

## PSYCHOLOGICAL SCIENCE

## Severe violations of independence in response inhibition tasks

Patrick G. Bissett<sup>1\*</sup>, Henry M. Jones<sup>1</sup>, Russell A. Poldrack<sup>1</sup>, Gordon D. Logan<sup>2</sup>

The stop-signal paradigm, a primary experimental paradigm for understanding cognitive control and response inhibition, rests upon the theoretical foundation of race models, which assume that a go process races independently against a stop process that occurs after a stop-signal delay (SSD). We show that severe violations of this independence assumption at short SSDs occur systematically across a wide range of conditions, including fast and slow reaction times, auditory and visual stop signals, manual and saccadic responses, and especially in selective stopping. We also reanalyze existing data and show that conclusions can change when short SSDs are excluded. Last, we suggest experimental and analysis techniques to address this violation, and propose adjustments to extant models to accommodate this finding.

## INTRODUCTION

An essential adaptive feature of cognition and action is that they can be controlled and directed toward the achievement of goals. However, goals can change immediately and completely, as when a green light turns red while driving. In this case, the current course of action (accelerating) must be stopped. In such instances, a behavioral kill switch, known as response inhibition, is a necessary mechanism for control. Response inhibition is also a necessary part of modifying action as goals change (1–3). Thus, response inhibition is a fundamental control mechanism (4–6) that affords behavioral flexibility whenever actions need to stop or change in accordance with changing goals or environmental conditions.

A primary paradigm used to understand response inhibition is the stop-signal paradigm (7), which usually involves making a choice response to a go task and attempting to stop that response when an infrequent stop signal occurs after a stop-signal delay (SSD). This paradigm has grown in its use and is a common tool across various disciplines including neuroscience, psychiatry, psychology, and more [see (8), appendix 1]. The main theoretical vehicle for understanding and analyzing data from the stop-signal paradigm is the independent race model (7, 9), which assumes that a go process begins when the go stimulus occurs and races independently against a stop process that begins when the stop stimulus occurs. Stop finishing first results in stop success (i.e., no response); go finishing first results in stop failure (i.e., an overt response that escapes inhibition).

The independent race model provides a theoretical framework for understanding the stop-signal task and captures the main features of stop-signal performance. First, as the SSD increases, the probability of stop failure should increase (7) as longer SSDs handicap the race in favor of the go process. Second, stop-failure reaction time (RT) should be faster than no-stop-signal RT (i.e., responses on trials without a stop signal) because the stop process cuts off the upper tail of the go RT distribution, and stop-failure RT should decrease with decreasing SSD, because shorter SSDs will cut off more of the upper tail (7, 10, 11). These predictions tend to be supported by data.

The independent race model (7, 9) assumes context independence, which means that the finishing time distribution of the go process is the same whether or not a stop signal is presented [ $P(T_{go} < t \mid \text{no stop signal}) = P(T_{go} < t \mid \text{stop signal})$ ]. This assumption is essential to the race model account of the major dependent variables in the stop task: the probability of inhibiting response at each SSD, RTs on stop-failure trials, and the finishing time of the stop runner in the race, stop-signal RT (SSRT). Context independence allows the model to use the observed go distribution on no-signal trials as an estimate of the distribution of go runners on stop-signal trials. Violations of context independence invalidate the application of the race model to the data and call into question conclusions based on race-model measures (see the “Violations contaminate main dependent variables” section in the Supplementary Materials). The independent race model also assumes stochastic independence, which means that the finishing times of the go and stop processes are independent on a given trial [ $P(T_{go} < t_{go} \text{ AND } T_{stop} < t_{stop}) = P(T_{go} < t_{go}) \times P(T_{stop} < t_{stop})$ ]. The present manuscript focuses on the assumption of context independence.

Stop-failure RTs have been used to assess context independence (7, 12). If mean stop-failure RT is faster than mean no-stop-signal RT, then the context independence assumption is assumed to hold, and the race model is applied to the data (7, 12). If stop-failure RT is longer than the race model predicts, then the context independence assumption is violated. In severe violations, stop-failure RT may be longer than no-stop-signal RTs, which is not possible in the independent race model. We report extensive evidence for such severe violations below.

Violations of context independence are not only problematic for the original independent race model (7) but also a general issue for extant models of stopping. Parameterized versions of the independent race model that are intended as process models (9) or measurement models (13) also assume context independence. Models aimed at the underlying physiology also assume context independence up to the point at which stop and go processes interact (14, 15). These models assume that the parameters that generate the distributions of stop and go finishing times are context independent in that they take the same values on stop and go trials. Therefore, for all of these models, assuming context independence is essential to fitting the data and essential to estimating the distribution of SSRTs. Violations thus have important consequences for a broad range of theories in this domain.

Copyright © 2021  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>Department of Psychology, Stanford University, 450 Jane Stanford Way, Building 420, Stanford, CA 94305, USA. <sup>2</sup>Department of Psychology, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37240, USA.

\*Corresponding author. Email: pbissett@stanford.edu

Severe violations also challenge extant models that assume “trigger failures,” which are complete failures to detect, discriminate, or respond to the stop signal, as if the stop process was not “triggered” on that trial (7, 16). They are assumed to occur on a subset of trials regardless of SSD, as if the subject neglected the stop goal. When a trigger failure occurs, the go process races alone, resulting in stop-failure RTs that are just as long as no-stop-signal RTs. Therefore, trigger failure models can accommodate stop-failure RTs that are faster than no-stop-signal RTs (if only a subset of stop-failure trials results from trigger failure), like the independent race model (7, 9), or as long as no-stop-signal RTs (if all stop-failure trials result from trigger failure), but not stop-failure RTs that are longer than no-stop-signal RTs. Hence, the presence of stop-failure RTs that are longer than no-stop-signal RTs would also challenge extant models of trigger failures (16).

Some previous work suggested that context independence may be violated at short SSDs. Colonus and colleagues (17–20) showed that violations tended to occur at SSDs of <200 ms. Logan and Cowan (7) found violations in the same range of SSDs. However, this work only included a total of 18 subjects across five published studies. Some of this work relies on model fitting to generate predictions for independence to compare with observed data, and that is only feasible in studies with many trials per participant. Few studies include that many trials, which led us to develop an alternative method that would apply more generally.

We evaluated violations of context independence with a novel method that compared observed stop-failure RTs to observed no-stop-signal RTs from the immediately preceding trial. Crucially, the race model predicts that stop-failure RTs should be progressively faster than no-stop-signal RTs as SSDs decrease. Therefore, if observed stop-failure RT is found to be longer than observed no-stop RT at short SSDs, then this is evidence of a severe violation of context independence. In addition, by comparing stop failures to their immediately preceding no-stop-signal trials, we are able to eliminate the contamination of slow fluctuations in RT and SSD that occur throughout the experiment (21–23). Furthermore, this method should eliminate the influence of any response slowing including proactive slowing (21, 24), as the stop-failure trial and the immediately preceding trial should be similarly influenced by such slowing. In the following section, we apply this method to 860,568 trials obtained from 675 subjects across 25 conditions in 14 datasets (see Table 1) to evaluate the ubiquity of violations and to inform the mechanisms underlying them. We reveal violations at short SSDs across fast and slow conditions, auditory and visual stop signals, manual and saccadic responses, and selective stopping. The data and all analysis codes are openly available (<http://doi.org/10.5281/zenodo.4432816>).

## RESULTS

### Violations at short SSDs in experiments with fixed SSDs

A common procedure for determining SSD is the 1 up 1 down tracking procedure (25), but this can result in a small number of trials at short SSDs. To evaluate violations across SSD, we designed two experiments that used a broad range of fixed SSDs (100 to 500 ms in fixed SSD 1 and 0 to 500 ms in fixed SSD 2; see Table 1). In these experiments, a set of SSDs is presented in random order with the same number of trials at each SSD. This ensures that short SSDs are as probable as intermediate and longer SSDs.

To evaluate the prevalence of violations at short SSDs, we plot the violation (mean stop-failure RT from trial N minus mean no-stop-signal RT from trial N-1) against SSD. The positive values demonstrate evidence of a severe violation of the context independence assumption of the independent race model.

In our fixed SSD 1 study (see Fig. 1A), we observed violations at the shortest 100-ms SSD, but not at the longer  $\geq 200$ -ms delays. To sample short SSDs with greater granularity, we included 11 SSD values from 0 to 500 ms in fixed SSD 2 (see Fig. 1B) and showed violations at the 0-, 50-, and 100-ms SSDs. In both fixed SSD studies, we ran a linear mixed effects model to evaluate whether violations were significantly greater than zero. In fixed SSD 1, the violation did not reach significance (see Fig. 1A), but in fixed SSD 2, we showed that the violations were not only positive but also significantly greater than 0 at the 0-ms SSD (see Fig. 1B).

