statistic can be problematic, we also report the 95% confidence interval obtained with the profile method. The fixed intercept was significantly different from 0, estimate = .21, $SE = .059$, $z = 3.50$, $p < .001$, indicating that participants chose go items above chance level, $OR = 1.23$, 95% CI [1.09, 1.38]. That is, participants had 1.23-times higher odds of choosing go items than choosing no-go items.

Exploratory. Participants made valid choices in the choice task: On filler trials, when both app icons on a choice trial were go items or both app icons were no-go items but they differed in value, participants chose the higher-valued icon 66.9% of the time ($SD = 18.8\%$, Figure 3). A generalized mixed-effects intercept-only model with a random intercept per participant shows that participants indeed chose high-value apps above chance level, estimate = 0.82, $SE = .11$, $z = 7.46$, $p < .001$, $OR = 2.27$, 95% CI [1.83, 2.83].

As a robustness check, we also analyzed the choices for go versus no-go items with a Bayesian one-sample t test comparing the mean proportion of choices for go items against 50%. This test indicated strong evidence that choices were different from chance level, $BF_{10} = 30.86$.

There were no visual or formal outliers to test for robustness of the model. We were also interested whether the effect of the training would differ for different choice RTs as previous research has observed that the choice effect becomes weaker the more time participants take (Chen et al., 2019). However, although the coefficient was negative, RTs were not a significant predictor of app choice, $\chi^2 = .223$, $p = .135$.

Mediation. Last, we tested whether evaluations mediated the effect of the training on choice. However, we could not conduct a direct test of the mediation as the evaluation data and the choice data had different structures. For the evaluations, each case (i.e.,

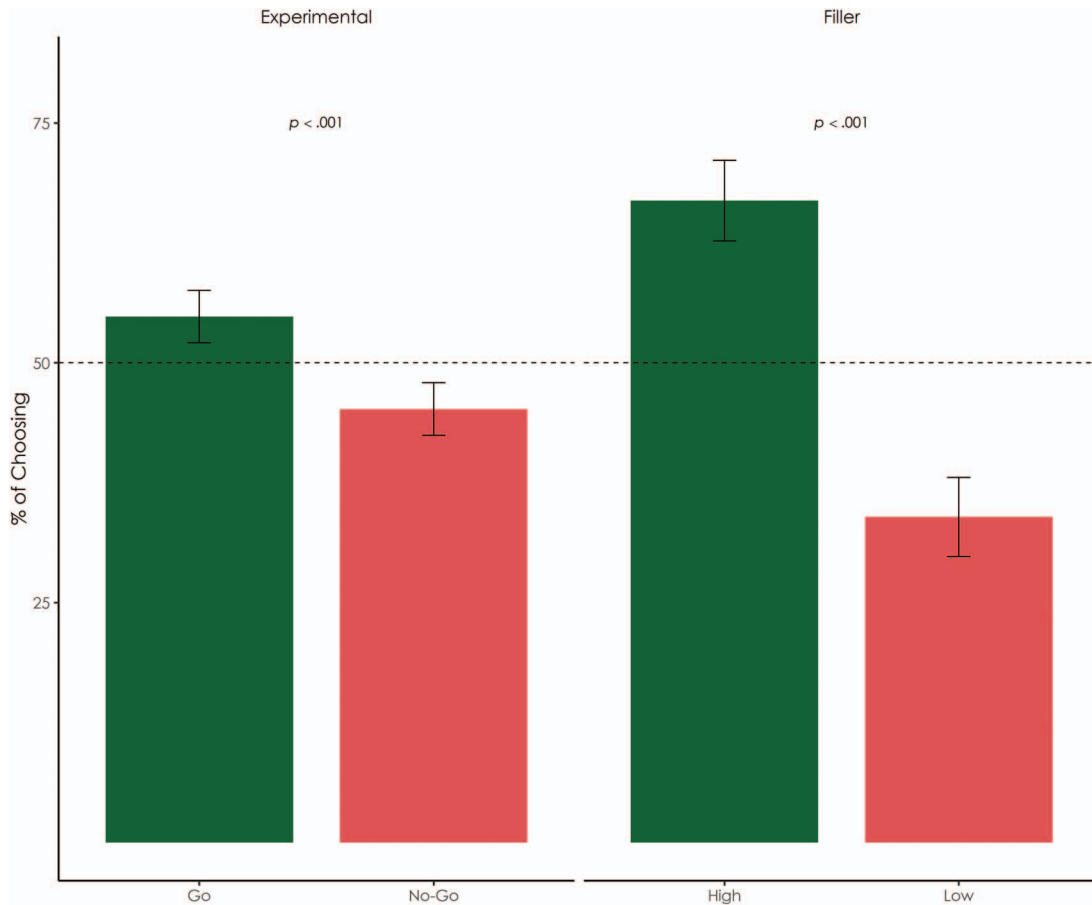


Figure 3. The left side of the graph shows the mean probability of choosing go and no-go apps in experimental trials. The right side of the graph shows the mean probability of choosing high-value and low-value apps in filler trials. *P* values reflect the overall tests of choosing go over no-go (left) and high value over low value (right) against chance level (50%, dashed line). Error bars reflect the 95% confidence interval. See the online article for the color version of this figure.

