1   QpFMRI: Bayesian adaptive stimulus presentation for real-time fMRI

2   Steven M. Weisberg[1] & Geoffrey K. Aguirre[2]

3   [1] University of Florida
4   [2] University of Pennsylvania

5   Author Note

6   Steven M. Weisberg, University of Florida, Department of Psychology Geoffrey K.

7   Aguirre, University of Pennsylvania, Department of Neurology, Center for Cognitive

8   Neuroscience. Pre-print submitted for peer review.

9   Correspondence concerning this article should be addressed to Steven M. Weisberg,

10 945 Center Dr., Gainesville, FL 32611. E-mail: stevenweisberg@ufl.edu

11   Abstract

12   QUEST+ (Q+, Watson, 2017) is an adaptive stimulus selection approach that works to

13   minimize uncertainty in a parameterized model of the responses. Unlike random stimulus

14   selection, Q+ stimulus selection dynamically select the next stimulus based upon the prior

15   responses of the subject, and an underlying stimulus-response model. While this approach

16   has been successful in behavioral experiments, to use Q+ in BOLD fMRI experiments,

17   experimenters require a 'real-time' measure of the BOLD response to stimuli as they are

18   presented and solutions that address idiosyncracies of the BOLD signal. We have created a

19   software toolbox that 1) extracts (or simulates) the BOLD fMRI signal from a brain region

20   as imaging volumes are acquired, 2) cleans and fits the growing time-series to estimate the

21   response on each trial, and 3) applies Q+ to the responsees to select the next stimulus. In

22   simulations that model empirical data for a simple visual contrast experiment featuring

23   stimuli of varying visual contrast values, we found that Q+ stimulus selection recovers

24   model parameters more accurately than random stimulus selection. We report the results

25   of stimulus selection approaches on simulated data, as well as how to optimize

26   experimental parameters (e.g., TR, trial length) to improve model fit. In addition, we

27   describe our solutions to the technical challenges posed by adapting Q+ for fMRI data,

28   describe the codebase we have written to implement the solution, and present sample uses

29   for Q+ stimulus selection in future real-time BOLD fMRI experiments.

30   *Keywords:* adaptive stimulation, fMRI, Quest+, real-time fMRI, vision neuroscience

31   Word count: X

32   QpFMRI: Bayesian adaptive stimulus presentation for real-time fMRI

33   Introductory text placeholder.

34 **General background of the problem**

35 **Specific problem to be solved**

36 **Here we . . .**

37   **Method**

38 **Approach**

39   Does the qpfMRI approach yield improved fit over random stimulus selection? As a

40 proof of principle, we created a forward of model primary visual cortical function using a

41 logistic model (Equation 1). We selected this model because it is relatively simple (only

42 two parameters vary between 0 and 1) and because logistic functions would be a plausible

43 model for visual cortex (e.g., in response to light flickering at varying frequencies) and thus

44 would make a good neurometric function. That is, a logistic function can model the

45 increasing neural signal in response to an increase in stimulus flicker. Let *slope* equal the

46 slope of the curve at the semi-saturation point. Let *semiSat* equal the point on the x-axis

47 at which the y-axis reaches 50%.Let x_0 be the value of a stimulus. Then, the logistic

48 function is defined as:

$$y_0 = 1 - \frac{1}{\left(e^{(slope*(x_0-semiSat))}\right)} \tag{1}$$

49 We developed a closed-loop pipeline, starting from known model parameters and resulting

50 in an estimate of those same parameters from simulated BOLD fMRI data. The broad

51 steps were as follows. 1) Select a set of model parameters for the logistic function. 2)

52 Simulate BOLD fMRI data in response to stimuli based on the known model parameters.

53 3) Use a forward model to generate simulated BOLD fMRI data. 4) Use a reverse model to

54 obtain estimates of the BOLD response for each trial. 4) Select the next stimulus based on

55 either random stimulus selection (control) or applying Q+ to the responses. In developing

56 simulations to achieve this goal, we solved several problems in applying Q+ to BOLD fMRI

57 data. To wit: BOLD fMRI data are (1) continuous, (2) noisy, (3) unitless, (4) have a

58 maximum response that is unknown *a priori* and (5) are subject to low-frequency temporal

59 variation. We address each of these in turn, but first provide a general overview of Q+.