### Severe violations at short SSDs across studies

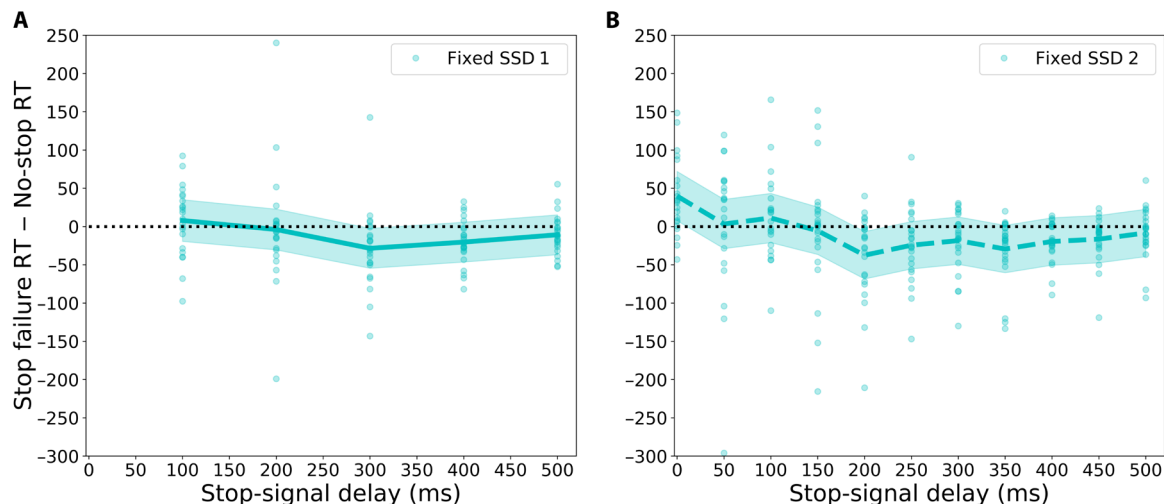
In our fixed SSD experiments, we showed numerical evidence of violations at shorter SSDs, with violations significantly greater than zero at the 0-ms SSD. However, the stop-signal literature is dominated by studies that use a 1 up 1 down tracking procedure (25) to determine SSDs. Therefore, we aimed to assess whether violations are also present when this procedure is used. We plot the violation against SSD across 25 conditions (see Fig. 2A and figs. S2, S4, and S6), including the two fixed SSDs from above for comparison. To summarize the prevalence of violations, we also plot the proportion of datasets (Fig. 2B) and individual subjects within each dataset (Fig. 2C and figs. S3 and S6) that show violations (i.e., numerically positive values from Fig. 2A) at a given SSD. These violations did not result from different subjects contributing to short and long delays (see fig. S1).

To evaluate whether violations were significantly greater than zero at short SSDs, we ran a series of linear mixed effects models. First, we ran linear mixed effects models separately on each condition, which revealed short SSD violations that were significantly greater than zero in at least one short SSD in 6 of the 25 conditions. However, short SSD trials are rare and stop-failure trials are rare at short SSDs, so we may not have the power to robustly evaluate the statistical significance of violations at short SSDs. To increase power, we ran a hierarchical linear mixed effect model that included all 25 conditions and revealed violations that were significantly greater than zero at short SSDs (see Fig. 2A, gray confidence band). This was the case even if stimulus selective stopping conditions were removed (see fig. S9). Therefore, the go process is slowed or impaired at short SSDs, which is inconsistent with not only the independent race model (7, 9) but also extant models of trigger failures (16). We will discuss the implications of this result in the Discussion.

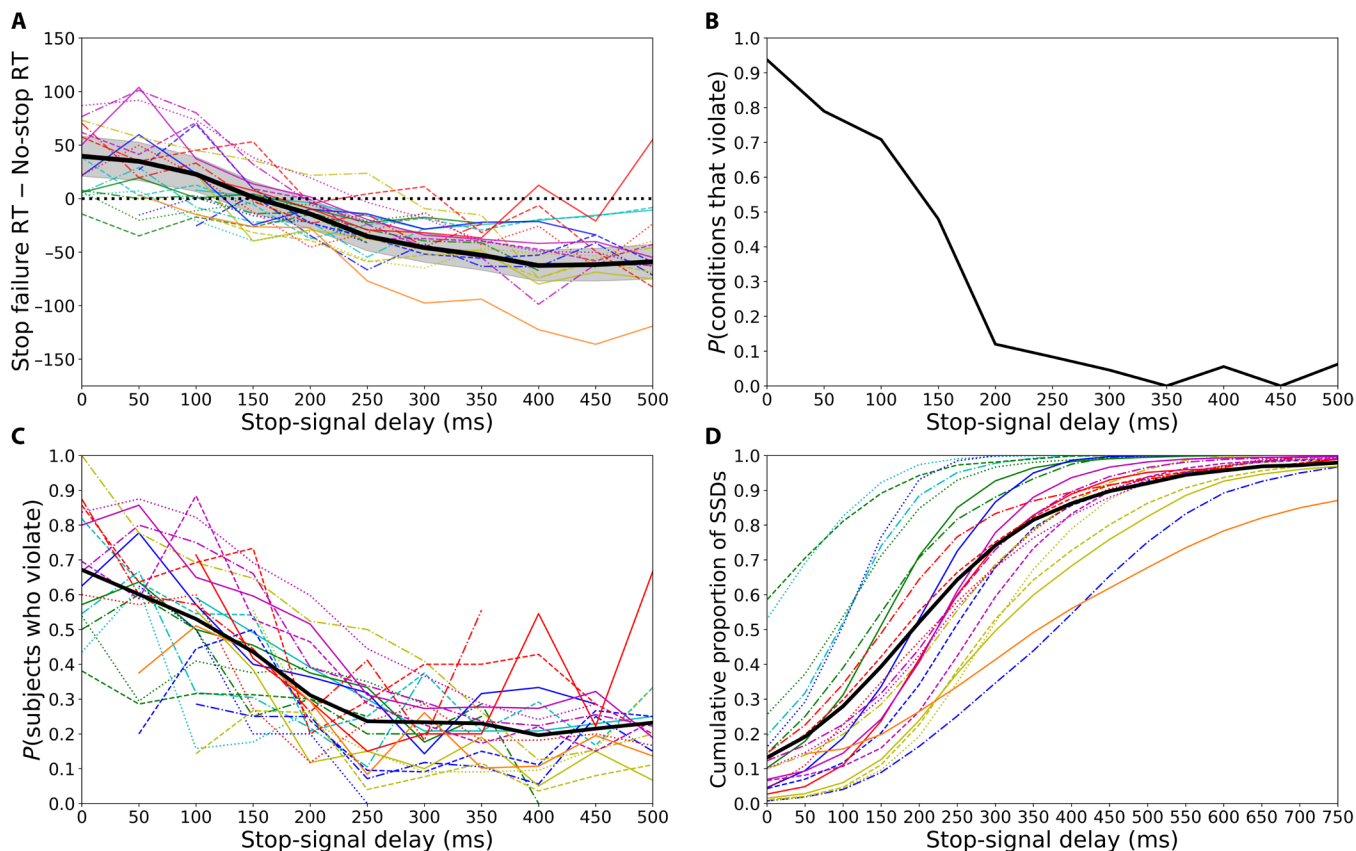
Together, Fig. 2 (A to C) reveals that violations are common at short SSDs and rare at long SSDs. Violations dominate at short SSDs, with mean violations being numerically positive at SSDs of <200 ms (see Fig. 2A). This result is notable because the independent race model predicts that stop-failure RTs should be progressively shorter than no-stop-signal RTs as SSD becomes shorter. Extant models of trigger failures predict that stop-failure RTs are less than or equal to no-stop-signal RTs, so these data are inconsistent with both the independent race model (7, 9) and trigger failure models (16). In addition, although most of the studies included in the present analysis used a commonly used SSD tracking procedure that was not designed explicitly to result in SSDs of <200 ms, these short SSDs made up

**Table 1. Basic information on analyzed datasets and conditions.** *N* is the number of subjects before excluding based on an insufficient number of trials at short SSDs (see the Supplementary Materials for details). Trial *N* is the number of total trials per subject.

Condition no.	Condition description	<i>N</i>	Trial <i>N</i>	Collection method	Stop modality	Stop rule	(Citation no.)
1	Fixed SSD 1	24	1200	In person	Auditory	Simple	(36)
2	Fixed SSD 2	24	2400	In person	Auditory	Simple	(36)
3	Deadline 3 (300 ms)	24	480	In person	Auditory	Simple	(36)
4	Deadline 1 (500 ms)		480	In person	Auditory	Simple	(36)
5	Deadline 1 (700 ms)		480	In person	Auditory	Simple	(36)
6	Deadline 2 (300 ms)	24	480	In person	Auditory	Simple	(36)
7	Deadline 2 (500 ms)		480	In person	Auditory	Simple	(36)
8	Deadline 2 (700 ms)		480	In person	Auditory	Simple	(36)
9	Stop probability 0.2	24	1200	In person	Auditory	Simple	(21), E. 1
10	Stop probability 0.4		1200	In person	Auditory	Simple	(21), E. 1
11	Turk simple 0.2	339	300	Mechanical Turk	Visual	Simple	(40)
12	Turk simple 0.4		300	Mechanical Turk	Visual	Simple	(40)
13	Turk stim selective		300	Mechanical Turk	Visual	Stim selective	(40)
14	Turk motor selective		300	Mechanical Turk	Visual	Motor selective	(40)
15	Saccadic eye movements	11	600	In person	Auditory	Simple	(36)
16	Between-subjects modality auditory 1	24	1200	In person	Auditory	Simple	(42) E. 1
17	Between-subjects modality auditory 2	24	1200	In person	Auditory	Simple	(42) E. 2
18	Between-subjects modality visual 1	32	1200	In person	Visual	Simple	(42) E. 3
19	Between-subjects modality visual 2	24	1200	In person	Visual	Simple	(42) E. 4
20	Between-subjects stimulus selective stop	24	1200	In person	Auditory	Stim selective	(27) E. 1
21	Within-subjects central go simple stop	24	520	In person	Auditory	Simple	(36)
22	Within-subjects central go selective stop		520	In person	Auditory	Stim selective	(36)
23	Within-subjects peripheral go simple stop		520	In person	Auditory	Simple	(36)
24	Within-subjects peripheral go selective stop		520	In person	Auditory	Stim selective	(36)
25	Variable difficulty	53	576	In person	Visual	Simple	(38)



**Fig. 1. Violations at short SSDs in experiments designed to evaluate violations.** Violations across SSDs in the fixed SSD 1 (A) and fixed SSD 2 (B) conditions using linear mixed effects modeling (corrected for multiple comparisons across SSD values). Positive values indicate violations of context independence. Shaded areas indicate the 95% confidence interval. Fixed SSD 1 (A) includes SSDs between 100 and 500 ms at 100-ms increments, and fixed SSD 2 (B) includes SSDs between 0 and 500 ms at 50-ms increments.



