bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Design issues and solutions for stop-signal data from the Adolescent Brain Cognitive Development [ABCD] study

Patrick G. Bissett, McKenzie P. Hagen, & Russell A. Poldrack

Department of Psychology, Stanford University

Corresponding Author: Patrick G. Bissett, pbissett@stanford.edu

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

The Adolescent Brain Cognitive Development (ABCD) study is an unprecedented longitudinal neuroimaging sample that tracks the brain development of over 10,000 9-10 year olds through adolescence. At the core of this study are the three tasks that are completed repeatedly within the fMRI scanner, one of which is the stop-signal task. In analyzing the available stopping experimental code and data, we identified a set of design issues that we believe significantly limit its value. These issues include but are not limited to: variable stimulus durations that violate basic assumptions of dominant stopping models, trials in which stimuli are incorrectly not presented, and faulty stop-signal delays. We present eight issues, show their effect on the existing ABCD data, suggest prospective solutions to the study organizers including task changes for future data collection, and suggest retrospective solutions for data users who wish to make the most of the existing data.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The Adolescent Brain Cognitive Development is the largest and most comprehensive long-term study of brain development and child health in the United States (Casey et al., 2018). The study includes 11,878 youth and their families and aims to understand the environmental, social, genetic and other biological factors that affect brain and cognitive development. This study was made possible by the Collaborative Research on Addiction at NIH (CRAN) including the National Institute of Drug Abuse, National Institute on Alcohol Abuse and Alcoholism, and National Cancer Institute in partnership with the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institute of Mental Health, National Institute of Minority Health and Health Disparities, National Institute of Neurological Disorders and Stroke, and the NIH Office of Behavioral and Social Sciences Research. CRAN granted $300 million in support to 21 research institutions across the United States to complete this study.

At the core of the ABCD study are the structural and functional MRI brain scans that occur biennially for each participant. The initial baseline scans have all been completed. The study organizers chose to include three cognitive tasks to be presented during fMRI acquisition: the monetary incentive delay task (Knutson et al., 2000), the emotional N-back task (Barch et al., 2013), and the stop-signal task (Logan & Cowan, 1984). In this manuscript we focus solely on the stop-signal task.

We analyzed behavioral data from the baseline scan of 7,321 of the 11,878 participants. This subset resulted from the following exclusion, which attempted to eliminate incomplete data or data that were already flagged as problematic by the ABCD study organizers. First, we attempted to download 8,811 participants from the “FastTrack Recommended Active Series” from the NIMH Data Archive, which has gone through some quality assurance for the associated imaging files by the ABCD study organizers. Of these, 8,776 files were successfully downloaded, but a subset did not include stop-signal data, leaving 7,906 subjects. Of these, only 7,347 included summary scores from the Stop Signal Task in the ABCD Data Release 2.0.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Finally, 26 subjects were removed who did not have two complete runs with 180 trials each, leaving us with a total of 7,321 complete datasets.

The stop-signal task is a primary paradigm used to understand response inhibition. It involves making a choice response to a go stimulus but attempting to stop that response when an infrequent stop signal occurs after a stop-signal delay (SSD). The dominant theoretical framework for understanding and interpreting stopping tasks is the Independent Race Model (Logan et al., 2014; Logan & Cowan, 1984), which assumes that a go process begins when a go stimulus occurs and races independently against a stop process that begins when the stop stimulus occurs. The go process finishing first results in a stop failure (an overt response), whereas the stop process finishing first results in stop success (no response).

The ABCD study is laudable for many reasons, not least of which is the dedication to making their data and materials openly available. This includes but is not limited to experimental code, trial-by-trial behavioral performance in each task, functional neuroimaging data, and structural neuroimaging data. Our research group aimed to use these data to understand the behavioral and neural underpinnings of response inhibition. In doing so, we first evaluated the experimental code and behavioral performance in the stop-signal task.

During our analyses, we found a nexus of issues with the ABCD experimental code and behavioral data. The issues are as follows: different go stimulus duration across trials, the go stimulus is sometimes not presented, faulty stop-signal delays, different stop signal duration for different SSDs, non-uniform conditional trial probabilities, trial accuracy incorrectly coded, SSD values start too short, and low stop trial probability. We judge these issues to vary from fundamental (e.g., different go stimulus duration across trials) to more minor (e.g., low stop trial probability). Indeed, some of the minor issues may reflect intentional design choices (e.g., low stop trial probability), but if so, we believe that those choices are suboptimal (for reasons that we lay out below). Additionally, we believe that the most fundamental issues, especially Issues 1-3, are incontrovertible errors, and in the aggregate we believe these eight issues significantly

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

compromise the utility of the stopping data in this study for understanding the construct of response inhibition.

We structure this paper as a series of issues. We order the issues roughly from what we judge to be most to least fundamental. For each, we outline the issue, demonstrate its effects in the ABCD data, suggest prospective solutions to the study organizers for future data collection, and suggest retrospective solutions to data users who want to make the most of the existing data.

Issue 1: Different go stimulus duration across trials

In order to try to lay out this most fundamental issue, we will first break down the trial structure (see Figure 1). 5/6 of all trials are go trials, in which a subject sees a go stimulus (a rightward- or leftward-pointing arrow), and makes one of two speeded responses based upon the direction of the arrow. The go stimulus is removed from the screen after 1000ms or when a response occurs, whichever comes first. On 1/6 of all trials, this go stimulus is replaced with the stop signal (a vertical arrow) after the go stimulus has been on the screen for the duration of the SSD; the stop signal is then presented for 300ms (but see Issue 4). Therefore, on go trials the go stimulus is on the screen for 1000ms or the response time (RT), whichever comes first, whereas on stop trials the go stimulus is on the screen for SSD or RT, whichever comes first, and is never concurrently on the screen with the stop stimulus. Mean go stimulus duration on go trials was 567ms (Standard deviation (SD) = 102ms), and mean go stimulus duration on stop trials was 226ms (SD = 113ms), so on average subjects had 341ms longer on go trials to apprehend the go stimulus (See Figure 2 for full distributions).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1. Stop-signal task trial structure

