Registered Replication of Wessel and colleagues [(2014)](https://www.zotero.org/google-docs/?ydg7g5) “Stimulus devaluation induced by stopping action”

Patrick G. Bissett, Jamie K. Li, Whitney Davidson, Corey N. White, Tom Schonberg, Russell A. Poldrack

1. Department of Psychology, Stanford University

2. Department of Psychology, Missouri Western State University

3. Department of Neurobiology and Sagol School of Neuroscience, Tel Aviv University

Abstract

There is a growing literature on inhibition induced devaluation, in which withholding a behavioral response in the context of a specific stimulus reduces the value of that stimulus and reduces the likelihood of choosing or using that stimulus in the future. Inhibition induced devaluation holds great promise for behavior change and improving public health, because a variety of public health problems, including unhealthy eating and substance abuse, involves the inability to control the use of something or inherent or perceived in value. Wessel and colleagues (2014) showed that when arbitrary stimuli were associated with response inhibition that subjects valued those stimuli less, suggesting that inhibition induced devaluation may be possible across a variety of stimuli. In two direct replications and seven conceptual replications of Wessel and colleagues (2014), we show no evidence for inhibition induced devaluation. This brings into question the robustness of inhibition induced devaluation, at least as implemented by Wessel and colleagues (2014) using arbitrary stimuli. We propose to supplement these nine datasets with three highly powered, direct registered replications of Wessel and colleagues (2014) ran independently across three research sites. All existing data are openly available (<https://osf.io/x38aj/>) and all data from proposed data will be made available.

Unhealthy behaviors, such as smoking, drug and alcohol abuse, overeating, and a sedentary lifestyle, account for a substantial proportion of the risk associated with preventable premature deaths in the United States [(Nielsen et al., 2018)](https://www.zotero.org/google-docs/?ljI85w). This has led to a pressing need for interventions that result in healthy behavior changes. One promising intervention is inhibition induced devaluation (IID), in which withholding a behavioral response in the context of a specific stimulus reduces the value of that stimulus, and even results in the subject being less willing to choose or use that stimulus in the future. IID holds great promise for improving public health, because a variety of public health problems can be distilled to the inability to control the use of something of inherent or perceived value. The promise of IID is that by devaluing a stimulus by associating it with inhibition, an individual will be able to more effectively control their urge to use the valuable but dangerous substance, facilitating better choices.

Inhibition is a fundamental mechanism of control that facilitates goal-directed behavior [(Logan, 1985)](https://www.zotero.org/google-docs/?ozI4Kz). In order to navigate a dynamic world, actions and goals must also have the flexibility to stop and change. In such instances, response inhibition is engaged [(Bissett & Logan, 2011, 2013; Camalier et al., 2007; Logan & Cowan, 1984)](https://www.zotero.org/google-docs/?zgg58Z).

Two main paradigms used to capture response inhibition are the go/no-go task [(Donders, 1868/1969)](https://www.zotero.org/google-docs/?ZLwh1l) and the stop-signal task [(Logan & Cowan, 1984)](https://www.zotero.org/google-docs/?1fxu6G). Both tasks involve the presentation of a series of stimuli, with one or more stimuli indicating the need to respond and a different set of stimuli indicating the need to withhold a response. The go/no-go task involves presenting only one stimulus on any given trial and subjects make a response (i.e., go) if that stimulus has a certain identity but do not make a response (i.e., no-go) if that stimulus has a different identity. However, theoretical [(Gomez, Ratcliff, & Perea, 2007)](https://www.zotero.org/google-docs/?RELhPW) and empirical work has questioned whether go/no-go tasks require response inhibition, at least under certain task parameters [(Wessel, 2018)](https://www.zotero.org/google-docs/?ZPvdsh). In contrast, the stop-signal task involves making a choice response to a “go” stimulus that is presented on every trial, except for on the subset of trials in which a second, subsequent “stop signal” is presented.

Both the stop and the go/no-go task have been used to implement IID. In the canonical implementation, certain stimuli are often or always presented in the context of withholding a motor response, and other stimuli are seldom or never presented in the context of withholding a motor response. In a subsequent phase of the study, subjects are either asked to do some form of value judgment on each class of stimuli, given the opportunity to consume the two classes of stimuli (e.g., if the classes are food), or both.

In general, stimuli that were associated with withholding a motor response are valued less and chosen less than stimuli not associated with withholding a motor response. Associating specific food stimuli with no-going [(Veling, Aarts, & Stroebe, 2013)](https://www.zotero.org/google-docs/?k9ppce) as well as stopping [(Lawrence, Verbruggen, Morrison, Adams, & Chambers, 2015)](https://www.zotero.org/google-docs/?71oWPA) reduces subsequent consumption of the trained food stimuli. Also, associating specific alcohol stimuli with no-going [(Houben, Nederkoorn, Wiers, & Jansen, 2011)](https://www.zotero.org/google-docs/?zrLedr) as well as stopping [(Jones & Field, 2013)](https://www.zotero.org/google-docs/?Tgu0cj) reduces subsequent consumption of the trained alcohol stimuli. Meta-analyses of food and alcohol stimuli [(Jones et al., 2016)](https://www.zotero.org/google-docs/?haeffi) and food stimuli [(Yang et al., 2019)](https://www.zotero.org/google-docs/?Gt3Bqz) also evidence for IID and suggest that go/no-go may be particularly effective for IID [(Yang et al., 2019)](https://www.zotero.org/google-docs/?L29h2k). A multi-experiment preregistered study ([Chen, Veling, Dijksterhuis, & Holland, 2016)](https://www.zotero.org/google-docs/?21abcD) suggests that IID is more effective when inhibition trials to be rare ([Chen et al., 2016)](https://www.zotero.org/google-docs/?e884Zw). However, some work urges caution on the efficacy of IID for behavior change [(Schonberg et al., 2014; Smith, Dash, Johnstone, Houben, & Field, 2017)](https://www.zotero.org/google-docs/?1KhPhr).