**Fig. 2. Severe violations of independence.** The thick black line is the mean across studies, and colored lines are individual conditions or studies. (A) Positive values indicate severe violations. The gray band around the mean is the 95% confidence interval of the results of the linear mixed effect model using data from all conditions and studies (corrected for multiple comparisons across SSD values). Figure S2 is identical but includes a legend. (B) Proportion of conditions that violate [i.e., positive values from (A) at each SSD]. (C) Proportion of individual subjects from each condition that violate the independence assumption at each SSD. Figure S3 is identical but includes a legend. (D) Cumulative proportion of SSDs in our 23 conditions that tracked SSD with a 1 up 1 down tracking algorithm. Note: A small proportion of stop trials had negative SSDs (0.02) or SSDs above 750 ms (0.007) and are not displayed here or included in any analyses of violations. Figure S7 is identical but includes a legend.

approximately half of all stop trials (40%  $\leq 150$  ms and 53%  $\leq 200$  ms; see Fig. 2D). Thus, we anticipate that this result may be relevant to most subjects and virtually all datasets in the stop-signal literature.

### Violations generalize across various common variables

To understand the mechanisms underlying these violations, we investigated whether they are influenced by common stop-signal task variables, which are reflected in the different colored lines in Fig. 2 (A, C, and D). Most variables including fast versus slow subjects, fast versus slow conditions, manual responses versus saccadic eye movements, and auditory versus visual stop signals did not interact with the violation presented in Fig. 2A (see Table 2 and the Supplementary Materials for statistical details). The preceding results provide clarity on one central point: The violations tend to occur across these manipulations. Therefore, these severe violations appear to result from short SSDs, rather than as a result of speed, specific effector use, or stimulus modality.

There are factors that do modulate the violation: Extreme go speed pressure tends to reduce it (see Table 2, row 2), and slower selective stopping tends to increase it (see Table 2, rows 6 to 8). Therefore, the violation appears to increase with the time that a subject is concurrently processing both the go and the stop process, an explanation that we will return to in the Discussion. However, note the equivocal Bayesian results that suggest similar evidence for the null and alternative hypothesis (see Table 2), so additional research is necessary to draw any strong conclusion about the effect of go speed pressure or selective stopping on the size of the violation. Related,

although some of the largest violations are in stimulus selective stopping conditions, as we mention above, the violation across all conditions is still significantly greater than zero if the four stimulus selective stopping conditions are removed (see fig. S9), showing that even in a smaller set of only 21 stopping conditions, the violations cannot be explained by the independent race model (7, 9) or extant trigger failure models (16).

### Removing short SSDs that are prone to violations can change conclusions

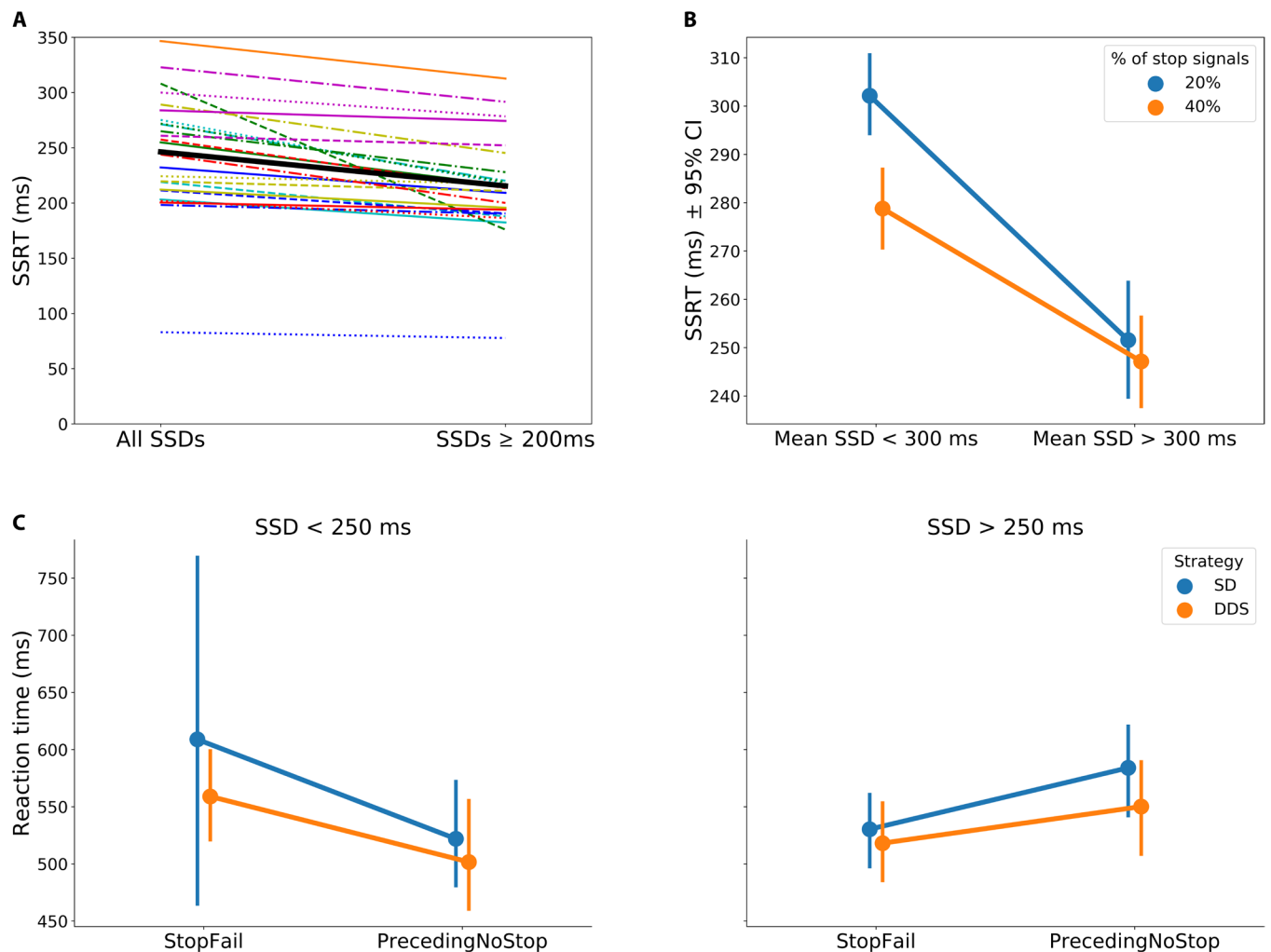
We have shown severe violations, but that does not mean that they are consequential. Perhaps short SSD trials or short SSD subjects could be removed and the same conclusions would be drawn from data. As we showed in Fig. 2D, short SSDs were approximately half of the stop trials in our data, showing that they are common and suggesting that if they are removed, this may be consequential. To test the latter, we compared SSRT estimates in our 25 conditions with short SSDs included or excluded. In 24 of our 25 conditions, SSRTs were significantly faster when short SSDs (<200 ms) were excluded (see Fig. 3A and fig. S8; also see fig. S5 for SSRTs across all SSDs in the fixed SSD 2 dataset). The only exception was the saccadic eye movement condition (see dotted blue line in Fig. 3A with SSRT of around 80 ms).

We also examined two example datasets to evaluate whether their conclusions would change if short SSDs were removed. First, an open question in the inhibition literature is whether reactive response inhibition, as measured by SSRT, is influenced by proactive control, which can be manipulated by increasing the probability of a stop signal

**Table 2. Which variables influence the violation?** Most variables do not affect the violation (although note that Bayes factors ( $BF_{10}$ ) in support of the null hypothesis in rows 1 and 3 to 5 were approximately 3, suggesting anecdotal evidence), but extremely short deadlines (row 2) reduce the violation and introducing stimulus or motor selectivity increases the violation (rows 6 to 8, though note equivocal Bayes factors). Numbers in parentheses in the condition column correspond to the condition column in Table 1. Statistics were based on the analysis of variance (ANOVA) interaction of the trial type (preceding no stop versus stop fail) and the condition on mean RT.  $\eta^2$  is a measure of effect size

	Comparison	Condition	Violation	P	$BF_{01}$	$\eta^2$
1	Fast vs. slow subjects	RT below median (1/2) vs. RT above media (1/2)	6 vs. 15 ms	0.549	3.034	0.0009846
2	Varying go response deadlines	300 ms (3/6) vs. 500 ms (4/7) vs. 700 ms (5/8)	-16 vs. -5 vs. 3 ms	0.011	2.478	0.008
3	Low vs. high stop probability	0.2 (9/11) vs. 0.4 stop probability (10/12)	8 vs. 27 ms	0.127	2.582	0.007
4	Saccadic vs. manual responses	Saccadic (15) vs. manual responses (9)	-9 vs. -3 ms	0.754	3.799	$9.382 \times 10^{-5}$
5	Auditory vs. visual responses	Auditory (16/17) vs. visual responses (18/19)	-22 vs. -13 ms	0.637	3.095	0.0007657
6	Stimulus selective vs. simple stopping	Stimulus selective (20) vs. simple stopping (9)	55 vs. -1 ms	0.064	0.802	0.018
7	Stimulus selective vs. simple stopping	Stimulus selective (13/22/24) vs. simple stopping (9)	49 vs. 24 ms	0.018	0.794	0.014
8	Motor selective vs. simple stopping	Motor selective (14) vs. simple stopping (11/12)	69 vs. 36 ms	0.019	1.467	0.015





**Fig. 3. Fundamental conclusions can change when short SSDs are removed.** Note: Error bars in (B) and (C) represent 95% confidence intervals from bootstrapping. (A) SSRT is significantly faster in 24 of our 25 conditions when short SSDs (< 200 ms) were removed. Figure S8 is identical but includes a legend. (B) SSRT is only faster in the 40% stop signal than the 20% stop signal condition in subjects with predominantly shorter (mean < 300 ms) SSDs. (C) Violations are present at shorter SSDs (< 250 ms; left) but not at longer SSDs (> 250 ms; right) for all subjects in stimulus selective stopping, bringing into question the putative differences in strategies (SD, stop then discriminate; DDS, dependent discriminate then stop) proposed by Bissett and Logan (27).