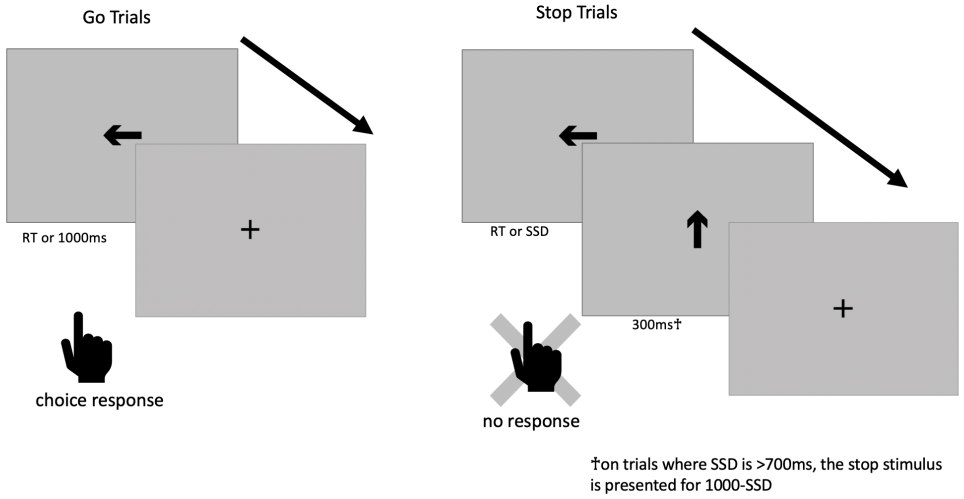
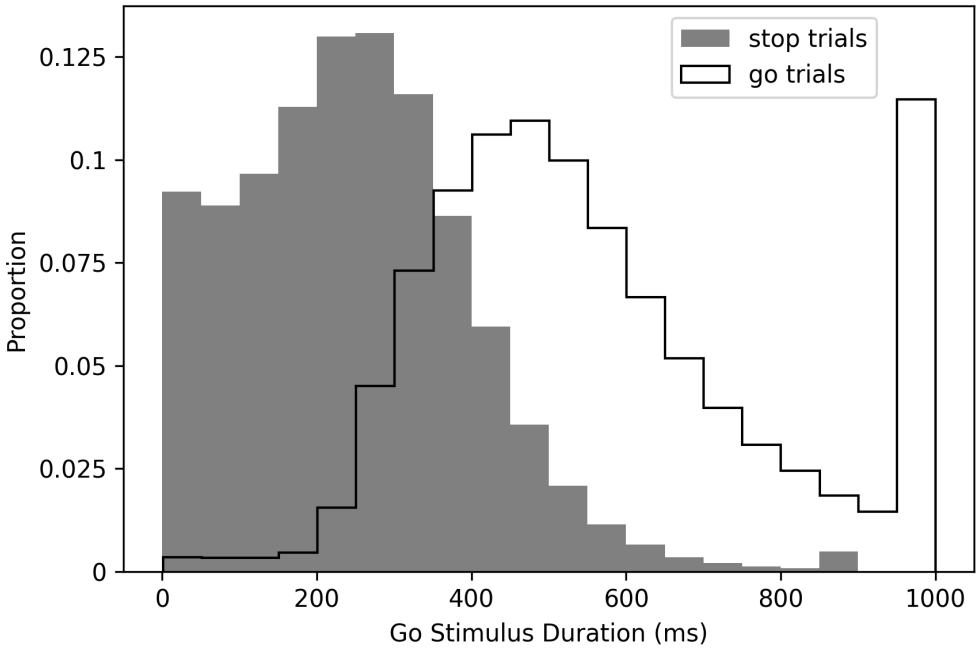


Figure 2. Proportion of Go Stimulus Durations on Go and Stop Trials



The main dependent variable in the stop-signal task is stop-signal reaction time (SSRT), which quantifies the latency of inhibition, but the calculation of SSRT requires application of a model because there is no overt response associated with a successful stop and thus stopping

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

latency cannot be directly measured. The dominant framework for capturing stop-signal performance and estimating SSRT is the Independent Race Model (Logan et al., 2014; Logan & Cowan, 1984). The Independent Race Model assumes context independence, which means that the go process and its finishing time is not affected by the presentation of the stop signal. Context independence is essential for calculating SSRT, because context independence allows one to assume that the full distribution of responses on go trials can stand in for the (unobservable) full distribution of go processes on stop trials. Additionally, violations of context independence contaminate other major dependent variables in the stop-signal task, including the inhibition function and stop-failure RT (Bissett et al., 2019; Logan & Cowan, 1984).

In both simple (e.g., Froeberg, 1907) and choice (e.g,. Kaswan & Young, 1965) reaction time, there is evidence that shorter duration stimuli yield slower responses. This effect of stimulus duration on response speed occurs even for stimuli presented for hundreds of milliseconds (Kahneman & Norman, 1964; Kaswan & Young, 1965). This relates to Bloch's Law (Bloch, 1885), which states that intensity and duration can be traded off for shorter duration stimuli. For example, reducing stimulus duration by half is equivalent to reducing intensity by half. This also relates to Pieron's law, which states that response times decrease with stimulus intensity (Pieron, 1914). Therefore, having shorter durations go stimuli on stop than go trials is akin to having a lower intensity go stimulus on stop than go trials trials, which slows RT. Taken together, a long history of work suggests that (all else being equal) shorter duration visual stimuli tend to yield slower responses.