app icon) had one condition assigned. For the choices, each case (i.e., choice trial) presented two apps and thus two conditions alongside each other. As a consequence, we could not assess mediation in a single statistical test. Instead, we preregistered to employ the approach of joint significance: If the effect of the independent variable (i.e., the training) on the mediator (i.e., the evaluations) is significant, and the effect of the mediator on the outcome (i.e., choosing go over no-go) is significant, it follows that the indirect effect is likely nonzero (Kenny, Kashy, & Bolger, 1998). Although joint significance does not provide a single estimate of the indirect effect size, it performs well in assessing mediation compared to other procedures such as bootstrapping (Hayes & Scharkow, 2013). This approach also follows recent suggestions to report individual paths of the components of mediation models (Yzerbyt, Muller, Bataillier, & Judd, 2018). Consequently, we first tested whether the training would influence evaluations; afterward, we assessed whether the evaluations predicted choice.

Confirmatory. To test the first path of the mediation model, we estimated a maximal mixed-effects model with the condition of app icon as predictor and the posttraining evaluations as outcome.

We preregistered to use the posttraining evaluations as mediator because (a) we already established an effect of the training on the evaluations controlling for the pretraining evaluations and (b) we wanted to avoid using a difference score (go over no-go) of a difference score (post minus pre) to aid interpretation. As predicted, no-go items ($M = 20.75$, $SD = 23.21$) were rated significantly lower than go items ($M = 26.56$, $SD = 22.45$), $F(1, 78.91) = 10.38$, $p = .002$.

To test the second path of the mediation model, we estimated a maximal mixed-effects model with the difference score of posttraining evaluations as predictor and choice of go icons as outcome. Specifically, because each trial presented two app icons, we calculated the difference between posttraining evaluations of the go item and the no-go item (go minus no-go). Thus, positive values reflect how much higher participants evaluated the go item compared to the no-go item on that specific choice trial. In line with the effect of the training on posttraining evaluations, go items received higher ratings than no-go items ($M = 5.77$, $SD = 19.21$). This analysis enabled us to carry out the test of joint significance: For the first path of the mediation model, we tested the effect of training condition on posttraining evaluations; if those posttraining

evaluations predicted whether people chose go items over no-go items, we could have confidence that evaluations indeed mediated the effect of training on choice. Importantly, if the intercept in the model did not remain significantly different from 0 (i.e., chance level), while including evaluations as predictor, this would indicate full mediation because the evaluations can account for all variance in choices.

Note that for the reasons described above, we again had to deviate from the preregistration and obtain p values with likelihood ratio tests. As predicted, higher evaluations for go items compared to no-go items were a significant predictor of choosing go items over no-go items, estimate = .015, $SE = .002$, $\chi^2(1) = 54.03$, $p < .001$, $OR = 1.015$, 95% CI [1.011, 1.018]. Thus, with each point that participants rated go items higher than no-go items, they had 1.015-times higher odds of choosing go items over no-go items. Although these odds may seem small, evaluations were made on a visual analogue scale ranging from -100 to 100 . To get a better understanding of the effect size, the odds ratio for the average difference between go and no-go items was $1.015^{5.77} = 1.09$. If participants showed a difference of one standard deviation in their rating of go and no-go items, they would have $1.015^{19.21} = 1.33$ higher odds of choosing go over no-go items, larger than the overall effect of the training on choice; see Figure 4.

Interestingly, the intercept was also significant, estimate = .16, $SE = .06$, $z = 2.82$, $\chi^2(1) = 7.59$, $p = .006$, $OR = 1.17$, 95% CI [1.05, 1.31]. Even after accounting for the effect of evaluations, participants still had 1.17-times higher odds of choosing go items over no-go items. Hence, evaluations appear to mediate the effect of the training on choices only partially.

Exploratory. A Bayesian repeated-measures ANOVA indicated strong support for the main effect of condition on posttraining evaluations, $BF_{10} = 411$. All analyses reported to test the mediation were robust to removal of outliers.

The difference in evaluations between go and no-go apps did not fully explain the choices participants made. We were thus interested in exploring another potential explanation. According to the stimulus-stop account, participants can learn to associate the act of stopping with no-go apps (Best et al., 2016; Verbruggen & Logan, 2008). Choosing go apps over no-go apps could thus also be caused by a trained stopping response. Consequently, participants would be slower and less likely to choose no-go apps than go apps because they must overcome the learned stopping response to no-go apps. This stimulus-stop association account could complement the evaluation account; see Figure 5. Thus, according to the

stimulus-stop account, the influence of the training on choices is the result of an association between an object and a learned stopping response (Best et al., 2016; Verbruggen & Logan, 2008), but it makes no prediction about GNG training influencing evaluations.