Perhaps most excitingly, work by Wessel and colleagues ([Wessel, O’Doherty, Berkebile, Linderman, & Aron, 2014)](https://www.zotero.org/google-docs/?be5l6T) showed that if arbitrary stimuli (colored shapes) are associated with specific monetary values, then a subset of these arbitrary stimuli are associated with inhibition, subjects selectively devalue them. This study is exciting because it suggests that IID may not only apply to stimuli with *a priori* value (e.g., palatable foods) but may be an effective strategy to experimentally devalue any stimuli. Once a stimulus is associated with inhibition, it appears to be subjectively less valuable, and therefore less likely to be used and (possibly) abused.

In two direct replications of their Experiment 2, our research group was unable to replicate the [Wessel et al](https://www.zotero.org/google-docs/?VmsopG). [(2014)](https://www.zotero.org/google-docs/?QXlklp) IID results. Additionally, in seven conceptual replications we were unable to replicate IID. This brings into question the robustness of IID, at least as implemented by Wessel and colleagues. In the following section, we lay out the results of nine pilot studies. We lay the two direct replications out in detail, as they are most directly pertinent to the replicability of Wessel et al. [(2014)](https://www.zotero.org/google-docs/?I7cVxP). We also briefly highlight the evidence for IID in the seven conceptual replications. These studies were not robustly powered and they varied on several dimensions, so we present these as pilot research. We propose to supplement this pilot work with robustly powered registered, direct replications of [Wessel et al. (2014](https://www.zotero.org/google-docs/?aXJtXC)) to be performed at three different universities. In our direct replications and our proposed experiments, our code, procedure, exclusion, and analysis plans follow [Wessel et al., 2014](https://www.zotero.org/google-docs/?rD9kxO) unless otherwise specified. All 9 existing datasets are openly available on the open science framework (<https://osf.io/x38aj/>) and the 3 new datasets will be made openly available at or before publication.

**Wessel et al., 2014 Experiment 2 Details**

Wessel et al. [(2014)](https://www.zotero.org/google-docs/?QRAFtA) used a 3-phase procedure in which a set of 8 colored shape stimuli were associated with different reward values (phase 1), then a subset of the colored shape stimuli occurred on trials with stop signals and another subset did not (phase 2), and finally subjects are asked to rate each shape stimulus in an auction procedure (phase 3). In Experiments 1 and 2 of the Wessel et al. [(2014)](https://www.zotero.org/google-docs/?rDiAxs) paper as well as in a more recent and larger group [(Wessel, Tonneson, & Aron, 2015)](https://www.zotero.org/google-docs/?Q4sNE4), they showed that subjects assigned lower monetary values to shapes that were associated with inhibition than shapes that were not. We focused on [(Wessel et al., 2014)](https://www.zotero.org/google-docs/?iz1rZM) Experiment 2 because it involved a minor change to Experiment 1 that involved more opportunities to correct subjects who implement a waiting-strategy of slowed responses. We first attempted two direct replications of [Wessel et al. (2014](https://www.zotero.org/google-docs/?WoBUrn)) experiment 2, the second of which was pre-registered on the Open Science Framework before acquiring the data (<https://osf.io/kq5xd>).

**Pilot Studies**

**Method**

**Direct Replication Replication 1 (N = 31)**

**Subjects.** 31 subjects (16 female) were recruited from the Stanford Psychology Department Sona Paid subject pool for a single 80-minute testing session. Subjects were paid $12 + the reward from a randomly selected 5 trials from phase 1. Mean age was 25.6 (+/- 5.9) years. 18 of the 31 subjects were Stanford University students.

**Procedure.** The experiment was run in PsychoPy (<http://www.psychopy.org/>) and matched the procedure from Wessel et al. (2014) Experiment 2. Tones were presented through noise cancelling headphones.

The following is an explanation of the verbal feedback given to the subject by the experimenter. In phase 2, the subject was required to exit the subject room to alert the experimenter that they have completed a block. At this time, the Experimenter would check the feedback to ensure that RTs were remaining fast (400-650ms), probability of responding given a stop signal was close to .5 (.4-.6) and SSD was > 100ms. If subjects deviated from these ideal values specified by Wessel et al. [(2014)](https://www.zotero.org/google-docs/?Oxdn2V), they would be encouraged to focus on balancing the important and competing goals of going and stopping. A predominant way in which subjects would deviate from these goals was by having RT longer than 650ms, and when this happened subjects would be told to try to reduce their RTs to below 650ms while balancing going and stopping. The subject did not report to the experimenter in between blocks in phases 1 or 3.

**Direct Replication 2 (N = 30)**

**Subjects.** 31 subjects (19 female) were recruited from the Stanford Psychology Department Sona Paid subject pool for a single 80-minute testing session. One subject was removed for failure to follow instructions. Subjects were paid $12 + the reward from a randomly selected 5 trials from phase 1. Mean age was 21.4 years with a standard deviation of 5.7 years (one subject did not report age). 22 of the 30 subjects were Stanford University students.

**Procedure.** The experiment was run with the Psychophysics Toolbox in Matlab. We used the following code provided by Jan Wessel:<https://github.com/janwessel/stopdeval>. The stimuli and timing matched those presented in Wessel et al. (2014) Experiment 2. Tones were presented through noise cancelling headphones.

We followed the procedure from Wessel et al. [(2014)](https://www.zotero.org/google-docs/?oIPwQJ) Experiment 2 exactly except in the following ways. We used the updated debriefing questions in the readme file of the code in Jan Wessel’s Github specified above. The questions were as follows: 1. Did you notice anything in the reaction time experiment?, 2. Did you think you were better for some symbols than for others?, 3. Some symbols were worth more than others, could you try to name any symbols you thought had particularly low or high values?, 4. Did you notice anything in the stop-signal experiment?, 5. Did you think some symbols were paired with stopping more often than others?, 6. Some symbols were paired with stopping more often. Could you try to name them?

We used the same instructional procedure as in Direct Replication 1 (i.e., having subjects come out after each short block in phase 2 so that the experimenter can intervene if their performance deviates from the values specified by [Wessel et al., 2014)](https://www.zotero.org/google-docs/?pNK4cq).