60 **Quest +**

61   Q+ (David H. Brainard, 2017; Watson, 2017) is a multidimensional Bayesian method

62 for adaptive stimulus presentation. Q+ operates by computing the conditional probability

63 of a behavioral outcome for each set of stimulus parameters. Q+ is a generalization on

64 QUEST (Watson & Pelli, 1983), which estimates a psychometric function on the basis of a

65 single stimulus dimension and provides an adaptive testing procedure to fit the function.

66 Whereas QUEST could only compute one stimulus dimension, one psychometric function

67 parameter, and two outcome measures, Q+ allows for multiple stimulus dimensions,

68 psychometric parameters, and many outcome measures. Here, we further expand Q+ to

69 model the BOLD fMRI signal based on a neurometric function, which fits stimulus

70 dimensions to changes in the BOLD response. To help explain, we contrast a BOLD fMRI

71 experiment with a color categorization behavioral experiment. In the color category

72 experiment, a participant is presented with one patch of color of varying wavelengths at a

73 time and must select which color category (e.g., red, blue, green, or yellow), the color

74 belongs to.

75   **Challenge 1: The BOLD signal is continuous.**   Q+ operates over discrete

76 categorical outcomes, such as selecting amongst a set of color names for a stimulus. For

77 stimuli in the middle of the red spectrum, roughly 100% of responses will be "red." But as

78 the color patch drifts closer to yellow, this percentage will decrease. But the "outcomes"

79 for the BOLD fMRI signal are continuous. There is no principled way to divide the BOLD

signal up into categories, and arbitrarily small categories will increase computation time dramatically. To solve this, we define a fixed number of outcome categories onto which we map the amplitude of the BOLD response.

Challenge 2: The BOLD signal is noisy and outcomes are sparse. Q+ esimtates parameters of a function that predict the proportion of trial outcomes that fall in each category. For example, from wavelength (the stimulus dimension) Q+ will generate a prediction for the percentage of trials that will be described as red, green, blue, or yellow. Although there is uncertainty in the response over trials (i.e., presented with the same wavelength patch, participants will sometimes call the color red and other times call the color green), each individual response is assumed to be certain. That is, we do not assume that a participant who has said "green" means to say "red" 10% of the time. The BOLD signal, on the other hand, is noisy. In response to a given stimulus, the BOLD signal will assume a range of values, normally distributed around the population mean (given some assumptions.) That is, the BOLD signal is a response that itself has uncertainty built in. In other words, physiologic noise in the BOLD fMRI response causes outcomes to be spread across categories around the veridical response value. Given unlimited amounts of data, this would not be a problem because the mean value of the BOLD response for a given stimulus would converge on the true value. But, unlike the color study, which could have 100s of trials, BOLD fMRI trials typically number in the 10s, thus responses are potentially sparse across the many outcome categories. To solve this, the outcome for each stimulus is determined by the parameters oof the neurometric function, plus a parameter we call σ (*sigma*) that controls the width of a Gaussian that distributes outcomes across adjacent categories. Q+ estimates the value of σ as well.

Challenge 3: The BOLD fMRI signal is unitless. Q+ operates on outcomes with defined boundaries. The BOLD fMRI signal is unitless and thus interpreted relative to a baseline stimulus. To solve this, the outcome amplitude is calculated relative to a specified baseline stimulus, which is required to be presented at the beginning of the

experiment.