To test the stimulus-stop account, we examined whether participants were slower on trials on which they chose no-go apps compared to trials on which they chose go apps. Slowing in their choices would reflect a direct association between stopping and no-go apps. Descriptively, participants were slightly slower when choosing no-go apps ($M = 784.80$, $SD = 104.74$) than go apps ($M = 773.54$, $SD = 107.35$). In a maximal mixed-effects model, this difference was not significant, $F(1, 75.52) = 1.62$, $p = .207$. A Bayesian paired-samples t test indicated moderate support for a lack of a difference, $BF_{01} = 3.61$.

In line with previous research (e.g., Martino, Fleming, Garrett, & Dolan, 2013; Zoltak et al., 2018), participants made faster choices on filler trials ($M = 750$, $SD = 105$), where value differences between choice alternatives were large ($M = 23.69$, $SD = 22.75$), than on experimental trials ($M = 773$, $SD = 96$), $F(1, 79) = 12.04$, $p < .001$, where value differences between choice alternatives were small ($M = 5.77$, $SD = 19.21$). Accordingly, we might see evidence for stimulus-stop associations only after accounting for within-trial value differences between choice alternatives. Predicting choice RT in experimental trials with both the posttraining value difference and condition showed a significant negative relation between value difference and RT. An increase of one standard deviation in absolute value was associated with an 11-ms faster choice ($SE = 4.8$), $F(1, 43.55) = 5.09$, $p = .029$. Importantly, the effect of condition on choice RT remained non-significant, $F(1, 75.14) = 1.59$, $p = .211$. Thus, we show that value difference between choice alternatives influences choice RT when comparing filler trials to experimental trials. Within-trial value differences are also associated with choice RT. Yet even after taking value differences within experimental trials into account, we did not find evidence for the stimulus-stop account.

Likewise, participants were descriptively slightly slower when rating no-go apps ($M = 3064$, $SD = 1095$) than go apps ($M = 3094$, $SD = 1265$) in the posttraining evaluations across both experiments. This difference, however, was again not significant, $F(1, 1009) = 0.59$, $p = .444$, with strong support for a lack of a difference, $BF_{01} = 26.56$. Together, these results do not support stimulus-stop associations as a potential explanation.

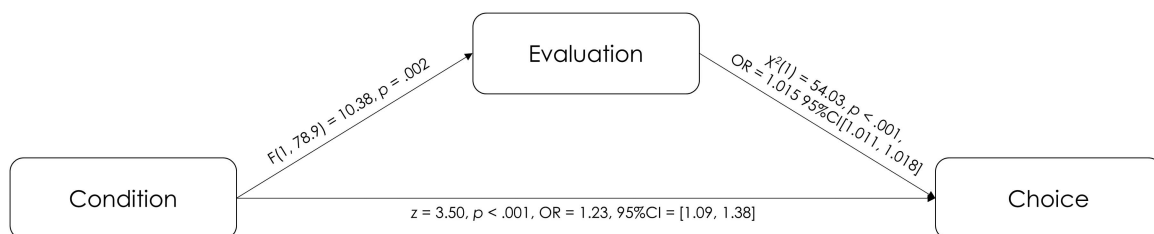


Figure 4. A visualization of the mediation model that we ran. The direct path from condition to choice represents the test of the training on choices. The path from condition to evaluation represents a test of the training on posttraining evaluations. The path from evaluation to choice represents the effect of the difference in posttraining evaluations between a go and a no-go app on choice. OR = odds ratio; CI = confidence interval.

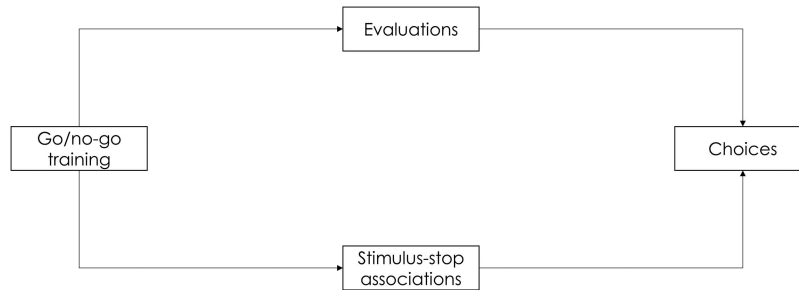


Figure 5. A visualization of the two accounts that explain the effect of go/no-go (GNG) training on choices. The evaluation account suggests that the training influences choices by decreasing liking of no-go objects. The stimulus-stop account suggests that the training establishes an association between objects and the response of stopping.