**Conceptual Replication 1 (N = 28)**

Detailed methods are presented in the supplement. In brief, following up on work by [Guitart-Masip et al., 2012](https://www.zotero.org/google-docs/?trqoIk), we presented arbitrary color patches that were associated with either acting to gain reward a monetary reward, acting to avoid monetary punishment, withholding an action to gain a monetary reward, or withholding an action to avoid a monetary punishment. This replaced the initial Wessel et al. (2014) phase of associating specific stimuli with value. In a subsequent phase, certain of these colors were associated with going and others with stopping. Specifically, go stimuli were presented in 1 of 5 colors and stop stimuli were presented in a different 1 of 5 colors on any given trial. Finally, subjects completed a two-alternative forced choice in which they were instructed to choose their preferred color (see Supplement for details). We focus on these forced choice results. This forced choice procedure replaced the auction procedure from Wessel et al., 2014. More go than stop color choices is taken as evidence of IID.

**Conceptual Replications 2-5 (N’s = 33, 23, 19, and 65 respectively)**

We detail methods in the supplement. To briefly summarize, in a line of four studies, we investigated whether having a greater number of associations, or instances ([Logan, 1988)](https://www.zotero.org/google-docs/?Tgts8e), of a stimulus being associated with stopping may increase the IID effect. Like Conceptual Replication 1, there was no initial valuation phase. Go stimuli were presented in one of three colors: the frequent go color (~76% of all go trials), the intermediate go color (~14% of all go trials), and the infrequent go color (~10% of all go trials). Stop stimuli were presented in one of three other colors: the frequent stop color (~57% of all stop trials), the intermediate stop color (~29% of all stop trials), and the infrequent stop color (~14% of all stop trials). As in Conceptual Replication 1, there was a final forced choice section where subjects had to choose between pairs of colors. Subjects choosing the stimuli associated with going more frequently than those associated with stopping would be consistent with IID. Conceptual Replication 2 was done at Stanford University. Conceptual Replications 3-5 were done online on Amazon Mechanical Turk. Conceptual Replications 2 and 5 were pre-registered here as NSSF and NSSF2, respectively: <https://osf.io/rgw2z/>.

**Conceptual Replication 6 (N = 34)**

We investigated whether we were not observing IID because subjects failed to stop on approximately half of the stop trials in our preceding IID studies. One important independent variable in the stop-signal task is the stop-signal delay (SSD), which is the delay between the presentation of the go stimulus and the stop signal. When SSD is long, subjects tend to fail to stop their responses, as the go process is near completion when the stop process starts. When SSD is short, subjects tend to succeed at stopping, as the go process has only recently begun when the stop process starts [(Logan & Cowan, 1984)](https://www.zotero.org/google-docs/?hoOAiW). In stop-signal studies, it is common to adjust the stop-signal delay up after stop success and down after stop failure to yield a 50% probability of a stop signal [(Levitt, 1971)](https://www.zotero.org/google-docs/?C3Zt69). This is the procedure that we used for our preceding 7 IID pilot samples. In order to test whether we would observe IID if subjects almost always stopped to a certain stop color, we used two stop colors presented at two different SSDs: one with the usual 1 up 1 down tracking algorithm and the other with a 0 SSD. On trials in which the 0 SSDs stop color appears, subjects should almost always succeed at stopping (stop accuracy was 97.5% in this sample at 0 SSD). Therefore, if IID requires a consistent mapping between color, the stop process, and the successful stop outcome then we should see IID for this 0 SSD stop color but not for the tracked SSD. However, this did not prove to be the case, as subjects chose the 0 SSD shape more than either the tracked SSD shape or the go shapes, so we collapsed across the 0 SSD and tracked SSD conditions for the conceptual replication results in Figure 2. We pre-registered this study here: <https://osf.io/5cdxf/>.

**Conceptual Replication 7 (N = 35)**

We investigated whether a go/no-go task [(Donders, 1868/1969](https://www.zotero.org/google-docs/?XKryki)) may be a more effective task to yield IID. Some work suggests that the stop-signal task involves a different type of inhibition, namely *cancellation,* than go/no-go, namely *restraint* [(Schachar et al., 2007)](https://www.zotero.org/google-docs/?elTPD1). Perhaps restraint is more effective at yielding IID. Additionally, as mentioned above, some existing work suggests that the go/no-go task is more efficacious than the stop-signal task to yield IID [(Yang et al., 2019)](https://www.zotero.org/google-docs/?zWJ5NA). Therefore, instead of associating specific stimuli with the go or the stop stimulus, we attempted to associate specific stimuli with the go or the no-go signal.

**Results**

**Direct Replications 1 and 2: Exclusion Criteria and the Role of Explicit Knowledge**

Wessel et al., (2014) used exclusion criteria that entailed excluded subjects who explicitly learned the contingency between specific stimuli and stopping. According to their criteria, 11 of 31 subjects would be excluded in our Direct Replication 1 and 18 of 30 subjects would be excluded in Direct Replication 2. We present the stopping results in Table 1 with all subjects and the exclusive subset that only includes the 32 subjects passing the [Wessel et al., 2014](https://www.zotero.org/google-docs/?zcDR05) criteria). We focus on the auction results in text (the p-values preceding the “/” include all subjects and the p-values following the “/” are the exclusive subset. See *Exclusion Criteria and the Role of Explicit Knowledge* in Supplement for more details.

**Auction Phase (Phase 3)**

**Direct Replication 1.** We ran a 2 (Stopping: Stop vs. Non-Stop Shapes) x 4 (Value: .5 dollars, 1 dollar, 2 dollars, and 4 dollars) rmANOVA that mirrored the results presented in Wessel et al. (2014) Figure 2 (see Figure 1a). There was no main effect of Stopping (*p* = .9/*p* = .58), a significant main effect of value (*p* < .001/*p* = .002), and no interaction (*p* = .59/*p* = .66). There was a similar number subjects who assigned lower values to stop than non-stop stimuli (15 of 31/11 of 20) as subjects who assigned lower values to non-stop than stop stimuli (16 of 31/9 of 20). This suggests that subjects learned the different values from phase 1, but there was no evidence for IID or an interaction of IID with value.

