Optimal go/no-go ratios to maximize false alarms

Michael E. Young¹ · Steven C. Sutherland² · Anthony W. McCoy¹

Published online: 29 June 2017
© Psychonomic Society, Inc. 2017

Abstract Despite the ubiquity of go/no-go tasks in the study of behavioral inhibition, there is a lack of evidence regarding the impact of key design characteristics, including the go/no-go ratio, intertrial interval, and number of types of go stimuli, on the production of different response classes of central interest. In the present study we sought to empirically determine the optimal conditions to maximize the production of a rare outcome of considerable interest to researchers: false alarms. As predicted, the shortest intertrial intervals (450 ms), intermediate go/no-go ratios (2:1 to 4:1), and the use of multiple types of go stimuli produced the greatest numbers of false alarms. These results are placed within the context of behavioral changes during learning.

Keywords go/no-go · Signal detection theory · Optimal design · Inhibition · Self-control · EEG

Go/no-go tasks (Drewe, 1975; Luria, 1973) involve a series of decisions in which participants are tasked with responding to one class of stimuli (the go stimuli) but not to another class of stimuli (the no-go stimuli). Because go trials outnumber no-go trials in the typical design, the task elicits a general tendency to respond that often results in inadvertent responses in the presence of a no-go stimulus. As a result, go/no-go tasks are frequently used to assess behavioral inhibition, under the assumption that people who have a higher false alarm rate

(responding to the no-go stimuli) have difficulties with inhibiting a prepotent response.

To assess sensitivity to the stimulus characteristics differentiating the go from the no-go stimuli and to measure response bias or the general tendency to respond, data from a go/no-go experiment can be subjected to a signal detection theory (SDT) analysis (Green & Swets, 1966; Macmillan & Creelman, 2004; Swets, Dawes, & Monahan, 2000). In SDT, responses are placed into four classes, (a) hits or making the response in the presence of a go stimulus, (b) misses or failing to respond in the presence of a go stimulus, (c) false alarms or responding to a no-go stimulus, and (d) correct rejections or not responding in the presence of a no-go stimulus. High sensitivity is evidenced by a large number of hits and correct rejections relative to misses and false alarms. Bias is evidenced by a greater or lesser tendency to respond relative to the response likelihood that is consistent with the ratio of go to no-go trials (a bias of zero). In general, difficulty discriminating stimuli will decrease sensitivity, and shifts in the costs/benefits of the four response classes or changing the base rate of the two stimulus classes (go vs. no-go) will alter the response bias (Swets et al., 2000).

In a survey of the literature, we were surprised by the considerable variability in how go/no-go tasks are implemented. The intertrial interval, or ITI, varies across experiments (from as little as 600 ms to 3,000 ms or more; e.g., Eimer, 1993; Garavan, Ross, Murphy, Roche, & Stein, 2002; Menon, Adleman, White, Glover, & Reiss, 2001), the trial-to-trial ITI is sometimes constant and sometimes not (e.g., Eimer, 1993; Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003), and the ratio of go to no-go trials differs across a wide range (most commonly from 3:1 to 1:1, but it can be as large as 15:1 or as low as 1:4; e.g., Criaud & Boulinguez, 2013; Eimer, 1993; Garavan et al., 2002; Kiefer, Marzinzik, Weisbrod, Scherg, & Spitzer, 1998;

✉ Michael E. Young
michaelyoung@ksu.edu

¹ Department of Psychological Sciences, Kansas State University, 492 Bluemont Hall, Manhattan, KS 66506-5302, USA

² University of Houston–Clear Lake, Houston, TX, USA

Mostofsky et al., 2003; Nieuwenhuis et al., 2003). Additionally, the number of different types of go or no-go stimuli has varied from 1:1 (e.g., M vs. W, Eimer, 1993), to 2:1 (e.g., green and gray go stimuli vs. yellow no-go stimuli; Chikazoe et al., 2009), to many:1 (all letters except X vs. X; Menon et al., 2001). Additional variations abound.

This lack of consistency across published studies makes it difficult for an investigator to determine how to design a go/no-go study for particular research purposes, and this often results in aborted attempts at producing the desired conditions. A task may be too easy for the participant, thus producing few mistakes that can be subjected to analysis, or the task may lack the right structure to produce the desired number of a particular type of mistake, such as misses or false alarms. Thus, we sought to identify the effects of two key task variables, difficulty and go/no-go trial ratio, on task sensitivity, response bias, and the production of false alarms. Our particular interest was in the conditions that increase the number of false alarms, in order to support future studies of the neural basis of errors. Approaches such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) require a sufficient number of trials of the target types to allow for proper analysis, and go/no-go tasks produce very few false alarms when the task is relatively easy or the go/no-go ratio is too low.