Challenge 4: The maximum response is not known *a priori*.   Similarly, the

maximum possible response is not known *a priori* and must be calculated from the

neurometric curve. Again, Q+ operates on outcomes with defined boundaries, so we must

impose a possible range on the BOLD signal from baseline to maximum. To solve this, the

outcome algorithm adjusts the mapping of BOLD fMRI response to the outcome

categories, updating an estimate of maximum BOLD response.

Challenge 5: BOLD fMRI data are subject to low-frequency temporal

noise.   Q+ accumulates evidence across trials. To obtain the amplitude of response on a

new trial, the growing BOLD fMRI signal is subject to pre-processing, including the

removal of low-frequency noise. This calculation alters the estimated response to prior

trials. To solve this, we iteratively re-train Q+ with all past trials up through the current

trial.

QPFMRI: An overview

The QPfMRI pipeline is detailed in Figure 1. We describe the pipeline with

simulated BOLD fMRI data where the model parameters can be specified in advance, but

the principles for implementing Q+ with real BOLD fMRI data will be similar. To

summarize, we specify a set of model parameters (Fig. 1A) then select trials and calculate

the estimated BOLD fMRI response based on the model with added simulated physiologic

noise. This results in a BOLD fMRI timeseries (Fig. 1B), from which we calculate the

estimated response to each trial. We then pass the series of trials and outcomes to Q+

(Fig. 1C), which provides an adaptive suggestion for the next trial. Finally, at the end of
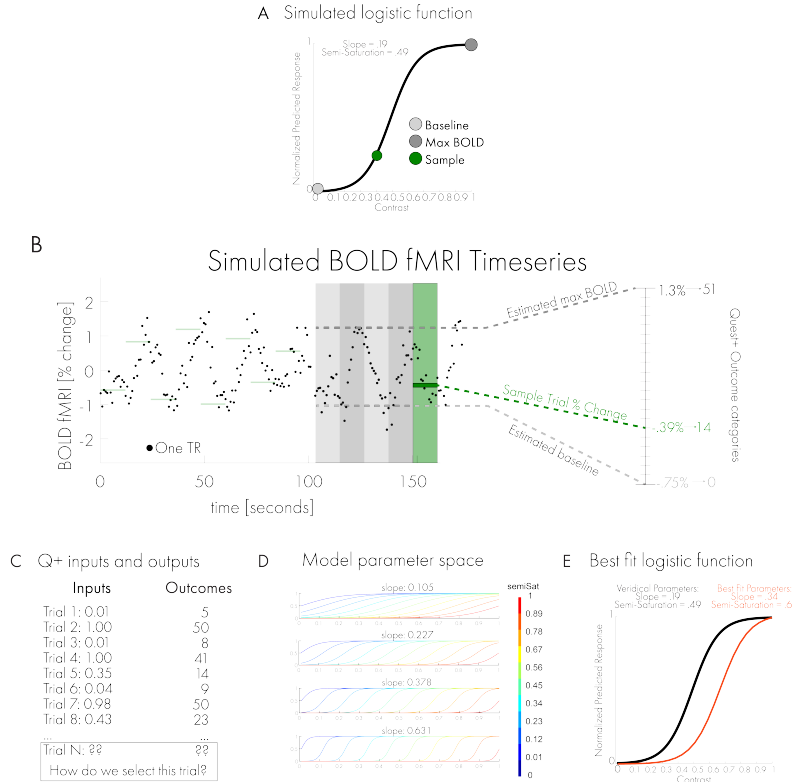
the simulation, we estimate model fit (Fig. 1E).