In the ABCD stopping experiment, go stimuli are presented for much shorter durations on stop than on go trials, so the work described in the preceding paragraph suggests that the go process will be faster on go trials than on stop trials. However, in order to extract SSRT, one must make the assumption that the go process is the same on go and stop trials (i.e., context independence). Therefore, we expect that context independence is violated in the ABCD dataset, which would contaminate major dependent variables in the stop signal tasks including

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SSRT estimates. Additionally, because go stimuli on stop trials are presented for a duration equal to SSD, the degree to which violations of context independence occur are likely to differ across SSDs. When SSD is short (e.g., 50ms), context independence may be more severely violated because the difference in go stimulus duration between stop (e.g., 50ms) and go (up to 1000ms) trials is so large.

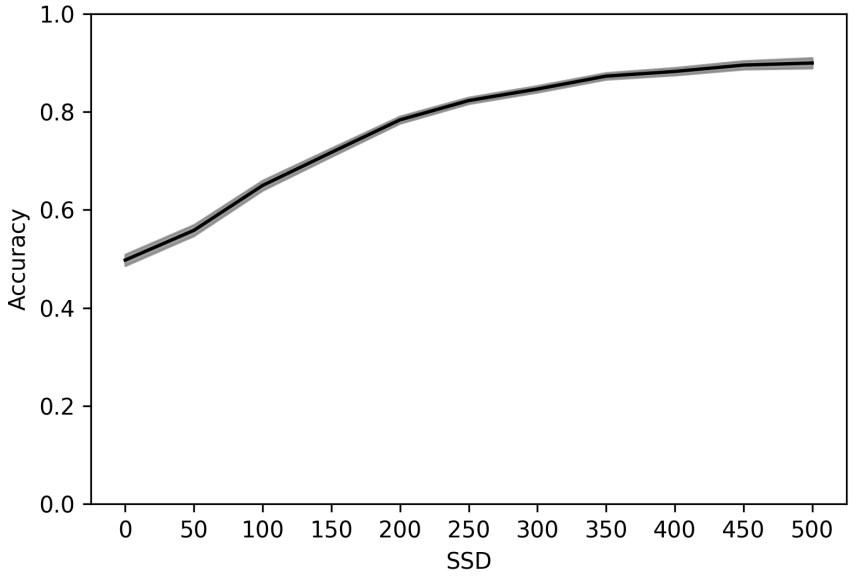
Evidence for Issue 1 in ABCD Data. The primary way to evaluate context independence is to compare reaction times on go trials to reaction times on stop-failure trials. If the former is longer than the latter then context independence is taken to hold (Bissett et al., 2019; Logan & Cowan, 1984; Verbruggen et al., 2019). On average across all subjects, stop-failure RT (*M* = 456ms, SD = 110ms) was shorter than overt responses on go trials (*M* = 544ms, SD = 95ms), 95% confidence interval of the difference [86.0, 89.2]. However, for 447 of the 7231 subjects (6.2% of all subjects), mean stop-failure RT was longer than mean RT on overt responses in go trials, suggesting that a subset of subjects violated context independence. Though note that the comparison of stop-failure and go RT is a conservative measurement that will only show violations of context independence if they are severe (Bissett et al., 2019).

In order to further evaluate whether the go process is impaired on stop trials as a result of the shorter go stimulus, we compared choice accuracy on stop-failure trials with choice accuracy on all overt (non-omission) go trials. Stop-failure trials had much lower accuracy (78%) than overt go trials (89%), 95% confidence interval of the difference [10.6%, 11.2%]. Additionally, choice accuracy on stop-failure trials was increasingly impaired at shorter SSD (see Figure 3). In other datasets, choice accuracy on stop-failure trials, even at shorter SSDs, tends to be similar to overt go trials (Bissett et al., 2019), which suggests that the impaired go accuracy on stop-failure trials in the present study results from the shorter go stimulus durations on stop trials in this task. Therefore, this lower choice accuracy is consistent with the go process

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

being fundamentally impaired on stop trials compared to go trials, particularly at short SSDs, violating the assumption of context independence.

Figure 3. Choice response accuracy on stop-failure trials across SSD. 95% confidence intervals are presented as gray confidence bands.



Prospective Suggestions For Issue 1. In order to address Issue 1, we would recommend that the ABCD study organizers present the go stimulus for a fixed period of time on every trial, perhaps 1000ms. When a stop signal occurs, it should not replace but it should be presented in addition to the go stimulus. If the study designers would like to keep all stimuli in the center of the screen, they could superimpose the stop stimulus around the arrow (e.g., a circle). Therefore, the go stimulus would be identical in form, size, and duration for all go and stop trials. This should eliminate any possibility that different go durations drive different go processes, violating context independence and contaminating dependent variables.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Retrospective Suggestions For Issue 1. The main reason that we have suggested that Issue 1 is the most fundamental is if one assumes that shorter go stimuli on stop trials yield slower, impaired go processes when compared with the longer go stimuli on go trials, which we believe is reasonable assumption given over 100 years of response time research (Bloch, 1885; Froeberg, 1907; Kahneman & Norman, 1964; Kaswan & Young, 1965; Pieron, 1914), then context independence is violated and major dependent variables in the stop-signal task are contaminated. This also influences common neuroimaging task contrasts like stop success versus go and stop failure versus go, as this introduces additional differences between trial types, including go stimulus duration and go stimulus reaction time, that will contaminate the ability of the contrast to isolate processes of interest like response inhibition. Given the above, if analyzing or disseminating existing ABCD stopping data, we would recommend caution in drawing any strong conclusions from the stopping data, and any results should be clearly presented with the limitation that the task design encourages context dependence and therefore stopping behavior (e.g., SSRT) and neuroimaging contrasts may be contaminated.