**Direct Replication 2.** We again ran a 2 (Stopping: Stop vs. Non-Stop Shapes) x 4 (Value: .5 dollars, 1 dollar, 2 dollars, and 4 dollars) rmANOVA that mirrored the results presented in Wessel et al. (2014) Figure 2 (see Figure 1b). There was no main effect of Stopping (*p* = .97/*p* = .29), a significant main effect of value (*p* < .001/*p* < .001), and no interaction (*p* = .41/*p* = .50).

As in Direct Replication 1, there was a similar number subjects who assigned lower values to stop than non-stop stimuli (14 of 30/3 of 12) as subjects who assigned lower values to non-stop than stop stimuli (16 of 30/9 of 12).Taken together with Direct Replication 1, we do not have evidence for IID in our two direct replications.

**Combined Direct Replications 1 and 2.** Almost half of our subjects did not pass the exclusion criterion proposed by [Wessel et al., 2014](https://www.zotero.org/google-docs/?IkdRzD), often because our subjects learned the explicit contingencies between specific stimuli and inhibition. We evaluated the 32 subjects from Direct Replications 1 and 2 who passed the exclusion criteria and the mean stopping-induced devaluation in this group was -.09, meaning that stop stimuli were valued slightly *higher* than go stimuli. Therefore, there was no evidence for IID even in our group of Direct Replication subjects who passed the exclusion criteria from [Wessel et al., 2014](https://www.zotero.org/google-docs/?HQh2Oe).

**Combined Analyses of 2 Direct Replications and 7 Conceptual Replications**

The primary question of interest is whether stimuli or stimulus features that are associated with inhibition have subsequently reduced value. To test this question, we computed a devaluation score for the two direct replications by computing the total auction value for the go stimuli and dividing it by the total value for both the stop and go stimuli. For the conceptual replications in which we replaced the auction by a forced choice, we computed the proportion of go choices when go stimuli were pitted against stop/no-go stimuli. Values above .5 signify a larger auction value or more frequent choices for go stimuli than no-go stimuli, which would be consistent with IID. Across these 9 studies and 298 subjects, the average value was .502 and the weighted average was .498. 150 of the 298 subjects had values <.5. Focusing exclusively on the two Direct Replications, the values were .499 and .503 for an average of .501. In short, there was no evidence for IID (see Figure 2).

**Proposed Registered Replication**

In our 9 pilot studies, we found no evidence for IID in direct and conceptual replications of Wessel et al. (2014). However, we encourage caution before drawing strong conclusions because of the following factors. First, as pointed out in direct communication with Jan Wessel and Adam Aron, two authors from [Wessel et al., 2014](https://www.zotero.org/google-docs/?YORgwe), the samples of subjects in our direct replications learned the explicit contingencies between stimuli and stopping more often than their original sample. This may influence the results, though some evidence suggests that explicit knowledge may increase IID [(Wessel et al., 2015)](https://www.zotero.org/google-docs/?tdm76z). If we remove subjects with explicit knowledge (see Combined Direct Replication 1 and 2 section in the main text and Exclusion Criteria and the Role of Explicit Knowledge in supplement) we are left with only 32 subjects total. Additionally, our seven conceptual replications changed multiple factors that may have influenced or reduced effects. We removed the phase where we associated specific stimuli with specific value. Instead, we attempted to change value of arbitrary features without the presence of this initial phase, which may be essential for IID. We also embedded color in the go or stop stimuli themselves, instead of as an accessory feature on go or stop trials. Finally, we replaced the auction phase with a forced choice procedure.

In order to overcome these mitigating factors, we propose to run a large-scale, pre-registered, multi-site direct replication of Wessel et al., 2014 Experiment 2, following the same procedure as our previous Direct Replication 2 (<https://osf.io/x38aj/>). The 9 pilot samples were small to medium in size (N’s = 19-68). In an *a priori* power analysis, we found that if we assume a small effect size (cohen’s d = .2), a sample size of 79 subjects achieves power of .95 for the main effect comparing the auction value of stop and non-stop stimuli (see openly available code here: <https://github.com/bissettp/PowerAnalysisIID>). This main effect was the primary evidence for IID in [Wessel et al. (2014](https://www.zotero.org/google-docs/?CkQLK6)), and a significantly lower auction value for stimuli associated with stopping would be evidence for IID. None of our existing samples achieve this power, and our two direct replications do not approach this power.

We will recruit subjects until we have 79 who satisfy the following exclusion criteria (see <https://osf.io/x38aj/> for exclusion scripts). In general, we will follow the [(Wessel et al., 2014)](https://www.zotero.org/google-docs/?ML12h5) exclusion criteria that we detail in the Exclusion Criteria and the Role of Explicit Knowledge section of the Supplement, with the following exceptions. Instead of calculating mean SSRT, we will calculate integration SSRT with replacement, as suggested by a recent consensus guide for the stop-signal task [(Verbruggen et al., 2019)](https://www.zotero.org/google-docs/?Na3UYz). Additionally, when computing SSRT, we will use the reaction times to stop shapes as the underlying go distribution on stop trials, as we did in the Supplement. We found that RTs tends to be slower to stop than non-stop shapes, so no-stop-signal RTs on only stop shapes should act as a better, more specific estimate of the underlying go distribution on stop trials.Subjects will be excluded if they took part in any of our previous IID samples. We will complete supplemental analyses that investigates whether IID occurs in excluded subjects and if included and excluded subjects are combined.

We propose to acquire these 79 usable subjects at each of our three independent sites (Stanford University, Tel Aviv University, and Missouri Western State University). We propose to acquire data from three sites in order to ensure that any of our findings are not idiosyncratic to Stanford University. Both of our direct replications came from Stanford University and we would like to ensure that our results generalize across populations. We believe that this replication effort is an essential complement to recent meta-analyses on IID (e.g., [Jones et al., 2016; Yang et al., 2019)](https://www.zotero.org/google-docs/?Jc27Nm), which can be driven by publication bias. In order to eliminate the possibility of coding discrepancies, we will use the Matlab code provided by Wessel and colleagues (<https://github.com/janwessel/stopdeval>).