Difficulty and go/no-go ratio

Our research study focused on two task variables, difficulty and the ratio of go to no-go trials. We used d' as our measure of sensitivity: the z -transformed hit rate minus the z -transformed false alarm rate (Macmillan & Creelman, 2004). A d' of zero indicates no ability to respond discriminately to the go and no-go stimuli. In addition to the expectation that d' would increase across trials, we also expected that increasing task difficulty would decrease d' . Difficulty was manipulated in two ways. In Experiment 1, participants were assigned to different inter-trial intervals, and we expected that a more rapid presentation of the go/no-go stimuli (i.e., shorter ITIs) would decrease d' . In Experiment 2, participants experienced different numbers of types of go stimuli for their assigned go/no-go ratio. For example, a participant may experience a 1:1 ratio of go to no-go stimuli but this would represent 50% go versus 50% no-go for a participant assigned to the 1v1 condition or 25% go-1 and 25% go-2 versus 50% no-go for a participant assigned to the 2v1 condition. Again, we predicted that as the task grew more difficult (number of types of go stimuli increased), the d' would decrease. It was not clear, however, whether changing the go/no-go ratio would have any effect on d' . It is possible that when no-go trials are rare that the resulting inattention to stimulus characteristics may decrease sensitivity.

Of greater interest to the production of false alarms is the response bias, which represents the tendency to choose one

response over the other. For example, some participants or conditions might show a strong tendency to respond, thus producing more hits, but consequently more false alarms as well, whereas other participants or conditions might produce more reluctance to respond, thus producing fewer hits but also fewer false alarms. We used SDT's c as a measure of response bias: $-0.5 \times [z(\text{hit rate}) + z(\text{false alarm rate})]$. This equation produces a c of zero if the false alarm rate is calculated as 1 minus the hit rate (e.g., a 75% hit rate accompanied by a 25% false alarm rate), thus indicating no bias toward one or the other type of error, because the false alarm rate would be equal to the miss rate. Values of c below zero indicate a bias to respond, and thus a higher relative false alarm rate, whereas values above zero indicate a greater reluctance to respond, and thus a higher relative miss rate. We expected that a greater ratio of go to no-go trials would produce a greater tendency to respond, and thus a negative c . This bias should grow stronger across trials, because participants would not initially have experienced enough trials to discern the go/no-go ratio. We did not expect that changing the difficulty would affect the response bias.

Regardless of any effects of go/no-go ratio and difficulty on d' and c , we were primarily interested in their effects on the overall number of false alarms, not just on the rate. Although increasing the go/no-go ratio was predicted to increase the false alarm rate, the corresponding decrease in the number of no-go trials might offset this increase. For example, Garavan et al. (2002) used a 14.75:1 go/no-go ratio and observed a false alarm rate of 45%, a remarkably high rate. Out of 1,000 trials, however, this rate would only produce about 29 false alarms, because there would only be 63 no-go trials. In contrast, the 4.55:1 ratio used by Mostofsky et al. (2003) produced a lower false alarm rate of 35%. However, this lower false alarm rate would produce a much larger total of 62 false alarms, because there would be 180 no-go trials out of 1,000. Although these studies suggest that there may be a desired go/no-go ratio to maximize false alarms, drawing a firm conclusion is confounded by the studies' use of different stimuli and ITIs, as well as some of their other characteristics. The optimal go/no-go ratio to maximize the number of false alarms could vary as a function of task difficulty. Our goal was to make direct comparisons of the effects of go/no-go ratio and task difficulty on d' , c , and the number of false alarms across a wide range of ratios and difficulties.

Experiment 1

Method

In Experiment 1 we examined the effect of the go/no-go ratio and the ITI on bias, sensitivity, and the number of false alarms. The go and no-go stimuli were identical except for color, with

the go stimuli in white and the no-go stimuli in red. One stimulus served as the go, and another served as the no-go, stimulus. Participants were instructed simply to press the space bar when the go stimulus appeared and to refrain from doing so when the no-go stimulus appeared.

Participants

The participants were 209 undergraduate students at a large American university. They received credit in their introductory psychology course for their participation.

Materials and design

Participants completed a go/no-go task programmed using PsyScope version XB57 (Cohen, MacWhinney, Flatt, & Provost, 1993). To assess the relationships between the independent variables and the dependent variables of interest, both ITIs and go/no-go ratios were randomly sampled from a uniform distribution (Young, Cole, & Sutherland, 2012). For each participant, we assigned one ITI and one go/no-go ratio; the ITI values ranged from 450 to 1,250 ms, and the programmed go/no-go ratios ranged from 1:1 to 10:1.