In addition, we suggest two other practical suggestions. First, in line with consensus guidelines (Verbruggen et al., 2019), we would recommend removing the subjects who have severe violations as evidenced by mean stop-failure RT > mean no-stop-signal RT (447 out of the 7231 that we analyzed, 6%). Second, given our suggestion that violations may be most severe at short SSDs (because the difference in go stimulus duration between go and stop trials is maximal at short SSDs), and given work suggesting violations of context independence may be most severe at short SSDs (Åkerfelt et al., 2006; Bissett et al., 2019; Colonius et al., 2001; Gulberti et al., 2014; Özyurt et al., 2003), we would recommend that any results be verified when only longer SSDs are used, perhaps only SSDs > 200ms. However, note that removing stop trials that have SSDs <=200ms would remove 51.5% of all stop trials, leaving less than an average of 30 stop trials per subject. Neither of these two suggestions resolve Issue 1 but they should eliminate the subjects and trials most likely to show severe violations, respectively. In

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

order to resolve Issue 1, the different go durations that encourage context dependence must be eliminated (see prospective suggestions), or new models for stopping must be developed to accommodate context dependence (Bissett et al., 2019), the latter of which we consider to be of utmost importance to advancing the stop-signal literature.

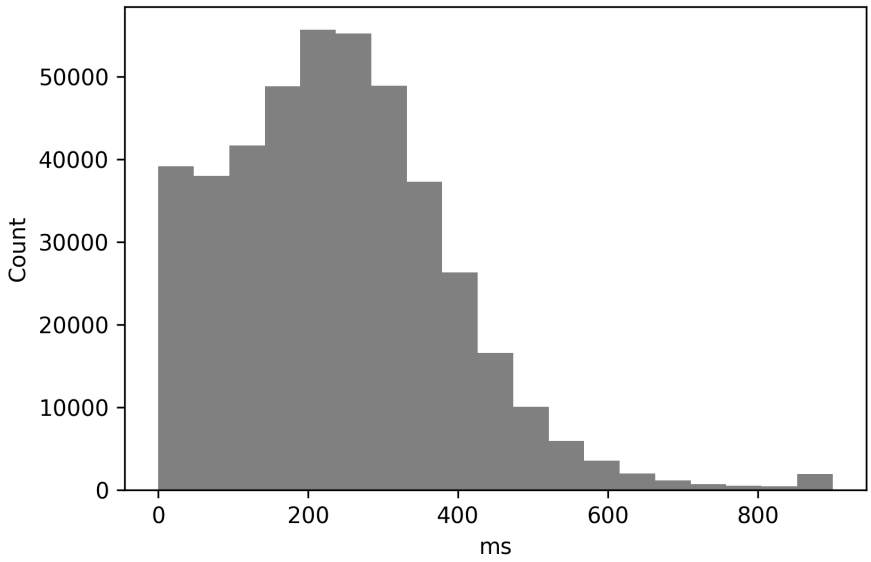
Issue 2: Go stimulus sometimes not presented

Issue 2 can be seen as a special case of Issue 1. When the stop signal occurs, it replaces the go stimulus, and SSD can reduce to 0ms. When the SSD is 0ms, a go stimulus does not occur on the stop trial. This is an issue because extant models of stopping assume a race between two processes (Logan & Cowan, 1984), and there is not a go stimulus to initiate a go process or to drive one of the two choice responses. The absence of a go stimulus may make successfully stopping trivial (as the go process never started) or may confuse subjects given these trials have a fundamentally different structure (i.e., one stimulus instead of two).

Evidence for Issue 2 in ABCD Data. 9% of all stop trials had an SSD of 0ms (see Figure 4). In a typical stop-signal experiment, a 0ms SSD trial means that the go and stop stimuli onset at the same time and are presented concurrently for a period of time. This tends to produce a very high stop success rate at a 0ms SSD. In these data, we found that the stop success rate at the 0ms SSD was 60.5%. This surprisingly low value may have resulted from subjects being confused as to how to approach these trials. It also may have been driven by subjects who stopped inhibiting their responses. Additionally, see Issue 3 for a contribution explanation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 4. Histogram of Stop-Signal Delays



Prospective Suggestions for Issue 2. The suggestion from Issue 1 will also naturally resolve Issue 2. When SSD is 0ms, the arrow go stimulus would be presented at the same time as a secondary stop signal presented around it.

Retrospective Suggestions for Issue 2. Given there is not a go stimulus to drive a choice towards a specific go response on 0ms SSD trials, we recommend that these 0ms SSD trials be removed from any behavioral or neuroimaging analyses. The complete lack of a go stimulus is at least as fundamental as 1, but this only occurs on 9% of all stop trials, so this issue can be addressed by removing these trials.

Issue 3: Faulty stop-signal delays

The ABCD study uses the common 1-up/1-down tracking algorithm (Levitt, 1971) to determine SSD, which involves increasing SSD by 50ms whenever a subject successfully stops and decreasing SSD by 50ms whenever a subject fails to stop. However, if SSD is 50ms and a

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

subject makes a response that is faster than 50ms (e.g., 25ms), then this triggers a glitch in the experiment code in which all subsequent stop trials have a response erroneously recorded at that same timestamp (e.g., 25ms). Therefore, all subsequent stop trials are treated as stop failures (because this one initial response is recorded for all subsequent stop trials), and SSD remains stuck at the minimum of 0ms for the remainder of that subject's stopping dataset.

Evidence for Issue 3 in ABCD Data. The triggering condition for this coding error is very specific and occurs somewhat rarely. In the entire sample, only 2.7% of subjects had this specific problem. However, this issue interacts with Issue 2 and partially explains the low stop accuracy at 0ms SSD. Once these trials are excluded, the stop accuracy at 0ms increases from 60.5% to 63.1%.

Prospective Suggestions for Issue 3. This appears to be a coding error, so we would recommend that the error be resolved by ensuring that any response on a given trial does not propagate forward incorrectly as also being a response with the same duration in subsequent trials.

Retrospective Suggestions for Issue 3. Given the rarity of this problem, in addition to the severity of its manifestation (i.e., it ensures that every subsequent stop trial is recorded as a stop-failure with a 0ms SSD, irrespective of actual behavior), we recommend removing the 2.7% of subjects who trigger this issue.