When constructing the choice trials, icons were matched on value. This matching reduced the relative difference between go and no-go apps; it is not possible, however, to match them exactly. Thus, preexisting value differences between go and no-go apps might contribute to the choices participants made. To explore this possibility, we predicted choices with the difference between go and no-go items on pretraining evaluations. Speaking to the success of our randomization, mean differences were close to zero ($M = -0.02$, $SD = 3.27$). They did significantly predict choices, estimate = .018, $SE = .004$, $\chi^2(1) = 22.06$, $p < .001$, $OR = 1.012$, 95% CI [1.011, 1.026]; the intercept remained significant, estimate = .21, $SE = .06$, $z = 3.51$, $\chi^2(1) = 11.52$, $p < .001$, $OR = 1.23$, 95% CI [1.09, 1.38]. When predicting choices with both pretraining and posttraining difference scores, both pretraining and posttraining were significant predictors. The effect of posttraining evaluations, estimate = .014, $SE = .002$, $\chi^2(1) = 43.77$, $p < .001$, $OR = 1.014$, 95% CI [1.010, 1.017], was descriptively larger than that of pretraining evaluations, estimate = .011, $SE = .003$, $\chi^2(1) = 8.55$, $p = .004$, $OR = 1.011$, 95% CI [1.004, 1.019]. The intercept was descriptively smaller but remained significant, estimate = .17, $SE = .06$, $z = 2.80$, $\chi^2(1) = 7.46$, $p = .006$, $OR = 1.18$, 95% CI [1.05, 1.33]. Therefore, variation in pretraining evaluations can explain choices to a degree. When controlling for pretraining evaluations, posttraining evaluations still predict choices. The remaining difference between go and no-go choices becomes smaller but stays significantly different from zero.

General Discussion

Motor response training tasks, such as GNG, have become popular tools to change behavior toward attractive objects (Allom et al., 2016; Aulbach et al., 2019; Jones et al., 2016; Stice, Lawrence, Kemps, & Veling, 2016; Turton et al., 2016; Yang et al., 2019). However, to date, we do not understand how they work (e.g., Aulbach et al., 2019). Possibly because of this lack of insight into the mechanism, GNG is still generally considered to improve inhibitory control (e.g., Bos et al., 2019; Forman et al., 2019), in line with the original idea (Houben & Jansen, 2011). However, on both conceptual and empirical grounds, changes in inhibitory control capacity induced by GNG cannot explain many effects of GNG on behavior (e.g., within-person effects, Chen et al., 2019). As a consequence, the inhibitory control training explanation has

been criticized as implausible (Enge et al., 2014; Inzlicht & Berkman, 2015; Veling et al., 2017).

Here, we offer an alternative by showing that evaluations play a critical role when explaining GNG effects. In two high-powered, preregistered experiments with established measures of explicit evaluations and consequential behavior, we demonstrated robust evidence that GNG influences evaluations of smartphone apps (Experiments 1 and 2) and that these evaluations mediate the effect of GNG on behavior 1 day later (Experiment 2). The experiments employed a within-participant design, which rules out improvements of inhibitory control capacity as an explanation for the preference of go over no-go objects. If GNG truly improved general inhibitory control, participants should have become better at inhibiting choosing all objects, leading to an equal likelihood of choosing go or no-go apps. We found that they were more likely to choose go apps, which was partially mediated by changes in app evaluations. In light of these findings, it might be time to start portraying GNG as a means to change object evaluations instead of an inhibitory control training.

Our experiments were designed to allow for causal claims about the mediation we observed. First, evaluations and behavior were entirely different measures, which rules out common method variance as an alternative explanation for mediation. Second, we introduced a temporal order: Evaluations succeeded the training, and choices succeeded evaluations. We observed effects of evaluations on consequential choices 1 day later. Third, for mediation to take place, we need evidence that we can manipulate the mediator and that it has the predicted effect (Bullock et al., 2010). Otherwise, confounding variables (e.g., attention; Quandt et al., 2019) might bias the results. In our case, we showed that we could manipulate the mediator in filler trials. We manipulated the value of apps, which had the expected effect of participants choosing apps of higher value. Therefore, we can be confident that evaluations indeed had a causal effect on choices.

It is noteworthy that we found effects on an unexamined category of objects, smartphone apps. Previous research mostly investigated the effects of GNG on objects with immediate bodily consequences, such as food, beverages, or sex (e.g., Chen et al., 2019; Ferrey et al., 2012; Houben et al., 2012). In contrast, there is a lack of evidence that GNG effects generalize to choices for objects that only have symbolic value. The effects on smartphone

apps we observe suggest that effects of GNG apply to a wide range of stimuli, similar to Pavlovian, evaluative, and operant conditioning (De Houwer, 2007). They also suggest that GNG might be a suitable intervention to reduce smartphone use given that previous interventions have shown to be ineffective (Parry & le Roux, 2019).

The mediation we observed was only partial and not full, as we predicted. Even after accounting for the evaluations, participants chose go over no-go apps. The strength of this choice became less pronounced when accounting for pretraining differences in value but remained above chance level. This effect suggests there is room for other influences besides evaluations. In an exploratory analysis, we did not find evidence that stimulus-stop associations contributed to participants choosing go over no-go apps (Best et al., 2016; Verbruggen & Logan, 2008), but our methods may not have been suitable to detect such an effect.