The word “PRESS” appeared in white for go trials and in red for no-go trials, always presented against a black background; thus, the only difference between go and no-go trials was the color of the text in which the stimulus was presented. The colors for our stimuli were selected to ensure that the stimuli were readily discriminable while minimizing potential effects of colorblindness. The go/no-go stimuli were always presented for 250 ms at the beginning of each trial. The total trial time ranged from 450 to 1,250 ms (250-ms stimulus presentation + an interstimulus interval ranging from 200 to 1,000 ms).

Procedure

To maximize sample size, participants completed the 5-min go/no-go task following a number of different studies being conducted in our lab. To ensure that there was no effect of the preceding study (e.g., differential effects of fatigue), the identity of the previous study was included as an independent variable in our analyses. This variable had no significant effect, and thus all results presented for this experiment will be collapsed across this factor.

Upon completion of the previous study, participants were given written instructions that outlined the task requirements. They were instructed to press the spacebar on their keyboard when the word “PRESS” was presented in white, and not to press the spacebar when the word “PRESS” was presented in red. Participants completed 150 trials. On any given trial the stimulus (go or no-go) was chosen randomly according to the assigned ratio and sampled with replacement to ensure that the

probability of each stimulus type remained consistent on each trial (e.g., .50 for the 1:1 ratio and .91 for the 10:1 ratio). Because the experienced ratio of go to no-go stimuli sometimes deviated from the programmed ratio for particular combinations, the ratio actually experienced by the participant was used in the data analysis.

The stimulus was presented in the center of the computer screen for 250 ms and was followed by a blank screen until the next trial was scheduled to begin. Responses were recorded during the entire trial (250-ms stimulus presentation + inter-stimulus interval). A failure to respond during the trial was counted as not responding. We did not eliminate trials with exceptionally fast reaction times, because these trials impact the d' score.

Results

Of the 209 participants, six were dropped from further analyses. Of those dropped, three responded much more often to the no-go than to the go stimulus (a negative d'), which indicates a failure to understand the instructions, whereas the other three showed no discriminative responding (d' values below 0.20).

Signal detection theory analysis

We used an SDT analysis of the effects of the experienced go/no-go ratio and ITI on d' and c to document the validity of our manipulations. This analysis was accomplished at the individual-trial level using a multilevel probit regression (DeCarlo, 1998). We included trial as a predictor and as a moderator to assess changes in d' and c as the experiment progressed. Both linear and logarithmic relationships between trial and behavior were examined, as well as both linear and logarithmic relationships between ratio and behavior (all of the log-transforms were natural logs). We considered both models that did not include trial as a predictor of sensitivity or bias and models that did not include sensitivity or trial as random slope effects. The best-fitting model of the ones tested (as assessed by AIC) was the model shown in Table 1 that included random effects of intercept across subjects and random subject slopes for the within-subjects variables type (go vs. no-go), which assessed sensitivity, and trial. The fixed-effects parameter estimates are represented visually in Fig. 1 as simple effects; to illustrate the effects of each variable (trial, ITI, and go/no-go ratio) on d' and c , each line represents the effect of the variable when the other variables are set to their average.

Changes in d' as a function of the predictors are represented at the top of Fig. 1. Not surprisingly, d' increased across trials, indicating that learning took place, and increased when the ITI was longer, suggesting that the slower pacing of trials produced greater sensitivity, and thus higher accuracy. The value of d' also decreased when the go/no-go ratio increased,

Table 1 Best-fitting parameter estimates from a multilevel probit regression of the trial-by-trial likelihood of responding on Experiment 1's go/no-go task (see DeCarlo, 1998; Wright & London, 2009)

Parameter	Estimate	SE	<i>z</i>	Prob > <i>z</i>
Intercept	0.65	0.03	20.36	<.001
Type	3.27	0.07	45.36	<.001
ITI	0.00042	0.00011	3.90	<.001
Log(Ratio)	0.343	0.039	8.83	<.001
Log(Trial)	0.179	0.025	7.06	<.001
Type × ITI	0.00161	0.00023	6.90	<.001
Type × Log(Ratio)	−0.590	0.084	−7.00	<.001
Type × Log(Trial)	0.221	0.034	6.43	<.001

Type is an effect-coded categorical variable: Go = +0.5, No-Go = −0.5. The predictors were mean-centered. The mean ITI was 833.7, the mean log(Ratio) was 1.44, and the mean log(Trial) was 4.07. The probit regression's intercept estimates negative *c*, so it must be negated to represent the *c* in signal detection theory.

suggesting that fewer exposures to the no-go trials negatively impacted sensitivity. However, in all cases the sensitivity was quite high, with *d'* typically averaging over 2.5.

Change in bias (as measured using *c*) as a function of the predictors is represented in the bottom part of Fig. 1. A bias of zero indicates that on average participants showed no propensity toward identifying the go versus the no-go stimulus. Negative bias values indicate that the criterion has shifted to the left in an SDT analysis, creating a bias toward responding.