Issue 4: Different stop signal duration for different SSDs

The Independent Race Model (Logan et al., 2014; Logan & Cowan, 1984) not only assumes *go* context independence, but also *stop* context independence, which means that the stop process is the same across SSDs. However, in an issue that mirrors Issue 1, the stop

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

signal is presented for different durations at different SSDS. If SSD is <= 700ms, then the stop signal is presented for 300ms. If SSD is >700ms, then the stop signal is presented for 1000-SSD, with a maximum SSD of 900 (and therefore a minimum stop duration of 100ms).

Evidence for Issue 4 in ABCD Data. SSDs > 700ms are rare in these data (see Figure 4, 0.8%). This rarity makes it difficult to compute SSRT exclusively at these longer delays to evaluate whether the stop process may be changed as a result of the stop signal being presented for a shorter duration at the longest SSDs.

Prospective Suggestions for Issue 4. We recommend a fixed stop-signal duration across all SSDs, perhaps 300ms.

Retrospective Suggestions for Issue 4. Given that SSDs that are > 700ms are so rare (0.8%), we suggest removing them from any analyses. This will eliminate their contamination on any averages without resulting in significant data loss.

Issue 5: Non-uniform conditional trial probabilities

In stop-signal tasks, the default way to determine trial sequences is to randomly or quasi-randomly (e.g., random without replacement from a fixed pool of trials to ensure the same number of stop trials in each block) select which subset of trials will include a stop signal. However, in fMRI task-based experiments, it is common to select trial sequences in order to optimize power of finding an effect or a difference between contrasting conditions (Durnez et al., 2017; Kao et al., 2009; Liu et al., 2001). Increasing detection power can be achieved by adjusting the conditional probabilities such that trials are more likely to follow the same trial type than change to a new trial type (i.e. the design becomes more block-like). However, this push for greater power needs to be weighed against known expectancy effects in the stop-signal task

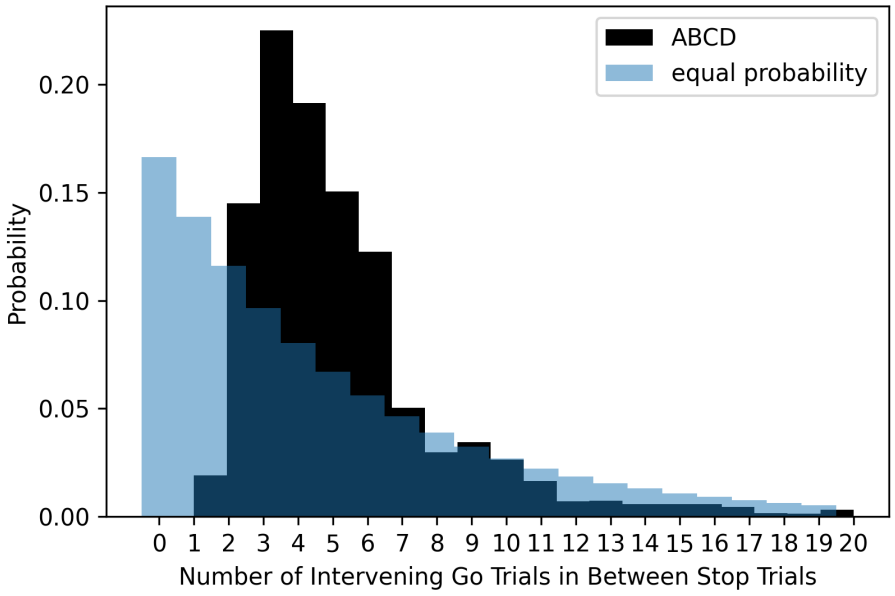
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Bissett & Logan, 2012b), which can change the processes involved in a task when subjects can predict what is coming. Indeed, some modern software (Durnez et al., 2017) allows the researcher to explicitly trade off the competing goals of power and avoiding trial-by-trial contingencies.

The ABCD stopping study has a highly non-uniform distribution of transition probabilities between go and stop trials. Given the overall probability of a stop signal is 1/6 or .167 and a uniform conditional probability across trials, the probability of a stop signal given the immediately preceding trial was a stop signal should be .167. In the ABCD task, this value is 0, as stop trials never repeat. Additionally, stop trials almost never occur if there was a stop trial two trials before (conditional probability of 1.8%). We plot the conditional probability of a stop signal given the most recent stop signal, both in the ABCD data and for an example where the probability of a stop signal was equal across all trials (see Figure 5). Subjects have the potential to learn this contingency that stop trials never repeat and seldom occur after one intervening go trial. If this is learned, this should speed RT and perhaps slow SSRT on the trial that follows a stop signal (Bissett & Logan, 2012b), which is the opposite to the sequential adjustment observed when conditional probabilities are equal across all trials (Bissett & Logan, 2011, 2012a; Rieger & Gauggel, 1999). Therefore, by making such drastic adjustments to the conditional probabilities, this may change the way that subjects balance going and stopping across trials.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

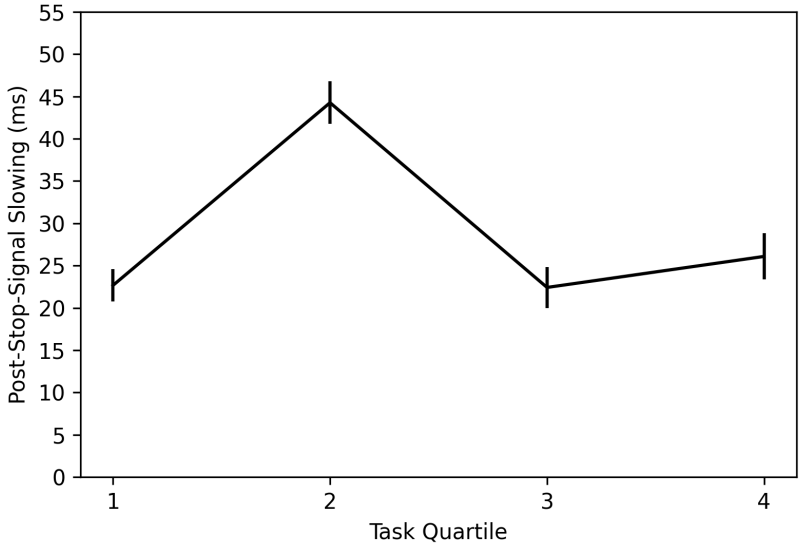
Figure 5. The probability of different numbers of intervening go trials between successive stop trials in the ABCD dataset (dark shading) versus expected probability distribution if stop trials were equally probable on all trials (light blue shading).