(21, 26). In a comparison of Table 1, conditions 11 versus 12, when we include all subjects, we found faster SSRT at the higher stop probability than the lower stop probability. However, this effect was only present in subjects with predominantly shorter SSDs, not those with predominantly longer SSDs (Fig. 3B). Therefore, removing subjects with short SSDs eliminated the effect of proactive control on SSRT.

Second, Bissett and Logan (27) suggested that performance in selective stopping tasks can be understood by categorizing subjects into different strategies, some of which are defined by violations of the race model (dependent discriminate then stop, which involves evaluating an ambiguous secondary signal before engaging inhibition if judged to be a stop signal) and others that are defined by context independence (e.g., stop then discriminate, which involves stopping before completing a more time-consuming discrimination process to evaluate an ambiguous secondary signal). Our reanalysis of Bissett and Logan's data (Table 1, condition 20) shows that subjects who were categorized into the strategy defined by violations had shorter SSDs ( $M = 198$  ms) than those categorized into the strategy

defined by independence ( $M = 340$  ms). In addition, there was a crossover interaction in which all subjects, irrespective of strategy, violated the race model when their SSDs were short but not long (Fig. 3C). Therefore, these results suggest that the apparent heterogeneity in strategies can be explained by whether a subject has predominantly short or long SSDs, bringing into question the individual differences in strategies proposed by Bissett and Logan (27). Together, the three preceding analyses provide a proof of concept that scientific conclusions can change when violation-prone short SSDs are removed.

## DISCUSSION

Independence between going and stopping is an essential assumption of the race models that are used to understand virtually every stop-signal dataset (7, 9). We show that violations of the race model are severe and can be consequential. The only necessary and sufficient condition for producing the violation was short SSDs, with

severe violations occurring often at SSDs of <200 ms and seldom at SSDs of  $\geq 200$  ms (Fig. 2, A to C). Therefore, data that include short SSDs, which is likely the case for nearly all published studies using the stop-signal task (Fig. 2D), may have come to erroneous conclusions based on invalid dependent measures (see Fig. 3).

### Toward models that can accommodate violations of independence

All models of the stop task assume context independence, so none of them can account for the violations we observed, as they are currently formulated. Newer models that parameterize the stop and go processes to improve measurement (13) or incorporate choice (9) also assume context independence, as does a recent model that assumes perfect negative dependence between stop and go (28). Recent models characterize the stop and go processes as stochastic accumulators that rise to separate thresholds, and successful inhibition occurs when the stop process finishes first (9), inhibiting the go process (14) or blocking its input (15), reversing its rate of growth before it reaches the response threshold. These models assume context independence for the model parameters that produce the distributions of go and stop RTs (accumulation rate, threshold, and nondesideration time), which implies assuming context independence for response time distributions. In addition, these violations do not appear to result from attentional blinks (29) or psychological refractory periods (30), as those phenomena involve impaired processing to a second (stop) stimulus, whereas the violation reflects impaired processing of the first (go) stimulus.

We have begun to explore modifications to extant models that might account for the violations, simulating the model with hand-picked parameters and assessing their predictions for mean stop-failure RTs and the probability of stop failure across SSDs. We considered two classes of models: ones that assume that short SSDs are special and violations only occur at short SSDs, and ones that assume that violations occur with some probability at all SSDs but are manifest only at short SSDs with our conservative criterion because they provide more time for the violations to take effect.

### Short SSDs are special

We considered several “short SSDs are special” models and were able to find sets of parameters for short and long SSDs that produced violations at short SSDs but not at long ones, as observed. The original independent race model (7), which addresses only the finishing time of the stop and go processes and not the computations that give rise to them, can produce violations if stop and go processes are both delayed at short SSDs. A more recent version of the independent race model (9), which assumes that stop and go processes are stochastic accumulators whose finishing time distributions are governed by rate of accumulation, threshold, and nondesideration time, could produce violations if stop and go nondesideration times were prolonged or stop and go accumulation rates were reduced at short SSDs. Such slowing of both go and stop could be driven by a lapse of attention. In our fixed SSD datasets (see the Supplementary Materials), choice accuracies on stop-failure trials were similar at short SSDs to those at long SSDs and go trials, which is more consistent with prolonged nondesideration time (which should not change choice accuracy) than reduced drift rates (which should reduce choice accuracy) at short SSDs. Perceptual fusion of go and stop stimuli could explain an increase in nondesideration time (31–34), and capacity sharing between go and stop processes could explain a reduction in

drift rate (9, 35). Logan and colleagues (9) suggested that go and stop do not share capacity, but this may be the case only at longer SSDs. The Boucher and colleagues’ interactive race model (14), which models stop and go processes as stochastic accumulators that interact with each other, could account for violations if both the stop accumulation rate was reduced and within-trial variability in the accumulation rate (the diffusion coefficient) was reduced on some proportion of the trials. The Logan and colleagues’ blocked input model (15), which is similar to the interactive race model but inhibits by reducing the go accumulation rate to zero, can also predict violations if the rate is reduced by a small amount and stop within-trial variability is reduced. Thus, there is potential to develop “short SSDs are special” models to account for the violations of context independence observed at short SSDs. These models would justify special treatment of data from short SSDs, either excluding the data or allowing different model parameters to deal with it.

### Variable potency of inhibition across all SSDs

Alternatively, violations may be produced by processes that occur at all SSDs but result in violations that are particularly severe at short SSDs. We hypothesized (36) that introducing variability in the potency of the stop process across all stop trials could produce violations that are only severe enough to be recognized with our conservative criterion at short SSDs. This possibility would present a more fundamental challenge to current theories [e.g., (7, 9)] and consensus stop analysis procedures (8), as it would not justify simply excluding short SSDs and would require fitting new models to measure SSRT and interpret stop task data.

We have explored this possibility in interactive race (14) and blocked input (15) models that implement weakened inhibition on some proportion of stop trials at all SSDs. Our interactive race model, which reduces stop accumulation rate to near zero and reduces stop within-trial variability substantially on some proportion of the trials at all SSDs, can produce the observed violations at short SSDs and apparent nonviolations at long SSDs. Our blocked input model, which varies go accumulation rate across trials and reduces stop within-trial variability, also produces the observed pattern of results. These models predict that violations will be smaller as SSD increases because longer SSDs reduce the time that the weakened inhibition can affect go accumulation. With short SSDs, the weakened inhibition takes effect shortly after go accumulation begins and affects it until go accumulation hits the threshold. With longer SSDs, the weakened inhibition takes effect well after go accumulation begins, when go activation is closer to the threshold, so there is less time for weakened inhibition to affect go RT. These models are consistent with our finding that the violation may be smaller with extreme go speed pressure, as there is less opportunity for a protracted interaction with a very fast go process, perhaps because of reduced go encoding time under time pressure (37). These models are also consistent with our finding that the violation may be larger in selective stopping, as complicating the stop process may encourage a weaker stop process that prolongs but does not fully inhibit the go process. These models are important because they argue against excluding short SSDs or subjects with short SSDs to salvage existing models. They urge new model development.

### Implications for trigger failures

Our results challenge extant trigger failures models (16) that assume that a trigger failure entails the go process racing alone, because

they predict stop-failure RTs that are faster or as slow as no-stop-signal RTs but not slower, and we show stop-failure RTs that are significantly slower than no-stop-signal RTs (see Fig. 2A). However, violations and trigger failures may be related. In particular, trigger failures may be graded rather than absolute such that partially triggering the stop process may yield weak inhibition and violations (see the “Implications for trigger failures” section in the Supplementary Materials). In addition, as mentioned by a reviewer, slower stop failures could arise if a trigger failure entails a lapse in processing of both the go and the stop process.

### Limitations

First, all of our modeling suggestions should be taken as preliminary, and additional modeling studies (including parameter search, model fitting, model comparison, parameter recovery, and model recovery) will be necessary to create a full model capable of validly estimating SSRT in the presence of such variable potency of inhibition. Second, some of our studies include a small number of short SSDs (see Fig. 2D), limiting the statistical power of some of our analyses, especially those in Table 2. This motivated our inclusion of the two fixed SSD datasets and the inclusion of a large number of datasets. Third, our analysis focuses primarily on data that we have acquired, which leaves open the possibility that idiosyncrasies in our experimental design or procedures could drive the observed violations. However, the presented experiments have been administered in various settings (Vanderbilt University, Stanford University, and online), using different code bases, and by different experimenters, reducing experimental similarities across our conditions. In addition, we included a dataset from Matzke and colleagues (38) in our main analyses. Last, previous work from Colonius and colleagues (17–20) has shown evidence for similar short SSD violations. For these reasons, we believe that the severe violations will generalize to other stopping datasets.

### Practical considerations

The context independence assumption is a keystone assumption for all modern models of response inhibition, and we have shown severe violations of this assumption. Therefore, to validly estimate SSRT, we believe that it is necessary to develop a new computational model for stopping that accommodates context dependence. We are currently developing such models, and we hope that our preliminary models presented above will also spur modeling work by others. Until such models are validated, users of the stop-signal task may consider the following strategies to limit the degree to which violations contaminate their results. Across all suggestions, we recommend evaluating violations at all SSDs of all subjects using the methods described above and the linked open-source code.