Our prediction, guided by our pilot data, is that we will not find inhibition induced devaluation in any of the three new samples. We will follow the same primary analysis as in [(Wessel et al., 2014)](https://www.zotero.org/google-docs/?uNU6tg) and our two main direct replications that implements a 2 (Stopping: Stop vs. Non-Stop Shapes) x 4 (Value: .5 dollars, 1 dollar, 2 dollars, and 4 dollars). A directional main effect of stopping with smaller auction values for stop than non-stop shapes would indicate IID. We anticipate strong evidence for the null hypothesis of no IID, which we will evaluate with equivalence testing [(Lakens, 2017)](https://www.zotero.org/google-docs/?zRsDPp) and Bayes Factors [(Rouder, Morey, Speckman, & Province, 2012)](https://www.zotero.org/google-docs/?NDz9I0). This result would be evidence against the Wessel et al. (2014) IID result.

References

[Bissett, P. G., & Logan, G. D. (2011). Balancing cognitive demands: Control adjustments in the stop-signal paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 392–404. https://doi.org/10.1037/a0021800](https://www.zotero.org/google-docs/?fwpYaU)

[Bissett, P. G., & Logan, G. D. (2013). Stop before you leap: Changing eye and hand movements requires stopping. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(4), 941–946. https://doi.org/10.1037/a0033049](https://www.zotero.org/google-docs/?fwpYaU)

[Camalier, C. R., Gotler, A., Murthy, A., Thompson, K. G., Logan, G. D., Palmeri, T. J., & Schall, J. D. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research*, *47*(16), 2187–2211. https://doi.org/10.1016/j.visres.2007.04.021](https://www.zotero.org/google-docs/?fwpYaU)

[Cavanagh, J. F., Eisenberg, I., Guitart-Masip, M., Huys, Q., & Frank, M. J. (2013). Frontal Theta Overrides Pavlovian Learning Biases. *Journal of Neuroscience*, *33*(19), 8541–8548. https://doi.org/10.1523/JNEUROSCI.5754-12.2013](https://www.zotero.org/google-docs/?fwpYaU)

[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](https://www.zotero.org/google-docs/?fwpYaU)

[Donders, F. C. (1868). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.](https://www.zotero.org/google-docs/?fwpYaU)

[Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*(3), 389–413. https://doi.org/10.1037/0096-3445.136.3.389](https://www.zotero.org/google-docs/?fwpYaU)

[Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, 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*(1), 154–166. https://doi.org/10.1016/j.neuroimage.2012.04.024](https://www.zotero.org/google-docs/?fwpYaU)

[Houben, K., Nederkoorn, C., Wiers, R. W., & Jansen, A. (2011). Resisting temptation: Decreasing alcohol-related affect and drinking behavior by training response inhibition. *Drug and Alcohol Dependence*, *116*(1–3), 132–136. https://doi.org/10.1016/j.drugalcdep.2010.12.011](https://www.zotero.org/google-docs/?fwpYaU)

[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. https://doi.org/10.1016/j.appet.2015.11.013](https://www.zotero.org/google-docs/?fwpYaU)

[Jones, A., & Field, M. (2013). The effects of cue-specific inhibition training on alcohol consumption in heavy social drinkers. *Experimental and Clinical Psychopharmacology*, *21*(1), 8–16. https://doi.org/10.1037/a0030683](https://www.zotero.org/google-docs/?fwpYaU)

[Lakens, D. (2017). Equivalence Tests: A Practical Primer for *t* Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177](https://www.zotero.org/google-docs/?fwpYaU)

[Lawrence, N. S., Verbruggen, F., Morrison, S., Adams, R. C., & Chambers, C. D. (2015). Stopping to food can reduce intake. Effects of stimulus-specificity and individual differences in dietary restraint. *Appetite*, *85*, 91–103. https://doi.org/10.1016/j.appet.2014.11.006](https://www.zotero.org/google-docs/?fwpYaU)

[Levitt, H. (1971). Transformed up-down methods for psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.](https://www.zotero.org/google-docs/?fwpYaU)

[Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica*, *60*, 193–210.](https://www.zotero.org/google-docs/?fwpYaU)

[Logan, Gordon D. (1988). Toward an Instance Theory of Automatization. *Psychological Review*, *95*(4), 492–527.](https://www.zotero.org/google-docs/?fwpYaU)

[Logan, Gordon D, & Cowan, W. B. (1984). On the Ability to Inhibit Thought and Action: A Theory of an Act of Control. *Psychological Review*, *91*(3), 295–327.](https://www.zotero.org/google-docs/?fwpYaU)

[Nielsen, L., Riddle, M., King, J. W., Aklin, W. M., Chen, W., Clark, D., … Weber, W. (2018). The NIH Science of Behavior Change Program: Transforming the science through a focus on mechanisms of change. *Behaviour Research and Therapy*, *101*, 3–11. https://doi.org/10.1016/j.brat.2017.07.002](https://www.zotero.org/google-docs/?fwpYaU)

[Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001](https://www.zotero.org/google-docs/?fwpYaU)

[Schachar, R., Logan, G. D., Robaey, P., Chen, S., Ickowicz, A., & Barr, C. (2007). Restraint and Cancellation: Multiple Inhibition Deficits in Attention Deficit Hyperactivity Disorder. *Journal of Abnormal Child Psychology*, *35*(2), 229–238. https://doi.org/10.1007/s10802-006-9075-2](https://www.zotero.org/google-docs/?fwpYaU)

[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](https://www.zotero.org/google-docs/?fwpYaU)

[Smith, J. L., Dash, N. J., Johnstone, S. J., Houben, K., & Field, M. (2017). Current forms of inhibitory training produce no greater reduction in drinking than simple assessment: A preliminary study. *Drug and Alcohol Dependence*, *173*, 47–58. https://doi.org/10.1016/j.drugalcdep.2016.12.018](https://www.zotero.org/google-docs/?fwpYaU)