Evidence for Issue 5 in ABCD Data. In order to investigate whether subjects are learning this trial contingency and using it to change behavior, we broke the stop trials in the task into quantiles over time (i.e., first 15 stop trials, 2nd 15, 3rd 15, and 4th 15) and computed post-stop-signal RT changes in each. If subjects are learning this contingency and changing their behavior in response to it, we should see post-stop-signal slowing (longer RT on the trial immediately following a stop signal compared to RT on the trial immediately preceding a stop signal) reduce over the experiment, perhaps giving way to post-stop-signal speeding towards the end of the experiment. This is not what we found (see Figure 6) as post-stop-signal slowing was around 20ms for each quartile except the 2nd, in which it was 40ms. Therefore, it appears that in this baseline session, subjects did not learn this contingency or did not change their behavior in response to it.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 6. Post-stop-signal slowing across the four quantiles of the ABCD stopping data. Error bars are 95% confidence intervals.



Prospective Suggestions for Issue 5. In order to eliminate the possibility of subjects learning this contingency and changing their behavior, we suggest that the ABCD study make the conditional probabilities fully uniform or at least much more uniform, most importantly allowing for the possibility of immediate repeats. Though learning effects were not apparent in the baseline session, subjects may learn this contingency and implement more severe behavioral changes in subsequent sessions as they have greater exposure to this contingency. Additionally, as subjects mature, they may develop the ability to learn this contingency and use it to change their behavior.

Retrospective Suggestions for Issue 5. The two trials after a stop signal never or virtually never include a stop signal, so they could have involved fundamentally different processes from normal go trials. In essence, if this contingency is learned, the trial after a stop trial becomes a

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

"certain-go" trial, which has been shown to have different behavioral performance and neural responses (Chikazoe et al., 2009). However, we did not find evidence that subjects learned and adjusted their behavior based upon this contingency in this baseline session. Additionally, removing the two go trials after a stop signal would involve huge data loss (1/3 of all trials), so this does not seem warranted. Therefore, we do not think that Issue 5 will affect data users, at least in the baseline session.

Issue 6: Trial accuracy incorrectly coded

In the stop-signal task, there are two types of trials, go and stop, and each can be correct or an error. On a go trial, subjects can be correct by making the appropriate choice response or make an error by either making the incorrect choice response or by omitting their response. On stop trials, subjects can be correct by omitting their response or be incorrect by making a choice response. In three out of four of these trial outcomes, there are errors in how the ABCD data categorizes trials.

Evidence for Issue 6 in ABCD Data. All 212,359 trials that are categorized as incorrect stops have one or more overt responses, so these all appear to be categorized correctly in the output. 10 of the 304,756 trials (.003%) classified as incorrect go trials appear to actually have the correct go response, but we were unable to ascertain why these trials were miscategorized. 49,150 of the 1,864,544 trials (2.6%) classified as correct go trials are incorrect go trials. This appears to result from overwriting the first response with any subsequent response that is recorded later in the trial, which we understand to be inappropriate for speeded tasks generally and perhaps especially for a stop-signal task in which one wants to measure the speed of the first go process that completes, not a subsequent go process. Finally, 911 of the 221,503 trials (.4%) classified as correct stops are incorrect stops, as they include one or more overt response. This final class of errors results in SSD increasing by 50ms when it should instead

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

decrease by 50ms. 127 of these 911 appear to result from a response occurring at the exact millisecond that one of the constituent stimuli occur in the trial. We were unable to ascertain why the other 784 of the 911 were recorded as correct stops, given there was one or more recorded responses on these stop trials.

Prospective Suggestions for Issue 6: Issue 6 appears to result from errors in the experimental code. Therefore, Issue 6 should be fixable by adjusting the experimental code. For example, in their code there are three constituent periods within a stop-signal trial (the SSD, the 100-300ms in which the stop-signal is on the screen, and the fixation period after the stop signal), and if the recorded RT = 0ms in all 3 (either because there was no response or if it occurred at the exact millisecond that the stimulus was presented), then trials are categorized as a correct stop. However, on 127 “correct stop” trials, there is a response at the exact millisecond that one of these stimuli is presented. It is still recorded as a response in the output, so instead of categorizing trials and increasing SSD by ensuring RT = 0ms outputs in each period, the trials should be categorized as a correct stop if no response is recorded in any of these three periods.