If stimulus-stop associations cannot explain the remainder of the GNG effect on choices, what other mechanism can? Recent work suggests that GNG tasks can influence the formation of episodic memories of the objects used during the task. Specifically, people have better recognition memory for go objects compared to no-go objects after the task, which is attributed to enhanced memory for action-relevant go stimuli (Yebrá et al., 2019). Episodic memory could influence choices in several ways: People may prefer items for which they have good episodic memory (e.g., Murty, FeldmanHall, Hunter, Phelps, & Davachi, 2016), or action-relevant objects grasp visual attention (Bamford, Klassen, & Karl, 2020) and visual attention can influence choices independent of stimulus evaluation (Gwinn, Leber, & Krajbich, 2019). Thus, GNG training may influence choices via two paths: by decreasing evaluations of no-go objects and by enhancing memory for action-relevant go objects (which does not enhance evaluations of go objects; see Experiment 1). Examining the remaining mechanisms underlying the effect of GNG on behavior, next to changes in evaluation, is an important next step for future research toward fully understanding motor response trainings.

The central role of evaluations raises several questions. First, it is important that future work examines which of the four accounts outlined in the introduction best explains how GNG changes object evaluations. For instance, from the perspective of operant evaluative conditioning (De Houwer, 2007; Eder et al., 2019), the affect from a conditioned response transfers to an object. It is possible that not going is naturally associated with negative affect (e.g., Clancy et al., 2019). This affect might then transfer to the object for which participants withheld a response. Such a transfer could happen via associations: The app, not going, and the unpleasantness of not going might all be associated in memory (Hommel, 2004) such that retrieving one part of this association (i.e., the app) automatically retrieves the other parts (i.e., the unpleasantness), thus decreasing liking for the app. Alternatively, an inferential account proposes that people infer their liking of objects from their own behavior (Van Dessel, Eder, & Hughes, 2018; Van Dessel et al., 2018b). According to this view, participants observed that they stopped for an app and that stopping is unpleasant, thus negatively adjusting their evaluation of the app (for a more elaborate discussion, see Eder et al., 2019). Uncovering the exact mech-

anism of how GNG influences evaluations may improve future applied training tasks to elicit stronger effects (e.g., Van Dessel et al., 2018a).

Second, there is strong evidence that some motor response training procedures can influence preferences even for months (i.e., cue-approach training; Salomon et al., 2018; Schonberg et al., 2014). It remains to be tested whether the changes in evaluations can account for effects of other motor response training tasks and whether they can account for long-term behavior change. Furthermore, our behavioral measure was strictly confined to choices that participants do not encounter in such a form in their everyday lives. There is a need for research testing whether decreased evaluations also have an impact on smartphone use in everyday life. This question is crucial as current evidence is inconsistent about how applicable motor response training is outside the lab (Jones et al., 2018; Lawrence et al., 2015).