[Veling, H., Aarts, H., & Stroebe, W. (2013). Using stop signals to reduce impulsive choices for palatable unhealthy foods. *British Journal of Health Psychology*, *18*(2), 354–368. https://doi.org/10.1111/j.2044-8287.2012.02092.x](https://www.zotero.org/google-docs/?fwpYaU)

[Verbruggen, F., Aron, A. R., Bissett, P. G., Brockett, A. T., Brown, J. W., Chamberlain, S. R., … Thakkar, K. N. (2019). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *ELife*, *8*, e46423.](https://www.zotero.org/google-docs/?fwpYaU)

[Wessel, J. R. (2018). Prepotent motor activity and inhibitory control demands in different variants of the go/no-go paradigm. *Psychophysiology*, *55*(3), e12871. https://doi.org/10.1111/psyp.12871](https://www.zotero.org/google-docs/?fwpYaU)

[Wessel, J. R., O’Doherty, J. P., Berkebile, M. M., Linderman, D., & Aron, A. R. (2014). Stimulus devaluation induced by stopping action. *Journal of Experimental Psychology: General*, *143*(6), 2316–2329. https://doi.org/10.1037/xge0000022](https://www.zotero.org/google-docs/?fwpYaU)

[Wessel, J. R., Tonneson, A. L., & Aron, A. R. (2015). Stimulus devaluation induced by action stopping is greater for explicit value representations. *Frontiers in Psychology*, *6*, 1640.](https://www.zotero.org/google-docs/?fwpYaU)

[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*. https://doi.org/10.1111/obr.12916](https://www.zotero.org/google-docs/?fwpYaU)

Table 1. Direct Replication Stopping Results.

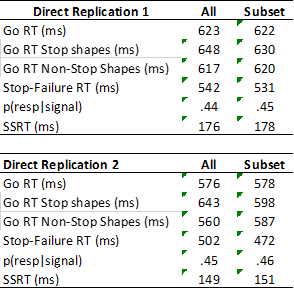
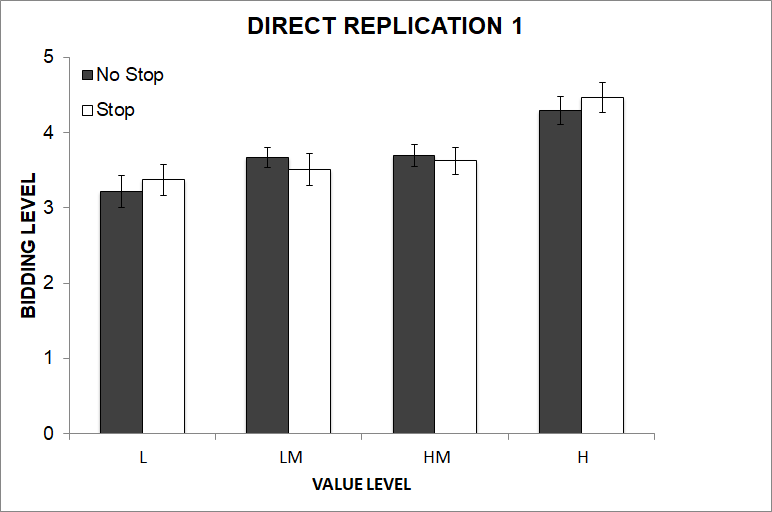


Table 1 Caption: p(resp|signal) means probability of responding given a stop signal.

Figure 1. Direct Replication Results.

Figure 1 Caption: Results presented in equivalent fashion to [Wessel et al., 2014](https://www.zotero.org/google-docs/?h3gVMc) Figure 2. We paraphrase their caption: Bidding data from the valuation phase in Direct Replication 1 (1a) and Direct Replication 2 (1b). Bidding levels range from low bids (1) to high bids (6), plotted by the actual value of each shape from the phase 1 learning (L = low, LM = low-medium, HM= high-medium, H = high). “No stop” and “Stop” refer to whether the shapes were paired with stop-signals or not in Phase 2 treatment. Error bars denote the standard error of the mean across subjects.

1a.



1b.

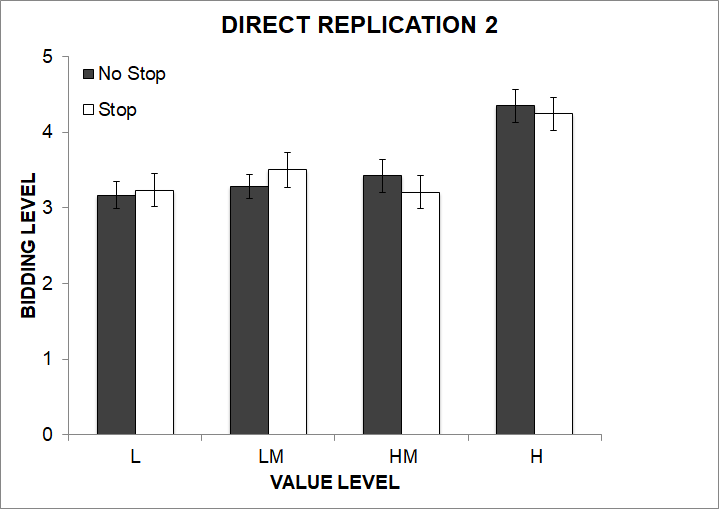
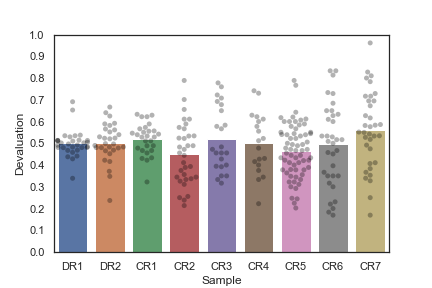


Figure 2. No devaluation across 9 pilot samples.

Figure 2 caption: .5 indicates equal auction value (DR1 and DR2) or choice probability (CR1-7) for stimuli associated with going and stopping. Therefore, values above .5 indicate inhibition induced devaluation. Average of 9 bars is .502 and weighted average all subjects is .498. Barplots indicate the mean for each of the 9 samples and each dot is the mean from an individual subject. DR = direct replication. CR = conceptual replication.