Given that most of the participants experienced a go/no-go ratio with more go than no-go trials, we would expect an overall bias toward identifying a go stimulus, and thus a negative *c*. The overall estimated *c* was a −0.65, consistent with this expectation. The bottom of Fig. 1 reveals that participants generally began the study with no discernible bias, but their bias increased as the experiment progressed and the participants learned the go/no-go ratio. As expected, we observed a much smaller bias when the go/no-go ratio was closer to 1, and the bias increased (i.e., *c* became more negative) as the ratio approached 10. The greater bias for longer ITIs was unexpected, but this may reflect the faster learning that is possible with a slower trial pacing, and the effect is likely analogous to the increase in bias as a function of trial.

The signal detection analysis confirmed that the task produced the expected behavior. Decreased difficulty, as the result of either learning or a slower trial pacing, increased sensitivity. Increasing the base rate of go trials (i.e., increasing the go/no-go ratio) produced a much stronger bias toward responding, and this bias grew stronger across trials. We now turn to the central issue of the effects of ITI and go/no-go ratio on the number of false alarms produced.

Number of false alarms

In our second analysis, we used a Poisson regression predicting the total number of false alarms as a function of go/no-go ratio (log-transformed, which produced a better fit),

Fig. 1 Experiment 1: The simple effects of trial, intertrial interval (ITI), and go/no-go ratio on sensitivity, *d'*, and bias, *c*, when the other variables are set to their mean (for ITI) or geometric mean (for ratio and trial).

Graphs show the estimated means ± one standard error derived from the probit regression analysis. A strong negative bias indicates a stronger propensity to “go”

ITI, and their interaction, as well as a quadratic go/no-ratio component and its interaction with ITI. We predicted that shorter ITIs would produce more false alarms, due to the poorer sensitivity producing more overall errors (see top of Fig. 1). We also predicted that a value of go/no-go ratio in the middle of the tested range would maximize the number of false alarms, because of the necessity of balancing the greater false alarm rate produced by an increased go/no-go ratio (see bottom of Fig. 1) against the resulting fewer opportunities to produce a false alarm due to the smaller number of no-go trials.

The best-fitting parameter values are shown in Table 2, and the relationships are depicted in Fig. 2. Consistent with our hypotheses, the number of false alarms increased when the ITI decreased, and an intermediate go/no-go ratio produced the greatest number of false alarms for all but the highest ITIs, which produced very few false alarms. By solving for the values that maximized the dependent variable, we discovered that the highest number of false alarms is predicted to occur for the shortest tested ITI (450 ms) and a go/no-go ratio near 4:1. These values are predicted to produce a mean of 6.4 to 8.7 false alarms per 150 trials (95% confidence interval of the mean). However, go/no-go ratios between 3 and 6 produce nearly the same numbers of false alarms for all but the longest ITIs.

Reaction times

Although the issue is not directly relevant to our research question, some readers might be interested in the impact of our variables on reaction times (RTs). Thus, we also performed an examination of these data using a multilevel gamma regression of RTs using a log link function; because this analysis was ancillary, we will forgo the typical complete presentation of statistical details. Not surprisingly, the best-fitting model documented shorter RTs for false alarms than for hits ($M = 287$ vs. 325 ms), consistent with other published studies, and shorter RTs when the ITI was 450 ms ($M = 272$ ms) than when the ITI was 1,250 ms ($M = 354$ ms). Finally, people

reacted more slowly when the go/no-go ratio was 1:1 ($M = 344$ ms) than when it was 10:1 ($M = 289$ ms). All reported means are marginal means in which the other variables were fixed at their average.

Discussion

Our results suggest that a go/no-go ratio between 3:1 and 6:1 produces the greatest number of false alarms at nearly all ITIs for these stimuli. Although shorter ITIs push the number of false alarms even higher, the acceptable ITI range might be restricted due to other task requirements. For example, an EEG or fMRI study might require a slower trial pacing in order to avoid overlap of stimulus- or response-initiated brain activity across trials, or short ITIs might create undue stress or chance-level responding in an elderly sample. Thus, in our second experiment we examined a second variable, in order to vary the difficulty of the go/no-go task by changing the number of types of go stimuli that must be discriminated from the no-go stimulus. Furthermore, along with the increase in the number of types of go stimuli, the new stimuli were more similar to the no-go stimulus.