Figure 1. An overview of the processing pipeline in converting a BOLD signal to a Quest+ outcome. A set of model parameters is chosen (A), resulting in a predicted value for the BOLD signal, normalized between 0 and 1. Trials are then selected from this curve, which can be baseline trials (light grey dot), where the expected BOLD response is expected to be the minimum. From these trials, a BOLD fMRI timeseries is simulated using the temporal fitting engine's forward model (B), which convolves the predicted neural response from the logistic function with a canonical hemodynamic function. Gaussian noise and pink noise are then added to this signal to simulate the physiologic noise typical of fMRI. The timeseries is then fit with a general linear model, resulting in the estimated response for each trial (green bands). The baseline response (light grey vertical panels) is estimated from the average of baseline trials, and the maximum BOLD response is also estimated (dark grey vertical panels) either from maximum BOLD trials or from the estimated maximum BOLD response based on the logistic curve. Then the Q+ outcome categories are created and each non-baseline and maximum BOLD trial is assigned to the outcome category that contains its estimated value. The bin numbers and outcome categories are then passed to Q+ (C), which calculates the conditional probabilities of the parameter values (the parameter space (D)), and returns a suggestion for the next trial to present, with the aim of minimizing the entropy among the choices in the parameter space. Finally, we estimate the best fitting model to the data, resulting in an estimate of each model parameter (E).

Simulation method

Forward model. For the forward model, we selected sample parameters that we expected to be well behaved (e.g., yield values between 0 and 1 for most of their range). Our choice of logistic function parameters and the domain of possible parameter values to search through were fixed through all simulations except where otherwise noted. We explain the flow of data using random stimulus selection. First, a random set of stimulus values is chosen of length N (where N is the number of trials in the experiment). We also

137 enforce a set number of baseline trials (where the expected BOLD response will be

138 minimal) and maximum BOLD trials (where the expected BOLD signal will be maximal).

139 Next, each stimulus value is passed to the logistic function using our chosen function

140 parameters. The output of that function (which has been normalized to be between 0 for

141 baseline and 1 for maximum response) is then convolved with a canonical hemodynamic

142 response function and adding 1/f noise (pink noise, consistent with BOLD signal). We also

143 add white noise to the simulated timeseries (the amount of white noise varied across

144 simulations). The resulting continuous timeseries is then downsampled to match the

145 number of TRs per trial in the simulation and mean-centered. These steps were

146 implemented with the temporal fitting engine

147 (https://github.com/gkaguirrelab/temporalFittingEngine).

148 **Reverse model.**   In the real-time simulation, the continuous timeseries is fed into

149 the reverse model one trial at a time. Each pass through the reverse model, the signal is

150 mean-centered. The timeseries is also reference coded, meaning all baseline trials are

151 averaged and that average is subtracted from the rest of the timeseries so that the

152 minimum value (theoretically) is zero. Next, we apply a general linear model (GLM) in

153 which the baseline trials are the baseline and every other trial is represented as a boxcar (a

154 vector of length $N*t$ with zeros everywhere except the cells for that trial, which contain

155 ones). The GLM is fit to the timeseries and the betas returned are then assigned to

156 outcome categories. At the end of each trial, the logistic function parameter values are

157 estimated and compared to the veridical logistic function parameters. Timeseries cleaning

158 (baseline and max BOLD), assigning outcomes, high-pass and low-pass filtering, removing

159 linear drift, mean-centering.

160 **Q+ stimulus selection.**   In behavioral experiments, the combinations of

161 parameters are those that define the model. Due to the challenges listed above for fMRI,

162 two additional parameters are searched for by Q+, which we call *sigma* and *beta*. Sigma

163 can be thought of as the estimate of the noise in the BOLD fMRI signal, though it is

164   implemented as the spread along the possible outcome categories (e.g., a small value of

165   sigma means that one stimulus value will correspond to only a few possible outcome

166   categories). Beta is a scaling parameter that is used to help estimate what the maximum

167   BOLD response is and approaches 1 as the maximum BOLD response is better estimated.

168   For each stimulus - outcome category pair, Q+ updates the log-likelihood for all possible

169   combinations of parameters. Depending on the previous stimuli, Q+ will select the next

170   stimulus by determining which result will most reduce the remaining entropy in the set of

171   parameter combinations.