One class of mitigating strategies involves adjusting experimental parameters to try to avoid the conditions and parameters that we found to result in the most severe violations. First, the violations appear to be most severe at SSDs of <200 ms, so study designers could restrict the range of SSDs to values of  $\geq 200$  ms. However, this may encourage RT slowing, as subjects may recognize that they cannot stop faster responses, and progressive RT slowing has been shown to contaminate SSRT estimates (39). If implementing this strategy, it will be important to emphasize in instructions and feedback that subjects should not slow their responses to wait for stop signals. This first strategy may also result in missing the intermediate portion of the inhibition function for faster subjects, which has been shown to be the most informative for constraining SSRT estimates

(12). Second, researchers could focus on simple stopping (stop all responses to one unambiguous stimulus) instead of the apparently more violation-prone selective stopping conditions. Selective stopping paradigms address selectivity in controlled behavior, as subjects stop only certain responses or responses to certain stimuli. This allows additional comparisons including between stop trials and high-level control conditions that involve similar attentional demands but putatively do not involve inhibition per se [e.g., ignore trials (27)]. Therefore, the empirical questions under investigation may require introducing selectivity, so abandoning selective stopping may not be possible. If selective stopping is used, particular care should be taken to test for violations and consider the implications of their presence. Third, study designers could implement short go response deadlines, which we showed somewhat reduced violations. However, a short go response deadline will be difficult to meet when the go task is more challenging. In addition, at very short deadlines, we found that some subjects required negative SSDs to successfully stop (see the two most positive lines in Fig. 2D at 0-ms SSD that correspond to the fastest 300-ms go response deadline conditions), and negative SSDs may change the subjective nature of the task (i.e., being told to stop something that has not started). Future work could evaluate whether the reduced violations at short SSDs was driven by the presence of negative SSDs. Fourth, short SSDs could be avoided by choosing a go task that yields longer RTs, perhaps by including a more difficult stimulus to response mapping. When coupled with the usual 1 up 1 down tracking algorithm (25), this may avoid short SSDs entirely (i.e., the lower end of the inhibition function will be  $\geq 200$  ms). However, we have not evaluated whether violations interact with experimental manipulations that would prolong go RT, like a more difficult stimulus to response mapping.

The second class of mitigating strategies involves analytical strategies to avoid short SSDs. First, short SSD trials could be excised and SSRT could be computed only from stop trials with SSDs above a threshold, perhaps  $\geq 200$  ms. Second, SSRT could be computed only for subjects with few or no short SSDs. However, these strategies share substantial shortcomings. Both would likely involve substantial data loss. For example, in our sample of 25 conditions, eliminating stop trials with SSDs of  $\leq 150$  ms would involve removing 40% of all stop trials. In addition, if using a tracking algorithm for SSD like the common 1 up 1 down algorithm (25), short SSDs arise from fast go RTs and slow SSRTs. When the go process is fast and the stop process is slow, the go process tends to win and SSD is reduced, resulting in short SSDs. Therefore, if subjects with short SSDs were removed, the sample would be biased toward subjects with slower go processes and faster stop processes. If short SSD trials were removed, subjects with more short SSD trials would have less data from which to base their SSRT estimates, making them less robust. In addition, this requires defining a threshold for what qualifies as “short.” In our analyses, we have primarily focused on SSDs of <200 ms because these were the SSDs that produced violations that were numerically greater than 0 ms in Fig. 2A. However, the requirement of numerically positive violations is a conservative criterion and may miss less severe violations.

We do not believe that any of these mitigating strategies are unqualified solutions, and many of the strategies (e.g., removing short SSD trials or subjects with short SSDs) implicitly suggest that short SSDs are special. As we show in our preliminary simulations above, the severe violations that are apparent at short SSDs may arise from processes that occur on all stop trials, so short SSDs may not be as



special as they appear. If this is the case, new modeling may be the only solution to extract trustworthy estimates of SSRT, as all stop trials involve dependence between go and stop. We believe that new modeling is an essential step to placing the stop-signal literature on sound theoretical ground that accommodates severe violations of context independence.

## MATERIALS AND METHODS

### Experimental design

Below, we describe the design of each of the 25 conditions included in our study. There were no prespecified components.

#### Condition 1

**Subjects.** Twenty-four adults recruited from the Nashville area were given \$12 for a single 1-hour session. All subjects had normal or corrected-to-normal vision. One subject was replaced for mean go RT more than 3 SDs above the group mean RT. In this and all subsequent data acquisition, we have complied with all relevant ethical regulations including study approval from the Vanderbilt or Stanford University IRB (institutional review board) and informed consent from each participant.

**Apparatus and stimuli.** The experiment was run on a Pentium Dual-Core PC running E-Prime 1 (pstnet.com). The stimuli were presented on a 19-inch cathode ray tube monitor. The go task was to respond to a single black shape (triangle, circle, square, or diamond) on a white background presented in the center of the screen. The height and width of each shape was 4 cm at the longest point. Subjects responded on a QWERTY keyboard. The stop signal was a 500-Hz tone (70 dB, 100 ms) presented through closed headphones.

**Procedure.** Each trial began with a 500-ms fixation cross, followed by the presentation of the go stimulus for 850 ms, and followed by a 1000-ms blank-screen intertrial interval (ITI). The go task was to respond as quickly and accurately as possible based on the identity of the centrally presented shape. Two of the shapes were mapped on the “z” key, and the other two were mapped onto the “m” key, and subjects responded with their left and right index fingers, respectively. The shape to key response mapping was counterbalanced across subjects.

A stop signal occurred on a random 20% of all trials, and subjects were instructed to try their best to stop their response when they heard it. There were five SSDs: 100, 200, 300, 400, and 500 ms. The SSD was randomly selected on each stop trial, with the only constraint being each was presented exactly 48 times for each subject.

Subjects were instructed to respond quickly and accurately to the shapes and then were given 12 trials of experimenter-supervised practice on trials without stop signals. Then, stop-signal trials were introduced, and subjects were instructed to also do their best to stop on stop-signal trials. They were given another 12 trials of practice that included two stop signals. After practice, subjects completed the main task of five blocks of 240 trials each. Between blocks, subjects were given feedback on the speed and accuracy of their no-stop-signal trials from the previous block.

#### Condition 2

**Subjects.** Twenty-four adults recruited from the Nashville area were given \$24 for a single 2-hour session. All subjects had normal or corrected-to-normal vision.

**Apparatus and stimuli.** The apparatus and stimuli for condition 2 matched those for condition 1.

**Procedure.** The procedure for condition 2 was the same as condition 1 with the following exceptions. The probability of a stop signal was 0.22 instead of 0.2. There were 11 SSDs with 48 stop trials each: 0, 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500 ms. There were 10 blocks of 240 trials each. At the end of the fifth block, subjects took a 5-min break before beginning the second half of the experiment.

#### Conditions 3 to 8

**Subjects.** Forty-eight subjects were recruited from the Nashville community and were compensated \$12 for a single 1-hour session. All subjects had normal or corrected-to-normal vision. Ten subjects were replaced, whose probabilities of successful stopping fell outside the 95% confidence interval of 0.5.

**Apparatus and stimuli.** The apparatus and stimuli for conditions 3 to 8 matched those for condition 1.

**Procedure.** The procedure was the same as condition 1 with the following exceptions. The deadline for the go response was manipulated by varying go-stimulus duration (300 ms for conditions 3 and 6, 500 ms for conditions 4 and 7, or 700 ms for conditions 5 and 8) and instructing subjects to respond before the go stimulus disappeared. There were six blocks of 240 trials each, and go-stimulus duration varied across the first three blocks in an order that was counterbalanced across subjects. The order of blocks for a given subject was the same for the first three blocks and the last three blocks. Each trial began with a 500-ms fixation display, followed by the go stimulus. In 24 of the subjects, a 1000-ms ITI followed the go stimulus (conditions 3 to 5), and in the other 24 subjects, the ITI was 1200, 1000, and 800 ms for go durations of 300, 500, and 700 ms (conditions 6 to 8), respectively. The performance in these two groups was very similar, so they were collapsed into one sample for the analyses presented in row 2 of Table 2.

On a random 25% of trials, a stop signal occurred that indicated that subjects should withhold their response for that trial. The SSD was varied with a tracking algorithm to achieve  $P(\text{respond}|\text{stop signal}) = 0.5$  (25). When subjects successfully inhibited, SSD increased by 50 ms; when subjects failed to inhibit, SSD decreased by 50 ms. There were three separate SSD tracking algorithms, one for each deadline.

Subjects were told to try to respond before the go stimulus left the screen and to sacrifice go response accuracy to respond before the deadline (although responses were still recorded after the deadline). After the instructions, subjects were given 24 trials of experimenter-supervised practice with the 500-ms deadline. After practice, subjects completed the main task. At the end of each block, subjects were given feedback on mean RT and mean accuracy from that block, as well as the percentage of trials in which they met the deadline.

#### Conditions 9 and 10

The methods are described in detail in experiment 1 of (21). The methods are similar to condition 1 except that SSD was tracked with a 1 up 1 down tracking algorithm (as in conditions 3 to 8). The go stimuli were four visual shapes (triangle, circle, square, or diamond) mapped onto two keypress responses. The stop signal was a 500-Hz auditory tone.

#### Conditions 11 to 14

**Subjects.** Six hundred sixty-two subjects were recruited from Amazon Mechanical Turk (MTurk) and were compensated \$6/hour to complete a 10-hour battery of 63 tasks at their own pace within the constraint that they completed within 1 week. Subjects had to have 95% of past MTurk assignments approved, have completed at least 2000 assignments, live in the United States, and be an adult. Five hundred

twenty-two subjects completed all tasks and passed basic quality assurance applied to all 63 tasks [in general, this included median response times longer than 200 ms, no more than 25% omission rate, accuracy higher than 60%, and no single response given on most trials; see (40) for details]. We also applied additional criteria specifically to the stop-signal tasks. One hundred twenty subjects were removed, whose probability of successful stopping fell outside the 95% confidence interval of 0.5 for at least one of conditions 11 to 14. We also applied the binomial test to ensure that accuracy was significantly higher than 0.5 in all conditions, which was not the case for 62 subjects in the noncritical go trials in the motor selective stopping task and another 1 subject on ignore trials in the stimulus selective stopping task, resulting in another 63 subjects being removed. This resulted in a final sample of 339 subjects.