Experiment 2

Experiment 2 examined the effect of the go/no-go ratio and the number of types of go stimuli on bias, sensitivity, and the number of false alarms. The go and no-go stimuli were individual capital letters, potentially making the discrimination more difficult. The no-go stimulus was always a single letter, whereas the number of go stimuli varied between subjects from one to two, three, or four letters. For the one-letter condition, the stimuli were also different colors in order to make the discrimination easier, to ensure a sufficient spread of difficulty across the four conditions. For the three- and four-letter conditions, the additional (third and fourth) letters were also more similar to the no-go letter, thus making discriminating between the go and no-go stimuli increasingly difficult.

Method

The basic structure of the task was identical to that used in Experiment 1, with the exception that the go and no-go stimuli were varied across conditions and the ITI was a constant 750 ms (250-ms stimulus presentation + an interstimulus interval of 500 ms). The letter “M” always served as the no-go stimulus. For participants in the 1v1 (easy) condition, the go stimulus was a white “T” and the no-go stimulus was a red “M.” For participants in the 2v1, 3v1 and 4v1 conditions, all stimuli were white, and the go stimulus was sampled equally (without replacement) from the following letters: respectively, “TH,” “THW,” or “THWN.” Thus, a participant in

Table 2 Best-fitting parameter estimates from a Poisson regression of the number of false alarms in Experiment 1

Term	Estimate	SE	z	Prob > z
Intercept	1.72	0.04	39.43	<.001
ITI	−0.00076	0.00018	−4.26	<.001
Log(Ratio)	0.012	0.049	0.24	.806
ITI × Log(Ratio)	0.00013	0.00019	0.72	.471
Log(Ratio) ²	−0.22	0.05	−4.19	<.001
ITI × Log(Ratio) ²	0.000040	0.00019	2.07	.038

The predictors were mean-centered. The mean ITI was 833.7 and the mean log(Ratio) was 1.44.

Fig. 2 Experiment 1: The simple effects of ITI and go/no-go ratio on the number of false alarms. The plots show the estimated numbers of false alarms as a function of go/no-go ratio (± 1 standard error) when ITI was at its minimum, average, or maximum

the 2v1, 1:1 ratio condition would receive 50% go stimuli (25% “T”s and 25% “H”s) and 50% no-go stimuli (“M”s). A participant in the 3v1, 4:1 ratio condition would receive 80% go stimuli (26.7% “T”s, 26.7% “H”s, and 26.7% “W”s) and 20% no-go stimuli (“M”s). Because the experienced ratio of go to no-go stimuli sometimes deviated from the programmed ratio for particular combinations, the ratio actually experienced by the participant was used in the data analysis.

Participants

The participants were 175 undergraduate students at a large American university. They received credit in their introductory psychology course for their participation.

Results and discussion

Of the 175 participants, seven were dropped from further analyses, all because of d' values below 0.20, indicating no significant differentiation between the go and no-go stimuli. The final distribution of participants to the four conditions was 1v1 ($N = 45$ after one dropped), 2v1 ($N = 43$ after three dropped), 3v1 ($N = 43$ after one dropped), and 4v1 ($N = 39$ after two dropped). Data collection stopped at the end of the semester.

Signal detection theory analysis

The same analytical approach was used here as in Experiment 1. The best-fitting model of the ones tested (as assessed by AICs) was the model shown in Table 3 and included the same random effects used in Experiment 1. The fixed-effects parameter estimates are represented visually in Fig. 3 as simple effects; to illustrate the effects of each variable (trial, number of types of go stimuli, and go/no-go ratio) on d' and c , each line represents the effect of the variable when the other variables are set to their average (the number of types of go stimuli

was treated as an effect-coded categorical variable, so its average occurs when these variables are all zero).

Changes in d' as a function of the predictors are represented at the top of Fig. 3. Not surprisingly, d' increased across trials, indicating that learning took place. Also as expected, d' decreased as the number of types of go stimuli increased, although only in the 1v1 condition was d' appreciably different. The sensitivity in this condition ($M = 2.7$) approached that observed in Experiment 1 ($M = 3.2$), which had also used only one type of go stimulus, but the sensitivities for the other conditions were much lower than those in our first experiment ($M = 1.6$), providing strong evidence that these conditions were more difficult. The value of d' also decreased when the go/no-go ratio increased, thus replicating the finding of Experiment 1.