Retrospective Suggestions for Issue 6: For data users, Issue 6 can largely be resolved by recategorizing each trial type (which we did for all analyses within this work), according to the rules specified in the paragraph where we introduced Issue 6, before analyzing the data. However, as we briefly mentioned above, one effect of this error cannot be undone, because 911 “correct stop” trials have overt responses, so they should have been categorized as incorrect stop trials and therefore SSD should have reduced by 50ms instead of being increased by 50ms. However, given how infrequent these are (.4% of all correct stops), this is unlikely to have a significant effect on the SSD tracking algorithm, and therefore we judge this issue to be more minor. These 911 should be recategorized as incorrect stops before analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

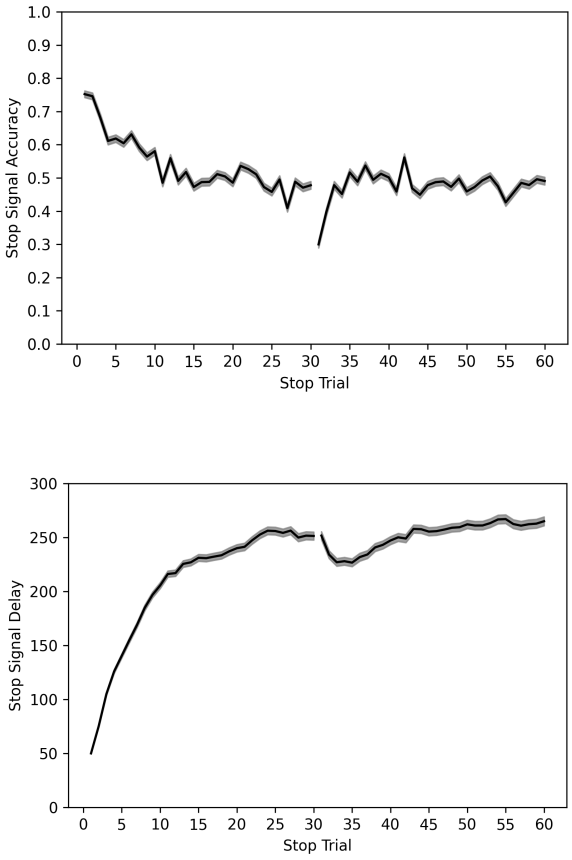
Issue 7: SSD values start too short

The 1-up/1-down SSD tracking algorithm (Levitt, 1971) is an efficient way to sample the intermediate part of the inhibition function, which is the most informative for constraining SSRT estimates (Band et al., 2003; Verbruggen et al., 2019). In order to sample the intermediate part of the inhibition function for the maximal number trials, the SSD should begin at a value such that go RT = SSD + SSRT. For example, if the expected mean RT in the sample is 500ms and the expected SSRT is 250ms, then SSD should start at 250ms. Otherwise, the initial stop trials will have a high stop success rate, which is less informative for constraining estimates of SSRT and could drive strategy shifts in the experiment (e.g., subjects might de-emphasize the stop process at the beginning of the experiment if they recognize that any stop signals will occur with such a short SSD that even a slower stop process could beat the go process).

Evidence for Issue 7 in ABCD Data. In the ABCD dataset, SSD starts at 50ms. We plot the stop accuracy and SSD (see Figure 7) across each of the 60 stop trials for all subjects. This shows that stop accuracy tends to be high (and SSD low) in the first ~10 stop trials but then stabilizes around .5, as desired, for the remaining ~50 stop trials. The one salient exception is trial 31, which is the first trial of their second block or session (some subjects do both blocks back-to-back and others do one block in one scanning session and the second block in a subsequent scanning session), which has a very low stop accuracy.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 7. Average stop-signal accuracy (a) and stop-signal delay (SSD, b) across subjects for each of the 60 stop trials. 95% confidence intervals are presented as gray confidence bands.



Prospective Suggestions for Issue 7. In order to create a more efficient design that eliminates this initial part of each section when many subjects have high stop success rates that are less informative for SSRT estimates, we recommend starting each session with a more canonical 250ms SSD. This is similar to the mean SSD (230ms) and final SSD (265ms) across subjects, so should sample the intermediate part of the inhibition function from the start for the most possible subjects.

Retrospective Suggestions for Issue 7. It is unclear whether subjects are recognizing these very short SSDs in the beginning of each session and adjusting their performance or

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

strategies in response to them. For example, as suggested above, subjects could deemphasize the stop process at the beginning of each session. Therefore, we would recommend a cautious approach of ensuring that any conclusions do not qualitatively change when initial trials (perhaps the first seven stop trials with stop success rates >.6) are removed from analyses.

Issue 8: Low stop trial probability

Stop probability is commonly .25, .33, or a similar value in stop-signal studies of the stop-signal task(Verbruggen et al., 2019). Increasing stop probability increases power for SSRT calculation, stop-based imaging contrasts, and other analyses (e.g., the inhibition function(Logan & Cowan, 1984)). Values greater than ⅓ are somewhat rare, perhaps because there is a belief that high stop probabilities may fundamentally change the stop process (but see Bissett & Logan, 2011; Logan & Burkell, 1986).

Evidence for Issue 8 in ABCD Data. This ABCD study uses a .167 stop probability, which is unusually low.

Prospective Suggestions for Issue 8. We suggest increasing the stop probability to .25, which would put this data acquisition more in line with the literature and would provide 50% more stop trials to increase analytical power without increasing the length of the data acquisition. We do not know of any research that suggests any benefits of any kind when reducing stop probability to as low as .167.

Retrospective Suggestions for Issue 8. We do not know of any published evidence that suggests that low stop trial probability may contaminate the data. Therefore, we do not believe that Issue 8 will affect data users, except that the low stop trial probability means that there are

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

fewer stop signals to constrain and provide accurate estimates of stopping performance including SSRT.

Conclusion

The ABCD study is an unprecedented investment towards understanding human cognition and brain development, and the stop-signal data is a key part of this dataset as one of only three tasks that are acquired within each neuroimaging session. Above, we laid out eight issues starting with more fundamental and moving towards more minor that, when taken together, we believe to significantly undermine existing ABCD stopping data. We have not evaluated the emotional N-back or monetary incentive delay tasks, but we hope that future work will review the experimental code and data from these tasks. Only a subset of the entire ABCD data have been acquired. We have offered prospective suggestions that we hope will be considered and implemented before future ABCD data is acquired. We recognize that implementing these changes will result in a lack of continuity between how the stop task is implemented across the scans for each subject, but we believe that these issues are significant enough that resolving them outweigh the benefit of task continuity. Additionally, some issues involve learning contingencies embedded within the experiment (e.g., Issue 5), so they could exacerbate as subjects have repeated exposures to the task over this longitudinal study. We have also provided retrospective suggestions to data users to contextualize any results that come from these existing data. These suggestions aim to help make the most of the ABCD stopping data, both past and future.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data and Materials Availability: The ABCD dataset is openly available through the NIH Data Archive. Analysis code is available at: https://github.com/mckenziephagen/ABCD_Stop_Signal.