**Apparatus and stimuli.** The experiments were run on the subject's PC operating on either Windows or OSX. The experiments were coded using jsPsych ([www.jspsych.org/](http://www.jspsych.org/)) and were published to MTurk via expfactory [[expfactory.org](http://expfactory.org) (41)]. Shapes differed between tasks and are listed below:

- 1) Simple stop signal: Pentagon, hourglass, teardrop, square
- 2) Motor selective stop signal: Circle, rhombus, L shape, triangle
- 3) Stimulus selective stop signal: Rectangle, oval, trapezoid, moon

Each shape was 275 pixels at the longest point, and all were black. Subjects responded on a QWERTY keyboard. Like the preceding conditions, conditions 11 to 14 involved a 4 to 2 stimulus to response mapping, with the responses being z and m on the keyboard. The stop signal was a 14-sided star, which was black in the simple stop and motor selective stop tasks and either blue or orange in the stimulus selective stop task. All stimuli were created with default shapes in PowerPoint.

The visual angle between the go and stop stimulus varied from  $<1^\circ$  to  $\sim 2^\circ$ . Based on feedback on a previous version of this work, we evaluated the possibility that violations in this dataset may be driven by smaller visual angles. To do so, we computed the violation separately for each go stimulus and found that violations were not significantly or even numerically larger when the visual angle between the go and stop stimulus was smaller, providing evidence against this possibility.

**Procedure.** Subjects completed a total of 63 tasks and surveys, of which three were stop-signal tasks. The order of tasks and surveys was randomized across subjects. Subjects were encouraged to spread work for the battery over the week and not do too many tasks in a row. The timing for each trial matched condition 1. Each stop-signal task will be discussed in more detail below.

Simple stop (conditions 11 and 12):

Subjects completed two types of practice: The first (20 trials) focused on speed and accuracy of shape-to-key mapping, while the second (12 trials) included stop signals. Subjects repeated practice blocks of each type until they completed five practice blocks or they met task-specific quality assurance (QA) thresholds. Between each practice block, subjects were given feedback on relevant QA thresholds. These thresholds are the following:

- 1) Average RT of less than 1000 ms
- 2) Go accuracy greater than 80%
- 3) Omit no more than 10% of all go trials
- 4) Stop accuracy is between 20 and 80%

There were 12 blocks and 50 trials in each block. Six blocks had a stop probability of 0.2 (condition 11), and the other six had a stop probability of 0.4 (condition 12). The order of conditions was counterbalanced across subjects, and subjects completed all six blocks for

one condition before they completed all six blocks for the other condition. As in conditions 9 and 10, subjects were not instructed about the manipulation of stop probability. Between blocks, subjects were given feedback on all relevant thresholds.

Stimulus selective stop (condition 13):

Subjects were instructed to try their best to stop their response if they saw a blue star but not if they saw an orange star. Therefore, stopping was selective to the stimulus color dimension. SSD was updated on trials that required subjects to stop their response, and ignore signal delay was yoked to SSD. There were six blocks and 50 trials in each block. In each block, stop signals occurred on a random 20% of trials, while ignore trials occurred on another random 20% of trials.

Trial timing and instructions were the same as conditions 11 and 12 except the first no-stop practice included 12 trials and the second practice that included stop and ignore trials included 30 trials. In addition, an additional QA threshold was applied such that subjects had to respond on more than 60% of ignore trials. Between blocks, subjects were given feedback on all relevant thresholds.

Motor selective stop (condition 14):

Subjects were instructed to try their best to stop their response if they saw a black stop signal and they were going to respond with the critical stop response (e.g., z); otherwise, they were instructed to ignore the stop signal and keep responding (e.g., if they were going to respond m). Therefore, inhibition was selective to a certain motor response (the critical response) but not the other (the noncritical response). SSD was updated on critical response stop trials. There were five blocks and 60 trials in each block. In each block, stop signals occurred on a random 40% of trials, half of which occurred on critical response trials and the other half on noncritical response trials. As in conditions 11 to 13, between blocks, subjects were given feedback.

### Condition 15

**Subjects.** Eleven young adults recruited from the Nashville area were given \$60 for five 1-hour sessions on five consecutive days. All subjects had normal or corrected-to-normal vision. Four subjects were replaced because their eyes could not be tracked satisfactorily, and two subjects were replaced because they did not complete all five sessions.

**Apparatus and stimuli.** The experiment was run on a PC running SR Research Experiment Builder software connected to a PC running EyeLink 2000. The stimuli were presented on a 19-inch cathode ray tube monitor displaying a  $1024 \times 768$  pixel resolution. The go task was to saccade to a black X presented on the right or left side of the screen. The X was 50 pixels by 50 pixels, and its center was positioned at coordinate  $172 \times 384$  if presented on the left and  $852 \times 384$  if presented on the right. The stop signal was either a 500-Hz tone, 750-Hz tone, or 1000-Hz tone (70 dB, 100 ms) for a given subject, and the tone choice was counterbalanced across subjects. The tone was presented through closed headphones.

Saccades were registered by EyeLink if above a velocity threshold of  $30^\circ/s$  (and remained above the threshold for 4 ms) or an acceleration threshold of  $8000^\circ/s$  per second. The minimum motion threshold was  $0.1^\circ$ . Saccades were registered as correct on a go trial if they landed within a circle around the target X with a radius of 170 pixels. Stop trials were registered as correct if no saccades were registered.

**Procedure.** Subjects completed five sessions across five consecutive days. The first session was a training session in which subjects completed only no-stop-signal trials, and they received trial-by-trial feedback as to whether their response was recorded as correct by the

eye tracker. This session was intended to train subjects to appropriately fixate and saccade. The final four sessions involved simple stopping to auditory stop stimuli, simple stopping to visual stop stimuli, stimulus selective stopping to auditory stop and ignore stimuli, and stimulus selective stopping to visual stop and ignore stimuli. The order of the final four sessions was counterbalanced across subjects. We focus on the results from the simple stopping to auditory stop stimuli session. Simple stopping to auditory stimuli is common, and the questions of whether modality and stimulus selectivity influence the violation are addressed in conditions 16 to 19 on larger datasets that involve manual responses. Subjects pressed the spacebar to begin each trial, which initiated drift correction and began a 500-ms fixation period before the target appeared for 1000 ms, followed by the 850-ms blank-screen ITI.

SSD was tracked with a 1 up 1 down tracking algorithm (25). Auditory stop signals were presented on 20% of all trials. Subjects were instructed to look promptly at the X when it appeared but try to remain fixated on the center of the screen if they heard a tone. After instructions, subjects were given 20 trials of practice. The main task included 10 blocks of 60 trials per session. At the end of each block, subjects were given rest but no feedback.

#### **Conditions 16 to 19**

The methods are described in detail in experiments 1 to 4 of (42). For the analyses in row 5 of Table 2, experiments 1 and 2 are combined to produce the auditory stop signals' dataset and experiments 3 and 4 are combined to produce the visual stop signals' dataset. To summarize, the methods were similar to fixed SSD 1 except SSDs were tracked with a 1 up 1 down tracking algorithm. The auditory stop signals were tones (experiment 1 and 2), and the visual stop signals were colored stars (experiment 3) or black bars presented above or below the go stimulus (experiment 4). Go responses were keypresses.

#### **Condition 20**

The methods are described in detail in experiment 1 of (27). To summarize, subjects responded to go stimuli that were black shapes on a white background, and the stop and ignore stimuli were auditory tones of different frequency. Go responses were keypresses.

#### **Conditions 21 to 24**

**Subjects.** Twenty-four young adults recruited from the Nashville area were given \$36 for two 90-min sessions on consecutive days. Two subjects were replaced, one for not showing for the second session and the other for having a probability of stopping outside the 95% confidence interval of the expected probability of stopping.

**Apparatus and stimuli.** The apparatus was the same as all previous keypress experiments, although the stimuli differed in the following ways. The go task in both sessions began with three "+" signs, one in the center of the screen flanked horizontally by one 2 inches to the left and one 2 inches to the right. The go task differed across sessions, and the order of sessions was counterbalanced across subjects. In one session, the central + changed to a "<" or ">," which informed subjects to respond z or m on the keyboard, respectively. In the other session, either the left or the right + changed to an X, which informed subjects to respond z or m, respectively. All stimuli were presented in 24-point font. Both 500-Hz tones and 750-Hz tones were presented through closed headphones. There were three conditions in each session: simple stopping with 20% stop signals, simple stopping with 40% stop signals, and selective stopping with 20% stop signals and 20% ignore signals.

**Procedure.** The procedure was the same as fixed SSD 1 with the following exceptions. SSD was tracked with a "1 up 1 down" track-

ing algorithm (25). Here, we only compared the 20% simple stopping condition to the selective stopping condition. The order of conditions was counterbalanced across subjects, but the order was the same for both the central and peripheral session for each subject.

In simple stopping blocks, subjects were instructed to stop if either the high (750 Hz) or the low (500 Hz) tone was presented. In the selective stopping block, subjects were instructed to stop to one of the two tones and ignore the other (which tone was the stop signal was counterbalanced across subjects). Subjects stopped to the same tone in both sessions.

Subjects were given 10 trials of experimenter-supervised practice on trials without stop signals. They were given another eight trials of practice that included stop signals for their first condition of the day and six trials of practice before starting each of the subsequent two sessions of each day. After the initial practice, subjects completed two blocks of 260 trials each for the first condition, then practiced the second condition and completed two blocks of 260 trials each for the second condition, and then practiced the third condition and completed two blocks of 260 trials each for the third condition. This procedure was repeated in the second session. Between blocks, subjects were given feedback on the speed and accuracy of their no-stop-signal trials from the previous block.