Table 3 Best-fitting parameter estimates from a multilevel probit regression of the trial-by-trial likelihood of responding on Experiment 2’s go/no-go task (see DeCarlo, 1998; Wright & London, 2009)

Parameter	Estimate	SE	z	Prob > $ z $
Intercept	1.03	0.03	33.50	<.001
Type	1.85	0.07	26.83	<.001
Condition[1v1]	−0.19	0.10	−1.90	.057
Condition[2v1]	0.11	0.10	1.10	.271
Condition[3v1]	−0.00	0.10	−0.03	.972
Log(Ratio)	0.410	0.043	9.44	<.001
Log(Trial)	0.051	0.018	2.88	.004
Type $\times$ Condition[1v1]	1.76	0.23	7.53	<.001
Type $\times$ Condition[2v1]	−0.57	0.23	−2.49	.012
Type $\times$ Condition[3v1]	−0.70	0.23	−3.06	.002
Type $\times$ Log(Ratio)	−0.346	0.099	−3.49	<.001
Type $\times$ Log(Trial)	0.127	0.030	4.23	<.001

Type is an effect-coded categorical variable: Go = +0.5, No-Go = −0.5. Condition is an effect-coded categorical variable across three dummy variables: 1v1 = {0.5, 0, 0}, and 2v1 = {0, 0.5, 0}, 3v1 = {0, 0, 0.5}, 4v1 = {−0.5, −0.5, −0.5}. The continuous predictors were mean-centered. The mean log(Ratio) was 1.40, and the mean log(Trial) was 4.07. The p values are derived using Wald tests.

Fig. 3 Experiment 2: The simple effects of trial, number of types of go stimuli, and go/no-go ratio on sensitivity, d' , and bias, c , when the other variables are set to their geometric mean (ratio and trial). Graphs show the

estimated means $\pm$ one standard error derived from the probit regression analysis. A strong negative bias indicates a stronger propensity to “go”

Change in bias (as measured using c) as a function of the predictors is represented in the bottom part of Fig. 3. Given that most of the participants experienced a go/no-go ratio with more go than no-go trials, we would expect an overall bias toward identifying a go stimulus, and thus a negative c . The overall estimated c was a -1.00 , consistent with this expectation, and stronger than the value observed in Experiment 1. The bottom of Fig. 3 reveals that participants generally began the study with a fairly strong bias, which increased slightly as the experiment progressed. As expected, the bias was much smaller when the go/no-go ratio was closer to 1, and it increased as the ratio approached 10. The bias in each of the four conditions that varied in their numbers of types of go stimuli were very similar but slightly lower in the 1v1 condition.

The signal detection analysis confirmed that the task produced the expected behavior. Decreased difficulty, as the result of either learning or the use of only one type of go stimulus, increased sensitivity. Increasing the go/no-go ratio produced a much stronger bias toward responding. We now turn to the central issue of the effects of the number of types of go stimuli and the go/no-go ratio on the number of false alarms produced.

Number of false alarms

In our second analysis, we used a Poisson regression predicting the total number of false alarms as a function of go/no-go ratio, condition, and their interaction, as well as a quadratic go/no-ratio component and its interaction with

condition. We predicted that the more difficult 2v1, 3v1, and 4v1 conditions would produce more false alarms, due to the poorer sensitivity in these conditions producing more overall errors (see the top of Fig. 3). We also predicted that a go/no-go ratio between 3:1 and 6:1 would produce the largest number of false alarms, consistent with Experiment 1.

The best-fitting parameter values are shown in Table 4, and the relationships are depicted in Fig. 4. Consistent with our hypotheses, the number of false alarms increased when there were more types of go stimuli, and an intermediate go/no-go ratio produced the greatest number of false alarms. Overall, we observed many more false alarms in Experiment 2 than in Experiment 1, and the use of multiple types of go stimuli resulted in a much greater number of false alarms than when there was only one. By solving for the values that maximized the dependent variable, we discovered that the highest number of false alarms was predicted to occur for the 4v1 condition and a go/no-go ratio of 2.5. These values are predicted to produce a mean of 19.6 to 25.9 false alarms per 150 trials (95% confidence interval of the mean). However, any go/no-go ratio between 1.5 and 3.5 is predicted to produce an average of 20 or more false alarms for the 2v1, 3v1, and 4v1 conditions.

Reaction times

To parallel the analysis of Experiment 1, we again performed an examination of RTs using a multilevel gamma regression

Table 4 Best-fitting parameter estimates from a Poisson regression of the number of false alarms in Experiment 2

Term	Estimate	SE	z	Prob > z
Intercept	2.87	0.03	106.90	<.001
Condition[1v1]	−0.40	0.05	−8.04	<.001
Condition[2v1]	0.10	0.04	2.31	.021
Condition[3v1]	0.08	0.04	1.90	.057
Log(Ratio)	−0.51	0.04	−11.88	<.001
Condition[1v1] × Log(Ratio)	0.15	0.09	1.68	.093
Condition[2v1] × Log(Ratio)	−0.09	0.07	−1.31	.190
Condition[3v1] × Log(Ratio)	−0.00	0.07	−0.08	.940
Log(Ratio) ²	−0.48	0.05	−9.94	<.001
Condition[1v1] × Log(Ratio) ²	−0.14	0.10	−1.46	.145
Condition[2v1] × Log(Ratio) ²	0.14	0.07	1.91	.056
Condition[3v1] × Log(Ratio) ²	0.12	0.08	1.49	.136

The log(Ratio) predictor was mean-centered. The mean log(Ratio) was 1.40. The *p* values are derived using Wald tests.

and a log link function. Again, the best-fitting model documented shorter RTs for false alarms than for hits ($M = 303$ vs. 331 ms), as well as shorter RTs in the 1v1 than in the 2v1, 3v1, and 4v1 conditions (M s = 295 , 322 , 332 , and 319 ms, respectively). Finally, people reacted more slowly when the go/no-go ratio was 1:1 ($M = 344$ ms) than when it was 10:1 ($M = 302$ ms). All reported means are marginal means in which the other variables were fixed at their average.