Supplement

**Exclusion Criteria and the Role of Explicit Knowledge**

We initially adopted the exclusion criteria specified by [Wessel et al., 2014](https://www.zotero.org/google-docs/?zKCXcT), which we understand to be: (1) stop-failure RTs >= no-stop-signal RT, (2) Any reportable explicit knowledge in phase 2, (3) SSRT < 100 ms, (4) the absolute value of the main effect of stopping devaluation > 1.5x the interquartile range. However, applying these criteria to both studies resulted in a much higher rate of exclusion than reported by Wessel et al. [(2014, 2015](https://www.zotero.org/google-docs/?ku2Yf7)), mainly resulting from many subjects having reportable explicit knowledge about the contingency between shapes and stopping in phase 2.

In Direct Replication 1, 4 subjects were full explicit learners of the stopping contingency (i.e., could name all 4 shapes associated with stopping) and another 3 were partial explicit learners (i.e., could name at least 1 shape associated with stopping). The 4 full explicit learners also had SSRTs < 100ms as computed by the mean method. We also computed SSRT by the integration method using only no-stop-signal RT to stop shapes. Many subjects had longer no-stop-signal RTs to stop shapes than non-stop shapes. This suggests that the underlying go process on stop trials is underestimated by the full no-stop-signal distribution, and is better approximated by the distribution of no-stop-signal RTs to stop shapes. When we calculate SSRT in this second way, only one subject had SSRT < 100 ms and this subject was an implicit learner in phase 2. This subject was also excluded from the “exclusive group”. One phase 2 explicit leaner had stop-failure RT longer than their overall mean no-stop-signal RT, but this does not seem to have resulted from a violation of the race model, as their no-stop-signal RT to stop shapes was longer than their stop-failure RT, so we did not exclude this subject. We also adopted the same outlier exclusion for phase 3 (i.e., the absolute value of the main effect of stopping devaluation > 1.5 X the interquartile range), which excluded an additional 3 subjects. 14 of the 31 subjects were partial explicit learners in phase 1, another 3 were full explicit learners, and the remaining 14 showed no explicit knowledge of phase 1. No subjects were excluded on phase 1 learning.

In Direct Replication 2, 7 subjects were full explicit learners and another 10 subjects were partial explicit learners of the stopping contingency. 4 of those full explicit learners and 3 of those partial explicit learners had SSRTs < 100ms using the mean method. When we instead computed SSRT by the integration method using only no-stop-signal RTs to stop shapes, four subjects had SSRTs < 100 ms, one phase 2 explicit learners, two phase 2 partial explicit learners, and one phase 2 implicit learners. Three subjects had stop-failure RT longer than their mean no-stop-signal RT (two phase 2 partial explicit learners and one phase 2 implicit learner), but none had stop-failure RTs longer than their mean no-stop-signal RT to stop shapes, so they were not excluded. 18 of the 30 subjects were partial explicit learners in phase 1, another 1 was a full explicit learner, and the remaining 11 showed no explicit knowledge of phase 1.

Therefore, if we exclude based upon the same criteria as Wessel et al. (2014), with the only difference being SSRT estimates were based upon the integration method using no-stop-signal RTs to stop shapes as the underlying go distribution, 11 of 31 subjects were excluded in Direct Replication 1 and 18 of 30 subjects were excluded in Direct Replication 2. We presented the results in the main text both with and without these exclusions, with the inclusive group mean preceding a “/” and the exclusive group following a “/”.

**Conceptual Replication 1 Method**

**Subjects.** 31 subjects were recruited from the Stanford Psychology Department Sona Paid subject pool for a single 60-minute testing session. Subjects were paid a base rate (and minimum) of $12 plus the bonus they earn in phase 1 learning for this ~60 minute task. Three subjects were removed because they were above our upper age threshold of 40, and for one of these subjects, no data was collected because a pop-up interfered with the experiment. For the remaining 28 subjects, 12 were males and 16 were female. The average age was 25.1 years (SD = 4).

**Procedure.** In phase 1, subjects were presented with 1 of 8 different colored squares in the middle of the screen on each trial. Two colors were associated with each of the following action-valence task rules: 1) act to gain reward, 2) act to avoid punishment, 3) withhold to gain reward, and 4) withhold to avoid punishment. We used two colors per association so that half of the colors were an irrelevant feature of the stop stimulus in the second phase, and the other half were an irrelevant feature of the go stimulus in the second phase. In order to ensure that associations were learned by all subjects in a relatively short period of time, subjects were explicitly instructed on the color-action-valence associations. Specifically, subjects were told to press the spacebar for colors associated with acting and withhold any response for colors associated with inaction. They were also told which color-action combinations would result in gaining reward and which color-action combinations would result in avoiding punishment.

After instruction, all subjects continued onto the main phase 1 “learning phase” in which performing the correct action to a specific color deterministically resulted in receiving the advantageous associated outcome (i.e., receive a 25 cent reward or the absence of a 25 cent punishment, depending on the color) and performing the incorrect action to the specific color deterministically resulted in the less advantageous associated outcome (i.e., the absence of a 25 cent reward or receive a 25 cent punishment). This phase is composed of 3 blocks of 24 trials, where each color is presented 3 times per block. Trial length is fixed at 3500ms (500ms fixation, 1500ms stim presentation, and 1500ms reward presentation).

In phase 2, subjects first practiced and then completed the stop task. The practice serves two purposes: (1) familiarize them with the stop signal task, and (2) introduce subjects to two neutral colors (one for go and one for stop) that were not part of phase 1 to counteract novelty and surprise effects when they are shown these colors in the main stop task. Introducing neutral colors allows us to compare phase 2 stop performance and phase 3 subjective values to two colors without action-valence task rules.

Practice for the stop task was divided in two parts. The first did not include stop signals (12 trials) and was intended to familiarize subjects with the stimulus response mapping of four stimuli (circle, rhombus, triangle, pentagon) to two responses (Z or M key). During practice, the go stimuli were presented in fifth go color (that did not have an action-valence task rule from phase 1). Trials include a 500ms fixation cross, 850 ms go stimulus presentation, and a 1000ms blank-screen ITI. The subject had a total of 1850ms to respond.