#### **Condition 25**

The methods are described in detail in (38). Briefly, subjects responded with the "Z" or "/" keys with their left or right index finger to indicate whether a random-dot kinematogram displayed 45° left or right upward global motion, respectively. There were two interleaved difficulty levels, with greater coherence in the easier condition. The stop signal was a gray square around the go stimulus. Stop signals occurred on 29% of all trials. For most stop-signal trials (86%), SSD was determined by a 1 up 1 down tracking algorithm (25) with a step size of 33 ms. On a small subset of stop trials (14% of all trials), SSD was presented at a fixed 50 ms. Given that most SSDs were determined by tracking, we present these results with the other tracking studies in the main text.

### **Statistical analysis**

#### **Computing the violation**

In each condition, we computed observed mean stop-failure RT and compared it to observed mean no-stop-signal RT on the trial immediately preceding each stop failure. This comparison was only made if the trial immediately preceding the stop-failure trial was a no-stop-signal trial and if this preceding trial was not an omission. Using the no-stop-signal RTs that immediately precede stop failures allows us to more accurately test for violations in experiments that use the typical staircased tracking algorithm for determining SSD. This is because go RTs fluctuate throughout the experiment, and SSD tends to fluctuate with it, so when RTs are fast, SSD tends to be short, and when RTs are slow, SSD tends to be long (21, 22). Therefore, to test for prolonged stop-failure RTs, which are evidence of violations of the race model, we assume that the no-stop-signal RT on the immediately preceding trial is the best baseline to compare the current stop-failure trial against. This procedure is not necessary for fixed SSD conditions, but most of our conditions and most of the experiments in the human stopping literature include a staircased tracking algorithm, so for consistency, we apply this procedure to all of our analyses, including our fixed SSD conditions.

We computed these stop failure and preceding no-stop-signal RTs for each subject at each SSD for which they have at least two of such

pairs of trials. In general, a cutoff point of one resulted in very noisy individual subject data, and cutoff points larger than two eliminated too many subjects. This resulted in only a subset of subjects contributing to the group average at a given SSD, because only a subset of subjects have two stop-failure trials at any given SSD. For motor selective stopping, only no-stop stop-failure trial pairs in which both used the stop (critical) response were included, given that responses were considerably faster for the response that is never stopped (mean correct noncritical go RT = 529 ms) than the response that can be stopped (mean correct critical go RT = 611 ms),  $t(338) = 24.4$ ,  $P < 0.001$ .

To test whether the results in Fig. 2 were driven by different sets of subjects within a given dataset contributing at short SSDs (<200 ms) and longer SSDs ( $\geq 200$  ms), we recalculated the same violation measure but only included subjects who contributed both short and long SSDs. To do this, we took the data from Fig. 2 and found the range of continuous SSDs for which at least five subjects from that condition contributed to each SSD and which contained at least one SSD < 200 ms and one SSD  $\geq 200$  ms. If these criteria were satisfied by multiple ranges (e.g., 50 to 200 ms and 150 to 350 ms), we chose the range with the shortest SSD for its lower bound (50 to 200 ms in the above example). This result is displayed in fig. S1, which shows that the violations at short SSD are not driven by different sets of subjects within a condition contributing to short and long SSDs. In addition, the fixed SSD studies (see Fig. 1) argue that the main results presented in Fig. 2 cannot be driven by different sets of subjects contributing to short and long delays, as all subjects experienced short and long delays.

Note that in analyses that combined data from multiple conditions across SSDs (i.e., Fig. 2 and figs. S1 to S4, S7, and S9), values for the fixed SSD 1 and variable difficulty conditions were interpolated to SSDs that were a multiple of 50 ms using a first-order spline to contribute to group averages at that SSD. When using linear mixed effects models to estimate violations at each SSD (i.e., the black lines in Fig. 2A and figs. S1, S2, S4, and S9), no interpolation was used. However, for visualization purposes, values for SSDs that were unique to the variable difficulty dataset (e.g., 133 ms, 167 ms, and all SSDs > 800 ms) were removed so that every SSD presented contained data from multiple conditions.

In addition, a reviewer suggested that our method for computing the violation may be contaminated by progressive RT slowing on repeated go trials. This suggestion would predict slower RT when go trials repeat. Across our 25 conditions, go trials that followed a go trial ( $M = 495$  ms) were faster than go trials that followed a stop trial ( $M = 515$  ms), consistent with the general finding of post-stop-signal slowing (21). In addition, we found that the RT on the go trial immediately preceding a stop trial ( $M = 501$  ms) and go trials immediately preceding a stop failure trial ( $M = 492$  ms) were not slower than the overall mean RT ( $M = 502$  ms), demonstrating that go trials that preceded stop trials were not unusually slow. Therefore, contamination by progressive RT slowing on repeated go trials cannot explain our severe violations.

We completed our Table 2 and Fig. 3 (B and C) analyses with JASP 0.12.2.0 (jasp-stats.org), an open-source project for flexible, intuitive frequentist and Bayesian analyses. In our Bayes factor computations, we used the default prior values in JASP: r scale fixed effects of 0.5, r scale random effects of 1, and r scale covariates of 0.354. We have shared the full input to and output from JASP for our analyses, as well as all raw data, at <http://doi.org/10.5281/zenodo.4432816>.

### **Analysis pertaining to violations generalizes across various common variables**

In our Table 2 analyses, we focused exclusively on SSDs of <200 ms, as this is the range under which we found evidence of severe violations in Fig. 2. When we compared the violation across conditions, we only included an SSD if it was <200 ms and if all conditions being compared had at least five subjects with at least two pairs of no-stop and then stop-failure trials at that SSD. Therefore, when two conditions were compared, they were compared across the same range of SSDs (to help ensure that any differences between conditions were not driven by differences in SSDs over which the violation was evaluated), but different comparisons in this paper were evaluated over different SSD ranges. Then, each subject in each condition was either included or excluded in the group average for that condition based on whether they had a sufficient number of no-stop and then stop-failure trial pairs in the chosen range for that study. As mentioned above, the criterion was the subject needed to have at least one SSD (within the chosen range for that study) with at least two no-stop followed by stop-failure trial pairs. If they had more than one SSD within the chosen range for that study, then the means of the violation (observed stop-failure RT minus observed preceding no-stop-signal RT) were averaged across the SSDs that passed the criterion of at least two stop-failure trials.

### **Linear mixed effects modeling**

To test whether the violations presented above were significantly greater than 0 at short SSDs, a linear mixed effects model was run using the `lmer()` function from the R package `lme4` (43). First, a violation analysis was run on each condition following the procedure described above, producing a violation for each subject at each SSD. One subject from conditions 21 to 24 was removed because they did not have more than one pair of no-stop + stop-failure trials at any SSD that was shared with at least four other subjects from their dataset. This produced violation data for 674 subjects. Following this, subjects with only one data point (i.e., a single SSD with more than one stop failure preceded by a non-omission go trial) were excluded. This resulted in 2 participants being excluded, 1 from the saccades condition and 1 from the variable difficulty condition, leaving 672 participants across 25 conditions. With this final dataset, a linear mixed effects model was run to estimate the mean violation at each SSD between 0 and 500, along with the bounds of a 95% confidence interval to be used for significance testing, which were adjusted for multiple comparisons using multivariate  $t$  distributions. Subjects and conditions were included as random intercepts in the model. The outputs are presented in Fig. 2A and revealed violations that were significantly greater than zero at SSDs of 0, 50, and 100 ms. Note that while all conditions were used in the fitting of the model and estimation of effects, only SSDs present in two or more conditions are included in the visualization to improve the readability of the figures. To determine whether these results were driven by stimulus selective stopping conditions, which have been known to produce violations in some subjects (27), this process was repeated excluding the stimulus selective stopping conditions (4 of 25), and the results are presented in fig. S9. We found that the lower bound of the 95% confidence interval was above 0 at SSDs of 0 and 50 ms, consistent with severe violations at short SSDs.

To test whether the results in Fig. 2A were driven by different sets of subjects within a given dataset contributing at short SSDs (<200 ms) and longer SSDs ( $\geq 200$  ms), we sparsified the data to subsets of SSDs that contained at least one SSD < 200 ms and one SSD  $\geq 200$  ms and



a set of five or more subjects present at each of these SSD for each condition, as described in the “Computing the violation” section. The sparsified data were run through a linear mixed effects model with subjects and conditions as random intercepts, again using SSDs between 0 and 500 ms and the same adjustment for multiple comparisons. The results are displayed in fig. S1, which shows that the violations at short SSD are not driven by different sets of subjects within a condition contributing to short and long SSDs, with violations remaining significantly above 0 at SSDs of 0 and 50 ms.