To conclude, the present research shows that changes in evaluations are a plausible mechanism that can explain how GNG influences behavior. Importantly, this mechanism offers a more adequate explanation than the consistent portrayal of GNG as inhibitory control training. We call for more research to investigate (a) how GNG influences evaluations and (b) the role of evaluations in other motor response trainings. Apart from theoretical insight, understanding whether motor response trainings have a common mechanism can inform decisions on whether to combine different response trainings in one intervention.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211. [http://dx.doi.org/10.1016/0749-5978\(91\)90020-T](http://dx.doi.org/10.1016/0749-5978(91)90020-T)
- Allom, V., Mullan, B., & Hagger, M. (2016). Does inhibitory control training improve health behaviour? A meta-analysis. *Health Psychology Review*, 10, 168–186. <http://dx.doi.org/10.1080/17437199.2015.1051078>
- 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, 179–208. <http://dx.doi.org/10.1080/17437199.2019.1571933>
- Bamford, L. E., Klassen, N. R., & Karl, J. M. (2020). Faster recognition of graspable targets defined by orientation in a visual search task. *Experimental Brain Research*, 238, 905–916. <http://dx.doi.org/10.1007/s00221-020-05769-z>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. <http://dx.doi.org/10.1016/j.jml.2012.11.001>
- Barton, K. (2018). MuMIn: Multi-model inference (Version 1.43.17) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=MuMIn>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Beauchamp, K. G., Kahn, L. E., & Berkman, E. T. (2016). Does inhibitory control training transfer? Behavioral and neural effects on an untrained emotion regulation task. *Social Cognitive and Affective Neuroscience*, 11, 1374–1382. <http://dx.doi.org/10.1093/scan/nsw061>
- 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, 115–137. <http://dx.doi.org/10.1037/xhp0000116>
- Blask, K., Frings, C., & Walther, E. (2016). Doing is for feeling. *Journal of Experimental Psychology: General*, 145, 1263–1268. <http://dx.doi.org/10.1037/xge0000211>
- Bos, J., Staiger, P. K., Hayden, M. J., Hughes, L. K., Youssef, G., & Lawrence, N. S. (2019). A randomized controlled trial of inhibitory control training for smoking cessation and reduction. *Journal of Consulting and Clinical Psychology*, 87, 831–843. <http://dx.doi.org/10.1037/ccp0000424>
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1, 9. <http://dx.doi.org/10.5334/joc.10>
- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, 98, 550–558. <http://dx.doi.org/10.1037/a0018933>
- Carbine, K. A., & Larson, M. J. (2019). Quantifying the presence of evidential value and selective reporting in food-related inhibitory control training: A *p*-curve analysis. *Health Psychology Review*, 13, 318–343. <http://dx.doi.org/10.1080/17437199.2019.1622144>
- CBS. (2018). *Internet; toegan, gebruik en faciliteiten [Internet; access, use and devices]*. Retrieved from <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83429NED/table?ts=1531831858987>
- 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*, 117, 721–740. <http://dx.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, 1687–1701. <http://dx.doi.org/10.1037/xge0000236>
- Clancy, E. M., Fiacconi, C. M., & Fenske, M. J. (2019). Response inhibition immediately elicits negative affect and devalues associated stimuli: Evidence from facial electromyography. In N. Srinivasan (Ed.), *Progress in brain research* (pp. 169–191). <http://dx.doi.org/10.1016/bs.pbr.2019.03.019>
- DeBruine, L., & Barr, D. J. (2019). Understanding mixed effects models through data simulation. *PsyArXiv*. <http://dx.doi.org/10.31234/osf.io/xp5cy>
- De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, 10, 230–241. <http://dx.doi.org/10.1017/S1138741600006491>
- Driscoll, R. L., de Launay, K. Q., & Fenske, M. J. (2018). Less approach, more avoidance: Response inhibition has motivational consequences for sexual stimuli that reflect changes in affective value not a lingering global brake on behavior. *Psychonomic Bulletin & Review*, 25, 463–471. <http://dx.doi.org/10.3758/s13423-017-1291-y>
- Eder, A. B., Krishna, A., & Van Dessel, P. (2019). Operant evaluative conditioning. *Journal of Experimental Psychology: Animal Learning and Cognition*, 45, 102–110. <http://dx.doi.org/10.1037/xan0000189>
- Ellis, D. A. (2019). Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior*, 97, 60–66. <http://dx.doi.org/10.1016/j.chb.2019.03.006>
- Enge, S., Behnke, A., Fleischhauer, M., Küttler, L., Kliegel, M., & Strobel, A. (2014). No evidence for true training and transfer effects after inhibitory control training in young healthy adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 987–1001. <http://dx.doi.org/10.1037/a0036165>
- Ferrey, A. E., Frischen, A., & Fenske, M. J. (2012). Hot or not: Response inhibition reduces the hedonic value and motivational incentive of sexual stimuli. *Frontiers in Psychology*, 3, 575. <http://dx.doi.org/10.3389/fpsyg.2012.00575>
- Forman, E. M., Manasse, S. M., Dallal, D. H., Crochiere, R. J., Loyka, C. M., Butryn, M. L., . . . Houben, K. (2019). Computerized neurocognitive training for improving dietary health and facilitating weight loss. *Journal of Behavioral Medicine*, 42, 1029–1040. <http://dx.doi.org/10.1007/s10865-019-00024-5>
- Green, P., & Macleod, C. J. (2016). simr: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7, 493–498. <http://dx.doi.org/10.1111/2041-210X.12504>
- Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014). Action versus valence in decision making. *Trends in Cognitive Sciences*, 18, 194–202. <http://dx.doi.org/10.1016/j.tics.2014.01.003>
- Guitart-Masip, M., Huys, Q. J. M., Fuentesmilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage*, 62, 154–166. <http://dx.doi.org/10.1016/j.neuroimage.2012.04.024>
- Gwinn, R., Leber, A. B., & Krajbich, I. (2019). The spillover effects of attentional learning on value-based choice. *Cognition*, 182, 294–306. <http://dx.doi.org/10.1016/j.cognition.2018.10.012>
- Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models—The R package pbkrtest. *Journal of Statistical Software*, 59, 1–30. <http://dx.doi.org/10.18637/jss.v059.i09>
- Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 24, 1918–1927. <http://dx.doi.org/10.1177/0956797613480187>
- Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, 8, 494–500. <http://dx.doi.org/10.1016/j.tics.2004.08.007>
- Houben, K., Havermans, R. C., Nederkoorn, C., & Jansen, A. (2012). Beer à no-go: Learning to stop responding to alcohol cues reduces alcohol intake via reduced affective associations rather than increased response inhibition. *Addiction*, 107, 1280–1287. <http://dx.doi.org/10.1111/j.1360-0443.2012.03827.x>
- Houben, K., & Jansen, A. (2011). Training inhibitory control: A recipe for resisting sweet temptations. *Appetite*, 56, 345–349. <http://dx.doi.org/10.1016/j.appet.2010.12.017>
- Hughes, S., De Houwer, J., & Perugini, M. (2016). Expanding the boundaries of evaluative learning research: How intersecting regularities shape our likes and dislikes. *Journal of Experimental Psychology: General*, 145, 731–754. <http://dx.doi.org/10.1037/xge0000100>
- Inzlicht, M., & Berkman, E. (2015). Six questions for the resource model of control (and some answers). *Social and Personality Psychology Compass*, 9, 511–524. <http://dx.doi.org/10.1111/spc3.12200>
- Johannes, N., Dora, J., & Rusz, D. (2019). Social smartphone apps do not capture attention despite their perceived high reward value. *Collabra Psychology*, 5, 14. <http://dx.doi.org/10.1525/collabra.207>
- Jones, A., Di Lemma, L. C. 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. <http://dx.doi.org/10.1016/j.appet.2015.11.013>
- Jones, A., McGrath, E., Robinson, E., Houben, K., Nederkoorn, C., & Field, M. (2018). A randomized controlled trial of inhibitory control training for the reduction of alcohol consumption in problem drinkers. *Journal of Consulting and Clinical Psychology*, 86, 991–1004. <http://dx.doi.org/10.1037/ccp0000312>
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). New York, NY: Oxford University Press.
- Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13, 1292–1298. <http://dx.doi.org/10.1038/nn.2635>