General discussion

The go/no-go task has been in steady use for decades because of its utility as both a platform for studying the basic mechanisms of discrimination, learning, and impulsivity and an assessment tool for individual differences. A greater understanding of the effects of such key manipulated variables as the ITI, go/no-go ratio, and number of types of go stimuli will improve its utility. Although our study focused on creating the best conditions to maximize the number of false alarms by manipulating three target variables, other factors would also affect

performance. In Experiment 1, we used two very discriminable stimuli (white vs. red) presented for 250 ms, but the use of less discriminable stimuli (Kok, 1986; Nieuwenhuis, Yeung, & Cohen, 2004), shorter presentation durations (Eimer, 1993; Garavan et al., 2002; Menon et al., 2001), and greater variability in the go or no-go stimuli (Langenecker, Zubieta, Young, Akil, & Nielson, 2007; Simmonds, Pekar, & Mostofsky, 2008) would decrease d' , and thus increase the numbers of both false alarms and misses. Making the task more difficult does raise the possibility that individual differences could emerge that are not due to basic cognitive differences in impulsivity, but rather may reflect differences in perceptual, attentional, or working memory ability (Criaud & Boulinguez, 2013). Perceptual deficits become of greater importance when the participants of interest are elderly, very young, or belong to populations with specific deficits (Casey et al., 1997; Langenecker & Nielson, 2003). In addition, increasing a task's difficulty may result in a weaker bias, because it may take participants longer to learn the go/no-go ratio (e.g., the effect of ITI documented at the bottom of Fig. 1).

Our results suggest that a go/no-go ratio of 3:1 produces a large number of false alarms, regardless of the ITI or the number of types of go stimuli. Although shorter ITIs push the number of false alarms even higher, the acceptable ITI range might be restricted by other task requirements. For example, an EEG or fMRI study might require a slower trial pacing in order to avoid an overlap of stimulus- and response-initiated brain activity across trials, or short ITIs might create undue stress or chance-level responding in an elderly sample. Thus, in our second experiment we demonstrated that increasing the number of types of go stimuli produced a sharp increase in the number of false alarms while also slightly reducing the optimal go/no-go ratio from 4.0 to 2.5. However, we must caution that the increased difficulty of discriminating two or more types of go stimuli from a no-go stimulus may tax working memory. Thus, the use of only two types of go stimuli may be preferred, in light of the small differences among the 2v1, 3v1, and 4v1 conditions.

We focused on identifying the conditions that maximize the number of false alarms because of their central interest to EEG

Fig. 4 Experiment 2: The simple effect of the go/no-go ratio on the numbers of false alarms (± 1 standard error) at each level for the number of types of go stimuli

research on the error-related negativity, in which brain activity accompanying a hit is compared to activity accompanying a false alarm (Nieuwenhuis et al., 2004; Scheffers, Coles, Bernstein, Gehring, & Donchin, 1996). When a task produces a small number of false alarms, the small sample size undermines statistical robustness. We should note that the design of a go/no-go study is much less important when the focus is on hits and correct rejections, because these response classes occur at a very high rate across a wide range of designs. Although misses can also be rare and may be of interest to researchers, the conditions that create more misses are clear—increased difficulty creates more errors, and a high go/no-go ratio creates more opportunities for misses, because of the increase in the number of go trials (although it is possible that misses might be even more likely to occur if the number of no-go trials outnumbered the number of go trials). Only for false alarms does the bias created by a higher go/no-go ratio (which increases the rate of false alarms) create the unfortunate side effect of fewer no-go trials on which false alarms can occur.

Although this was not a central goal of our study, our documentation of the changes in d' and c across the experiment should encourage researchers that use the task to assess behavioral inhibition to consider at which points during the experiment behavior is assessed. There was considerable behavioral change during the first 15 trials of our study, with substantive increases in both sensitivity and bias. Because many researchers approach EEG and fMRI data by aggregating across the entire experiment, there is the risk of including trials in which a response bias has not yet emerged. The same caution is relevant when go/no-go tasks are used for clinical assessment. If a response bias emerges over time, it might behoove assessment professionals either to consider eliminating the early portion of an assessment or to determine the rate at which performance changes.