An additional linear mixed effects model was run on each condition using the same method as above (i.e., calculating a violation for each subject at each SSD between 0 and 500 and removing subjects with only one data point). Subjects were included as random intercepts in the model, and the bounds of the confidence intervals were adjusted using the same method. Investigation of the lower bounds of the confidence intervals revealed violations at one or more short SSD (i.e., SSDs < 200 ms) in 6 of the 25 conditions. The outputs of these individual models are presented in Fig. 1 (A and B) and fig. S6.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/12/eabf4355/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

## REFERENCES AND NOTES

- P. G. Bissett, G. D. Logan, Stop before you leap: Changing eye and hand movements requires stopping. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 941–946 (2013).
- C. R. Camalier, A. Gotler, A. Murthy, K. G. Thompson, G. D. Logan, T. J. Palmeri, J. D. Schall, Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Res.* **47**, 2187–2211 (2007).
- F. Verbruggen, D. W. Schneider, G. D. Logan, How to stop and change a response: The role of goal activation in multitasking. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 1212–1228 (2008).
- G. D. Logan, Executive control of thought and action. *Acta Psychol.* **60**, 193–210 (1985).
- G. D. Logan, Taking control of cognition: An instance perspective on acts of control. *Am. Psychol.* **72**, 875–884 (2017).
- A. Miyake, N. P. Friedman, M. J. Emerson, A. H. Witzki, A. Howerter, T. D. Wager, The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cogn. Psychol.* **41**, 49–100 (2000).
- G. D. Logan, W. B. Cowan, On the ability to inhibit thought and action: A theory of an act of control. *Psychol. Rev.* **91**, 295–327 (1984).
- F. Verbruggen, A. R. Aron, G. P. H. Band, C. Beste, P. G. Bissett, A. T. Brockett, J. W. Brown, S. R. Chamberlain, C. D. Chambers, H. Colonius, L. S. Colzato, B. D. Corneil, J. P. Coxon, A. Dupuis, D. M. Eagle, H. Garavan, I. Greenhouse, A. Heathcote, R. J. Huster, S. Jahfari, J. L. Kenemans, I. Leunissen, C.-S. R. Li, G. D. Logan, D. Matzke, S. Morein-Zamir, A. Murthy, M. Paré, R. A. Poldrack, K. R. Ridderinkhof, T. W. Robbins, M. Roesch, K. Rubia, R. J. Schachar, J. D. Schall, A.-K. Stock, N. C. Swann, K. N. Thakkar, M. W. van der Molen, L. Vermeulen, M. Vink, J. R. Wessel, R. Whelan, B. B. Zandbelt, C. N. Boehler, A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *eLife* **8**, e46323 (2019).
- G. D. Logan, T. Van Zandt, F. Verbruggen, E.-J. Wagenmakers, On the ability to inhibit thought and action: General and special theories of an act of control. *Psychol. Rev.* **121**, 66–95 (2014).
- G. D. Logan, in *Attention and Performance IX*, J. Long, A. D. Baddeley, Eds. (Erlbaum, 1981).
- G. D. Logan, W. B. Cowan, K. A. Davis, On the ability to inhibit simple and choice reaction time responses: A model and a method. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 276–291 (1984).
- G. P. H. Band, M. W. van der Molen, G. D. Logan, Horse-race model simulations of the stop-signal procedure. *Acta Psychol.* **112**, 105–142 (2003).
- D. Matzke, C. V. Dolan, G. D. Logan, S. D. Brown, E.-J. Wagenmakers, Bayesian parametric estimation of stop-signal reaction time distributions. *J. Exp. Psychol. Gen.* **142**, 1047–1073 (2013).
- L. Boucher, T. J. Palmeri, G. D. Logan, J. D. Schall, Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychol. Rev.* **114**, 376–397 (2007).
- G. D. Logan, M. Yamaguchi, J. D. Schall, T. J. Palmeri, Inhibitory control in mind and brain 2.0: Blocked-input models of saccadic countermanding. *Psychol. Rev.* **122**, 115–147 (2015).
- D. Matzke, J. Love, A. Heathcote, A Bayesian approach for estimating the probability of trigger failures in the stop-signal paradigm. *Behav. Res. Methods* **49**, 267–281 (2017).
- A. Åkerfelt, H. Colonius, A. Diederich, Visual-tactile saccadic inhibition. *Exp. Brain Res.* **169**, 554–563 (2006).
- H. Colonius, J. Özyurt, P. A. Arndt, Countermanding saccades with auditory stop signals: Testing the race model. *Vision Res.* **41**, 1951–1968 (2001).
- A. Gulberti, P. A. Arndt, H. Colonius, Stopping eyes and hands: Evidence for non-independence of stop and go processes and for a separation of central and peripheral inhibition. *Front. Hum. Neurosci.* **8**, 61 (2014).
- J. Özyurt, H. Colonius, P. A. Arndt, Countermanding saccades: Evidence against independent processing of go and stop signals. *Percept. Psychophys.* **65**, 420–428 (2003).
- P. G. Bissett, G. D. Logan, Balancing cognitive demands: Control adjustments in the stop-signal paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* **37**, 392–404 (2011).
- M. J. Nelson, L. Boucher, G. D. Logan, T. J. Palmeri, J. D. Schall, Nonindependent and nonstationary response times in stopping and stepping saccade tasks. *Atten. Percept. Psychophys.* **72**, 1913–1929 (2010).
- G. Dutilh, D. van Ravenzwaaij, S. Nieuwenhuis, H. L. J. van der Maas, B. U. Forstmann, E.-J. Wagenmakers, How to measure post-error slowing: A confound and a simple solution. *J. Math. Psychol.* **56**, 208–216 (2012).
- F. Verbruggen, G. D. Logan, Proactive adjustments of response strategies in the stop-signal paradigm. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 835–854 (2009).
- H. Levitt, Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* **49**, 467–477 (1971).
- G. D. Logan, J. Burkell, Dependence and independence in responding to double stimulation: A comparison of stop, change, and dual-task paradigms. *J. Exp. Psychol. Hum. Percept. Perform.* **12**, 549–563 (1986).
- P. G. Bissett, G. D. Logan, Selective stopping? Maybe not. *J. Exp. Psychol. Gen.* **143**, 455–472 (2014).
- H. Colonius, A. Diederich, Paradox resolved: Stop signal race model with negative dependence. *Psychol. Rev.* **125**, 1051–1058 (2018).
- J. E. Raymond, K. L. Shapiro, K. M. Arnell, Temporary suppression of visual processing in an RSVP task: An attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 849–860 (1992).
- H. Pashler, Dual-task interference in simple tasks: Data and theory. *Psychol. Bull.* **116**, 220–244 (1994).
- C. T. White, Temporal numerosity and the psychological unit of duration. *Psychol. Monogr. Gen. Appl.* **77**, 1–37 (1963).
- F. Maeda, R. Kanai, S. Shimojo, Changing pitch induced visual motion illusion. *Curr. Biol.* **14**, R990–R991 (2004).
- C. Spence, Crossmodal correspondences: A tutorial review. *Atten. Percept. Psychophys.* **73**, 971–995 (2011).
- M. O. Ernst, Learning to integrate arbitrary signals from vision and touch. *J. Vis.* **7**, 7 (2007).
- J. T. Townsend, F. G. Ashby, *The Stochastic Modeling of Elementary Psychological Processes* (Cambridge Univ. Press, 1983).
- P. G. Bissett, thesis, Vanderbilt University (2014).
- V. Lerche, A. Voss, Speed–accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates? *Behav. Res. Methods* **50**, 2568–2585 (2018).
- D. Matzke, S. Curley, C. Q. Gong, A. Heathcote, Inhibiting responses to difficult choices. *J. Exp. Psychol. Gen.* **148**, 124–142 (2019).
- F. Verbruggen, C. D. Chambers, G. D. Logan, Fictitious inhibitory differences: How skewness and slowing distort the estimation of stopping latencies. *Psychol. Sci.* **24**, 352–362 (2013).
- I. W. Eisenberg, P. G. Bissett, A. Zeynep Enkavi, J. Li, D. P. MacKinnon, L. A. Marsch, R. A. Poldrack, Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat. Commun.* **10**, 2319 (2019).
- V. V. Sochat, I. W. Eisenberg, A. Z. Enkavi, J. Li, P. G. Bissett, R. A. Poldrack, The experiment factory: Standardizing behavioral experiments. *Front. Psychol.* **7**, 610 (2016).
- P. G. Bissett, G. D. Logan, Post-stop-signal adjustments: Inhibition improves subsequent inhibition. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 955–966 (2012).
- D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* **67**, 1–48 (2014).
- D. Matzke, F. Verbruggen, G. D. Logan, in *Stevens’ Handbook of Experimental Psychology and Cognitive Neuroscience*, J. T. Wixted, Ed. (John Wiley & Sons Inc., 2018), pp. 1–45.

**Acknowledgments:** We would like to thank I. Eisenberg and J. Li for assistance with data acquisition. Preliminary research related to this paper was published as P.G.B.’s dissertation (36) under the guidance of G.D.L. **Funding:** This research was supported by grant number R01-EY021833-01 from the National Eye Institute. P.G.B. acknowledges support from the National Institute of Drug Abuse (F32DA041773). R.A.P. and P.G.B. acknowledge support from the



National Institute of Mental Health (R01MH117772). Data collection was also supported by the NIH Science of Behavior Change Common Fund Program through an award administered by the National Institute for Drug Abuse (NIDA) (UH2DA041713; principal investigators: L. A. Marsch and R.A.P.). **Author contributions:** P.G.B. and G.D.L. conceptualized the study. P.G.B. acquired the data and wrote the manuscript. P.G.B. and H.M.J. curated and analyzed the data. P.G.B., H.M.J., and R.A.P. visualized the data. P.G.B., R.A.P., and G.D.L. acquired funding for the research. R.A.P. and G.D.L. supervised the research. All authors developed the preliminary computational models and edited the manuscript. **Competing interests:** The authors declare that they have no financial or other competing interests. **Data availability:** All data needed to evaluate the conclusions in the paper are present in the paper and the Supplementary Materials. Additional data related to this paper is available on <http://doi.org/10.5281/zenodo.4432816>.

<http://doi.org/10.5281/zenodo.4432816>. All raw data, processed data, and analysis scripts are freely available at <http://doi.org/10.5281/zenodo.4432816>. Code for analyses, figure creation, table creation, and modeling are available at <http://doi.org/10.5281/zenodo.4432816>.

Submitted 26 October 2020

Accepted 29 January 2021

Published 17 March 2021

10.1126/sciadv.abf4355

**Citation:** P. G. Bissett, H. M. Jones, R. A. Poldrack, G. D. Logan, Severe violations of independence in response inhibition tasks. *Sci. Adv.* **7**, eabf4355 (2021).