- Lawrence, N. S., O'Sullivan, J., Parslow, D., Javaid, M., Adams, R. C., Chambers, C. D., . . . Verbruggen, F. (2015). Training response inhibition to food is associated with weight loss and reduced energy intake. *Appetite*, *95*, 17–28. <http://dx.doi.org/10.1016/j.appet.2015.06.009>
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139087759>
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–1502. <http://dx.doi.org/10.3758/s13428-016-0809-y>
- Marteau, T. M., Hollands, G. J., & Fletcher, P. C. (2012). Changing human behavior to prevent disease: The importance of targeting automatic processes. *Science*, *337*, 1492–1495. <http://dx.doi.org/10.1126/science.1226918>
- Martino, B. D., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*, 105–110. <http://dx.doi.org/10.1038/nn.3279>
- Mertens, G., Van Dessel, P., & De Houwer, J. (2018). The contextual malleability of approach-avoidance training effects: Approaching or avoiding fear conditioned stimuli modulates effects of approach-avoidance training. *Cognition and Emotion*, *32*, 341–349. <http://dx.doi.org/10.1080/02699931.2017.1308315>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64. <http://dx.doi.org/10.20982/tqmp.04.2.p061>
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (R Package version 0.9.12–2) [Computer software]. Retrieved from <https://cran.r-project.org/package=BayesFactor>
- Murty, V. P., FeldmanHall, O., Hunter, L. E., Phelps, E. A., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision-making. *Journal of Experimental Psychology: General*, *145*, 548–558. <http://dx.doi.org/10.1037/xge0000158>
- Orben, A., Dienlin, T., & Przybylski, A. K. (2019). Social media's enduring effect on adolescent life satisfaction. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 10226–10228. <http://dx.doi.org/10.1073/pnas.1902058116>
- Orben, A., & Przybylski, A. K. (2019). Screens, teens, and psychological well-being: Evidence from three time-use-diary studies. *Psychological Science*, *30*, 682–696. <http://dx.doi.org/10.1177/0956797619830329>
- Parry, D. A., & le Roux, D. B. (2019). Media multitasking and cognitive control: A systematic review of interventions. *Computers in Human Behavior*, *92*, 316–327. <http://dx.doi.org/10.1016/j.chb.2018.11.031>
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. <http://dx.doi.org/10.1016/j.jneumeth.2006.11.017>
- Python Core Team. (2018). Python: A dynamic, open source programming language [Computer software]. Retrieved from <https://www.python.org/>
- Quandt, J., Holland, R. W., Chen, Z., & Veling, H. (2019). The role of attention in explaining the no-go devaluation effect: Effects on appetitive food items. *Journal of Experimental Psychology: Human Perception and Performance*, *45*, 1119–1133. <http://dx.doi.org/10.1037/xhp0000659>
- Raymond, J. E., Fenske, M. J., & Westoby, N. (2005). Emotional devaluation of distracting patterns and faces: A consequence of attentional inhibition during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1404–1415. <http://dx.doi.org/10.1037/0096-1523.31.6.1404>
- R Core Team. (2018). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Salomon, T., Botvinik-Nezer, R., Gutentag, T., Gera, R., Iwanir, R., Tamir, M., & Schonberg, T. (2018). The cue-approach task as a general mechanism for long-term non-reinforced behavioral change. *Scientific Reports*, *8*, 3614. <http://dx.doi.org/10.1038/s41598-018-21774-3>
- Scholten, H., Granic, I., Chen, Z., Veling, H., & Luijten, M. (2019). Do smokers devalue smoking cues after go/no-go training? *Psychology & Health*, *34*, 609–625. <http://dx.doi.org/10.1080/08870446.2018.1554184>
- 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*, 625–630. <http://dx.doi.org/10.1038/nn.3673>
- Sheeran, P., Maki, A., Montanaro, E., Avishai-Yitshak, A., Bryan, A., Klein, W. M. P., . . . Rothman, A. J. (2016). The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: A meta-analysis. *Health Psychology*, *35*, 1178–1188. <http://dx.doi.org/10.1037/hea0000387>
- Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). Afex: Analysis of factorial experiments (R Package version 0.20–2) [Computer software]. Retrieved from <https://cran.r-project.org/package=afex>
- Stice, E., Lawrence, N. S., Kemps, E., & Veling, H. (2016). Training motor responses to food: A novel treatment for obesity targeting implicit processes. *Clinical Psychology Review*, *49*, 16–27. <http://dx.doi.org/10.1016/j.cpr.2016.06.005>
- Turton, R., Bruidegom, K., Cardi, V., Hirsch, C. R., & Treasure, J. (2016). Novel methods to help develop healthier eating habits for eating and weight disorders: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *61*, 132–155. <http://dx.doi.org/10.1016/j.neubiorev.2015.12.008>
- Van Dessel, P., Eder, A. B., & Hughes, S. (2018). Mechanisms underlying effects of approach-avoidance training on stimulus evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1224–1241. <http://dx.doi.org/10.1037/xlm0000514>
- Van Dessel, P., Hughes, S., & De Houwer, J. (2018a). Consequence-based approach-avoidance training: A new and improved method for changing behavior. *Psychological Science*, *29*, 1899–1910. <http://dx.doi.org/10.1177/0956797618796478>
- Van Dessel, P., Hughes, S., & De Houwer, J. (2018b). How do actions influence attitudes? An inferential account of the impact of action performance on stimulus evaluation. *Personality and Social Psychology Review*, *23*, 267–284. <http://dx.doi.org/10.1177/1088868318795730>
- Veling, H., Aarts, H., & Stroebe, W. (2013). Stop signals decrease choices for palatable foods through decreased food evaluation. *Frontiers in Psychology*, *4*, 875. <http://dx.doi.org/10.3389/fpsyg.2013.00875>
- Veling, H., Chen, Z., Huaiyu, L., Quandt, J., & Holland, R. W. (2019). Updating the p-curve analysis of Carbine & Larson with results from preregistered experiments. *Health Psychology Review*, *14*, 215–219. <http://dx.doi.org/10.1080/17437199.2019.1669482>
- Veling, H., Holland, R. W., & van Knippenberg, A. (2008). When approach motivation and behavioral inhibition collide: Behavior regulation through stimulus devaluation. *Journal of Experimental Social Psychology*, *44*, 1013–1019. <http://dx.doi.org/10.1016/j.jesp.2008.03.004>
- Veling, H., Lawrence, N. S., Chen, Z., van Koningsbruggen, G. M., & Holland, R. W. (2017). What is trained during food go/no-go training? A review focusing on mechanisms and a research agenda. *Current Addiction Reports*, *4*, 35–41. <http://dx.doi.org/10.1007/s40429-017-0131-5>
- Verbruggen, F., & Logan, G. D. (2008). Automatic and controlled response inhibition: Associative learning in the go/no-go and stop-signal paradigms. *Journal of Experimental Psychology: General*, *137*, 649–672. <http://dx.doi.org/10.1037/a0013170>
- Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, *22*, 490–497. <http://dx.doi.org/10.1177/0956797611400615>