Then subjects were instructed about stop signals before the second part of practice. This second practice was also 12 trials, with 4 stop trials. Stop stimuli were a seven pointed star presented around the go stimulus, and those seven pointed stars occurred in the fifth stop color (that did not have an action-valence task rule from phase 1). Stop stim were presented for 500ms at a 250ms fixed SSD.

After practice, subjects moved on to the main stop task. The go stimuli were presented equally across 5 colors (the 4 action-valence associations and the 1 neutral go color) and the stop stimuli were presented equally across the other 5 colors (the 4 action-valence associations and the 1 neutral stop color). This phase was composed of 6 blocks with 75 trials in each block, where 1/3 of all trials were stop trials. SSD varied within a range of 0-850ms according to a 1 up 1 down tracking algorithm (i.e., increased 50ms after successful stopping and decrease 50ms after failed stopping, [Levitt, 1971)](https://www.zotero.org/google-docs/?vbBJLw). Initially we were interested in the interaction between action, valence, and stopping, but in this manuscript we focus on the comparison between go and stop colors. In this design, the go colors were always associated with the go process but not with an overt response, as subjects responded on ~5/6 of all trials in which the go color was presented (⅔ of all trials werego trials and an additional half of the ~⅓ of stop trials still resulted in an overt response). Similarly, in this design, the stop colors were always associated with the stop process but not with withholding a response, as subjects successful stopped on ~½ of stop trials.

Phase 3 of the experiment involved a 2-alternative forced choice to probe subjective valuation of the 10 colors. On each trial, subjects were presented with 2 colored squares, one on the left and the other on the right side of the screen. Subjects were asked to choose the color that they found “more rewarding”, in line with [Cavanagh, Eisenberg, Guitart-Masip, Huys, & Frank, 2013](https://www.zotero.org/google-docs/?hpHcP8), by pressing either the left or right arrow key. There were a total of 90 trials, with each color pair evaluated twice. Trial length was fixed at 3500ms: 500ms for fixation and 3000ms stim presentation.

**Conceptual Replications 2-5**

**Subjects.** For Conceptual Replication 2, 38 subjects were recruited from the Stanford Psychology Department Sona Paid subject pool for a single 60-minute testing session. Five were excluded for being over the age of 40, resulting in 33 usable subjects. For these 33 subjects, 10 were male and 23 were female. The average age was 24.9 years (SD = 3.5). For Conceptual Replications 3-5, 32, 32, and 100 subjects, respectively, were recruited from Amazon Mechanical Turk. In Conceptual Replication 3, two subjects were removed for below chance go accuracy in at least condition, 5 subjects were removed for low stop whose probabilities of successful stopping fell outside the 95% confidence interval of .5 probability of stopping to a stop signal, and 2 subjects were removed for omitting responses on forced choice questions (mean omission rate across excluded subjects = 96%) for responding on very few of their forced choice questions (<10%), leaving a N=23. The average age was 32.8 (SD=5.2) and there were 12 males and 11 females. In Conceptual Replication 4, 10 subjects were removed for omissions on forced choice questions (mean omission rate across excluded subjects = 100%), and three for age over 40, leaving N=19. The average age was 31.4 (SD = 5.5) and there were 10 males and 9 females. In Conceptual Replication 5, 31 subjects were removed for omitting responses on forced choice questions (mean omission rate across excluded subjects = 81%), and four for age over 40, leaving N=65 with mean age 31.3 (SD=5.5) with 34 males and 31 females.

**Procedure.** Conceptual Replications 2-5 shared features with Conceptual Replication 1 save the following differences. There was no initial phase that associated stimuli with action, inaction, reward, or punishment. The go stimulus shape appeared in one of three color conditions: frequent go (672 trials), intermediate go (126 trials), and infrequent go (84 trials). On 1/3 of trials, a stop stimulus (a star) appeared around the go shape. This stop stimulus appeared in one of three color conditions: frequent stop (168 trials), intermediate stop (84 trials), and infrequent stop (42 trials). The six colors were blue, red, pink, yellow, purple, and green. After 882 stop-signal task trials (294 stop trials), subjects completed 90 trials of forced choice in which the two of the six colors embedded in the initial stop-signal phase were pitted against each other and subjects had to choose which color patch they preferred.

**Conceptual Replication 6**

**Subjects.** 36 subjects were recruited from the Stanford Psychology Department Sona Paid subject pool for a single 60-minute testing session. Two subjects were excluded for age over 40, and for the resulting 34 subjects, there were 11 males, 22 females, and 1 N/A. Their average age was 25.5 years (SD = 4.5).

**Procedure.** Conceptual Replication 6 shared features with Conceptual Replications 2-5 with the exception of the details within the main text for Conceptual Replication 6.

**Conceptual Replication 7**

**Subjects.** 37 subjects were recruited from the Stanford Psychology Department Sona Paid subject pool for a single 60-minute testing session. Two subjects were excluded, one for age over 40 and one for responding to only 31 of 60 forced choice trials (all other subjects responded on at least 51 forced choice trials, resulting in a 35-subject sample. Of the usable 35 subjects, there were 10 males and 25 females, with an average age of 23.4 years (SD = 3.6).

**Procedure.** Conceptual Replication 7 was similar to Conceptual Replication 6 with the following exceptions. One shape appeared on the screen on a given trial: circle, rhombus, triangle, or pentagon. If either of two of those stimuli appeared, subjects were instructed to make a spacebar response, but if either of the other two stimuli appeared, subjects were instructed to refrain from responding. Each shape was displayed in a unique color from the following four: blue, green, red, and purple. Shapes and colors were counterbalanced across subjects. In the go/no-go phase, each of the two go shapes were presented 288 times, and each of the two stop shapes were presented 144 times. In the forced choice phase, each color was presented The four colors were blue, red, purple, and green. After the stop-signal task trials, subjects completed 60 trials of forced choice in which the two of the four colors embedded in the initial go/no-go phase were pitted against each other and subjects had to choose which color patch they preferred.