Acknowledgements: We would like to thank Sage Hahn, Hugh Garavan, and their team for identifying an error in a previous version of our manuscript and code that resulted in an inflation in our stop-failure RT estimates. We would also like to thank Stephen J. Hanson, Paul Jaffe, Henry Jones, Luke Stoeckel, and Jason Zamorano for helpful comments on an earlier draft.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Åkerfelt, A., Colonius, H., & Diederich, A. (2006). Visual-tactile saccadic inhibition. *Experimental Brain Research*, *169*(4), 554–563. https://doi.org/10.1007/s00221-005-0168-x

Band, G. P. H., van der Molen, M. W., & Logan, G. D. (2003). Horse-race model simulations of the stop-signal procedure. *Acta Psychologica*, *112*(2), 105–142. https://doi.org/10.1016/S0001-6918(02)00079-3

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., & Van Essen, D. C. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189. https://doi.org/10.1016/j.neuroimage.2013.05.033

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

Bissett, P. G., & Logan, G. D. (2012a). Post-stop-signal adjustments: Inhibition improves subsequent inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 955–966. https://doi.org/10.1037/a0026778

Bissett, P. G., & Logan, G. D. (2012b). Post-stop-signal slowing: Strategies dominate reflexes and implicit learning. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 746–757. https://doi.org/10.1037/a0025429

Bissett, P. G., Poldrack, R. A., & Logan, G. D. (2019). *Title: Severe and pervasive violations of independence in response inhibition tasks*. Preprint at https://psyarxiv.com/kpa65/

Bloch, A.-M. (1885). Experiences sur la vision [Experiments in vision]. *Comptes Rendus de Seances de La Societe de Biologie, Paris*, *37*(28), 493–495.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., … Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Chikazoe, J., Jimura, K., Hirose, S., Yamashita, K. -i., Miyashita, Y., & Konishi, S. (2009). Preparation to Inhibit a Response Complements Response Inhibition during Performance of a Stop-Signal Task. *Journal of Neuroscience*, *29*(50), 15870–15877. https://doi.org/10.1523/JNEUROSCI.3645-09.2009

Colonius, H., Özyurt, J., & Arndt, P. A. (2001). Countermanding saccades with auditory stop signals: Testing the race model. *Vision Research*, *41*(15), 1951–1968. https://doi.org/10.1016/S0042-6989(01)00084-0

Durnez, J., Blair, R., & Poldrack, R. A. (2017). *Neurodesign: Optimal Experimental Designs for Task fMRI* [Preprint]. Neuroscience. https://doi.org/10.1101/119594

Froeberg, S. (1907). *The relation between the magnitude of stimulus and the time of reaction*.

Gulberti, A., Arndt, P. A., & Colonius, H. (2014). Stopping eyes and hands: Evidence for non-independence of stop and go processes and for a separation of central and peripheral inhibition. *Frontiers in Human Neuroscience*, *8*.

Kahneman, D., & Norman, J. (1964). The time-intensity relation in visual perception as a function of observer's task. *Journal of Experimental Psychology*, *68*(3), 215–220. https://doi.org/10.1037/h0046097

Kao, M.-H., Mandal, A., Lazar, N., & Stufken, J. (2009). Multi-objective optimal experimental designs for event-related fMRI studies. *NeuroImage*, *44*(3), 849–856. https://doi.org/10.1016/j.neuroimage.2008.09.025

Kaswan, J., & Young, S. (1965). Effect of luminance exposure duration, and task complexity on

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reaction time. *Journal of Experimental Psychology*, *69*(4), 393–400. https://doi.org/10.1037/h0021741

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI Visualization of Brain Activity during a Monetary Incentive Delay Task. *NeuroImage*, *12*(1), 20–27. https://doi.org/10.1006/nimg.2000.0593

Levitt, H. (1971). Transformed up-down methods for psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.

Liu, T. T., Frank, L. R., Wong, E. C., & Buxton, R. B. (2001). Detection Power, Estimation Efficiency, and Predictability in Event-Related fMRI. *NeuroImage*, *13*(4), 759–773. https://doi.org/10.1006/nimg.2000.0728

Logan, G. D., & Burkell, J. (1986). Dependence and Independence in Responding to Double Stimulation: A Comparison of Stop, Change, and Dual-Task Paradigms. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 549–563.

Logan, G. 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.

Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). *On the Ability to Inhibit Thought and Action: General and Special Theories of an Act of Control*. *121*(1), 66–95.

Özyurt, J., Colonius, H., & Arndt, P. A. (2003). Countermanding saccades: Evidence against independent processing of go and stop signals. *Perception & Psychophysics*, *65*(3), 420–428. https://doi.org/10.3758/BF03194573

Pieron, H. (1914). Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices. *Annee Psychol.*, *22*, 17–96.

Rieger, M., & Gauggel, S. (1999). Inhibitory after-effects in the stop signal paradigm. *British Journal of Psychology*, *90*(4), 509–518. https://doi.org/10.1348/000712699161585

Verbruggen, F., Aron, A. R., Bissett, P. G., Brockett, A. T., Brown, J. W., Chamberlain, S. R.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.08.084707; this version posted July 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Chambers, C. D., Colonius, H., Colzato, L. S., Corneil, B. D., Coxon, J. P., Dupuis, A., Eagle, D. M., Garavan, H., Greenhouse, I., Heathcote, A., Huster, R. J., Jahfari, S., Kenemans, J. L., … 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.