- Yang, Y., Shields, G. S., Wu, Q., Liu, Y., Chen, H., & Guo, C. (2019). Cognitive training on eating behaviour and weight loss: A meta-analysis and systematic review. *Obesity Reviews*, *20*, 1628–1641. <http://dx.doi.org/10.1111/obr.12916>
- Yarkoni, T. (2019). The generalizability crisis. *PsyArXiv*. <http://dx.doi.org/10.31234/osf.io/jqw35>
- Yebra, M., Galarza-Vallejo, A., Soto-Leon, V., Gonzalez-Rosa, J. J., de Berker, A. O., Bestmann, S., . . . Strange, B. A. (2019). Action boosts episodic memory encoding in humans via engagement of a noradrenergic system. *Nature Communications*, *10*, 3534. <http://dx.doi.org/10.1038/s41467-019-11358-8>
- Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology*, *115*, 929–943. <http://dx.doi.org/10.1037/pspa0000132>
- Zoltak, M. J., Veling, H., Chen, Z., & Holland, R. W. (2018). Attention! Can choices for low value food over high value food be trained? *Appetite*, *124*, 124–132. <http://dx.doi.org/10.1016/j.appet.2017.06.010>

Received June 13, 2019

Revision received May 8, 2020

Accepted May 10, 2020 ■



AMERICAN PSYCHOLOGICAL ASSOCIATION

APA Journals®

ORDER INFORMATION

Subscribe to This Journal for 2021

Order Online:

Visit at.apa.org/xge-2021
for pricing and access information.

Call **800-374-2721** or **202-336-5600**

Fax **202-336-5568** | TDD/TTY **202-336-6123**

Subscription orders must be prepaid. Subscriptions are on a calendar year basis. Please allow 4-6 weeks for delivery of the first issue.

All APA journal subscriptions include Online First journal articles and access to archives. Individuals can receive online access to all of APA's scholarly journals through a subscription to APA PsycNet® or through an institutional subscription to the APA PsycArticles® database.

Visit AT.APA.ORG/CIRC2021
to browse APA's full journal collection