In sum, the choice of the go/no-go task to achieve research or assessment goals in order to study behavioral inhibition merely represents the first step. Careful consideration of the specific characteristics of that task will determine the task's utility. Larger samples (of, e.g., false alarms) and more stable measures (which would be achieved after behavior approaches asymptote) create a more precise assessment that will provide a stronger basis for subsequent analyses. Alternatively, the rate of behavioral change may be a significant marker with predictive and theoretical significance. Excellent design and analysis always improve quality.

References

- Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., Giedd, J. N., ... Rapoport, J. L. (1997). A developmental functional MRI study of prefrontal activation during performance of a go-no-go task. *Journal of Cognitive Neuroscience*, 9, 835–847. doi:10.1162/jocn.1997.9.6.835
- Chikazoe, J., Jimura, K., Asari, T., Yamashita, K., Morimoto, H., Hirose, S., ... Konishi, S. (2009). Functional dissociation in right inferior frontal cortex during performance of go/no-go task. *Cerebral Cortex*, 19, 146–152. doi:10.1093/cercor/bhn065
- Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25, 257–271. doi:10.3758/BF03204507
- Criaud, M., & Boulinguez, P. (2013). Have we been asking the right questions when assessing response inhibition in go/no-go tasks with fMRI? A meta-analysis and critical review. *Neuroscience and Biobehavioral Reviews*, 37, 11–23. doi:10.1016/j.neubiorev.2012.11.003
- DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, 3, 186–205. doi:10.1037/1082-989X.3.2.186
- Drewe, E. A. (1975). Go-no go learning after frontal lobe lesions in humans. *Cortex*, 11, 8–16.
- Eimer, M. (1993). Effects of attention and stimulus probability on ERPs in a go/nogo task. *Biological Psychology*, 35, 123–138.
- Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: Inhibition, error detection, and correction. *NeuroImage*, 17, 1820–1829.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- Kiefer, M., Marzinzik, F., Weisbrod, M., Scherg, M., & Spitzer, M. (1998). The time course of brain activations during response inhibition: Evidence from event-related potentials in a go/no go task. *NeuroReport*, 9, 765–770.
- Kok, A. (1986). Effects of degradation of visual stimulation on components of the event-related potential (ERP) in go/nogo reaction tasks. *Biological Psychology*, 23, 21–38.
- Langenecker, S. A., & Nielson, K. A. (2003). Frontal recruitment during response inhibition in older adults replicated with fMRI. *NeuroImage*, 20, 1384–1392. doi:10.1016/S1053-8119(03)00372-0
- Langenecker, S. A., Zubieta, J., Young, E. A., Akil, H., & Nielson, K. A. (2007). A task to manipulate attentional load, set-shifting, and inhibitory control: Convergent validity and test-retest reliability of the parametric go/no-go test. *Journal of Clinical and Experimental Neuropsychology*, 29, 842–853.
- Luria, A. R. (1973). The frontal lobes and the regulation of behavior. In K. Pribram & A. R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 3–26). New York, NY: Academic Books.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Mahwah, NJ: Erlbaum.
- Menon, V., Adelman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a go/nogo response inhibition task. *Human Brain Mapping*, 12, 131–143.
- Mostofsky, S. H., Schafer, J. G. B., Abrams, M. T., Goldberg, M. C., Flower, A. A., Boyce, A., ... Pekar, J. J. (2003). fMRI evidence that the neural basis of response inhibition is task-dependent. *Cognitive Brain Research*, 17, 419–430.
- Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 17–26. doi:10.3758/CABN.3.1.17
- Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2004). Stimulus modality, perceptual overlap, and the go/no-go N2. *Psychophysiology*, 41, 157–160. doi:10.1046/j.1469-8986.2003.00128.x
- Scheffers, M. K., Coles, M. G. H., Bernstein, P., Gehring, W. J., & Donchin, E. (1996). Event-related brain potentials and error-

- related processing: An analysis of incorrect responses to go and no-go stimuli. *Psychophysiology*, 33, 42–53.
- Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of go/no-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia*, 46, 224–232. doi:[10.1016/j.neuropsychologia.2007.07.015](https://doi.org/10.1016/j.neuropsychologia.2007.07.015)
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1–26.
- Wright, D. B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology*, 62, 439–456. doi:[10.1348/000711008X327632](https://doi.org/10.1348/000711008X327632)
- Young, M. E., Cole, J. J., & Sutherland, S. C. (2012). Rich stimulus sampling for between-subjects designs improves model selection. *Behavior Research Methods*, 44, 176–188. doi:[10.3758/s13428-011-0133-5](https://doi.org/10.3758/s13428-011-0133-5)