172   **Simulation parameters.**   In addition to the model parameters, which Q+

173   attempts to find, we also explore a range of simulation parameters. These include one

174   parameter related to Q+ search, the number of outcome categories (7 or 15). Other

175   parameters we explored relate to the structure of the simulation itself: simulated noise (the

176   amount of white noise added to the simulated BOLD fMRI signal), the length of individual

177   trials (8 or 12 seconds), the length of the TR (800 or 1600 ms).

178   We did not vary several other parameters, including the stimulus domain (which

179   stimulus values could be selected), the value for the maximum BOLD response, and the

180   proportion of outcome categories reserved for the top and bottom (which we call headroom

181   and set at 10%). We also did not vary the model parameter domains (i.e., the values that

182   each model parameter could take on). We arrived at these choices through trial and error,

183   which is a compelling reason itself to use the Q+ fMRI simulation toolbox to set

184   parameters for fMRI experiments. Some choices (like increasing the number of outcome

185   categories, or the size of the parameter and stimulus domains) drastically increase the time

186   to run Q+, limiting its utility in real-time fMRI experiments. Others, like the choice of

187   headroom, yield strange or nonsensical results if they are set too low (0%) or too high

188   (50%).

189   **Data and code availability.**   The data and code for the analyses described in the

190   results is available in a git repository

(https://github.com/smweis/Weisberg_and_Aguirre_2020). The QpfMRI toolbox is available for download and use (https://github.com/smweis/qpfmri).

Results

Our main approach is to compare simulated BOLD fMRI data timeseries that were generated using random stimulus selection with timeseries that were generated with Q+ optimization. We first step through one sample simulation, showing trial-by-trial results. We then describe simulations over a range of parameters comparing random stimulus selection with Q+ stimulus selection. The results demonstrate improved convergence on model parameters when Q+ controls stimulus selection. The results also suggest optimal parameters for real-time fMRI design using Q+ for TR and trial duration.

First, we describe one set of 100 6-minute simulated fMRI runs comparing Q+ with random stimulus selection for one set of model parameters and one set of simulation parameters. The first set of simulations illustrate the advantage of Q+ stimulus selection in the case of simulation parameters that optimize for realistic and desired fMRI conditions (fast TR, long trials, medium noise) and fast Q+ performance (few outcome categories).Next, we evaluate the range of performance for Q+ compared to random stimulus selection across a several control parameter choices. Finally, we explore performance within one set of control parameters across a range of model parameters.

For these next sets of simulations, we used the following control parameters: TR = 800ms; number of trials = 30; trial length = 12s. We set the estimated maximum BOLD to 1.5% signal change and initialized the maximum BOLD value to 1%. We enforced baseline and maximum BOLD trials to alternate for the first six trials and then mandated a baseline trial every subsequent 5th trial and a maximum BOLD trial every 10th trial. We chose the lowest level of noise (.05 SD).

Simulating one set of model parameters

To determine whether Q+ improves model fit over random stimulus selection, we selected model parameters and control parameters that we expected would behave well. We chose model parameters from the middle of the distribution (slope = .41, semi-saturation = .57), and which led to high differentiation among our stimulus domain. That is, variation in the range of possible stimulus values would lead to a unique model parameter solution, which would be discoverable by the Q+ algorithm. As stated above, we chose control parameters optimizing for desired fMRI conditions. The results of the simulations are summarized in Figure 2, Figure 3, and Figure 4. Qualitatively, Q+ stimulus selection led to a narrower distribution around the model parameters (Figure 2), with fewer outlying results. Q+ stimulus selection led to a higher overall fit as measured by R-squared (Figure 3). Over 50 simulations per type, the fit for Q+ stimulus selection was significantly higher, Wilcoxon signed-rank test $Z(98) = 5.13$, $p < 0.0001$ (Q+: $M = 0.9986$, $SD = 0.0012$; Random: $M = 0.9817$, $SD = 0.0348$).
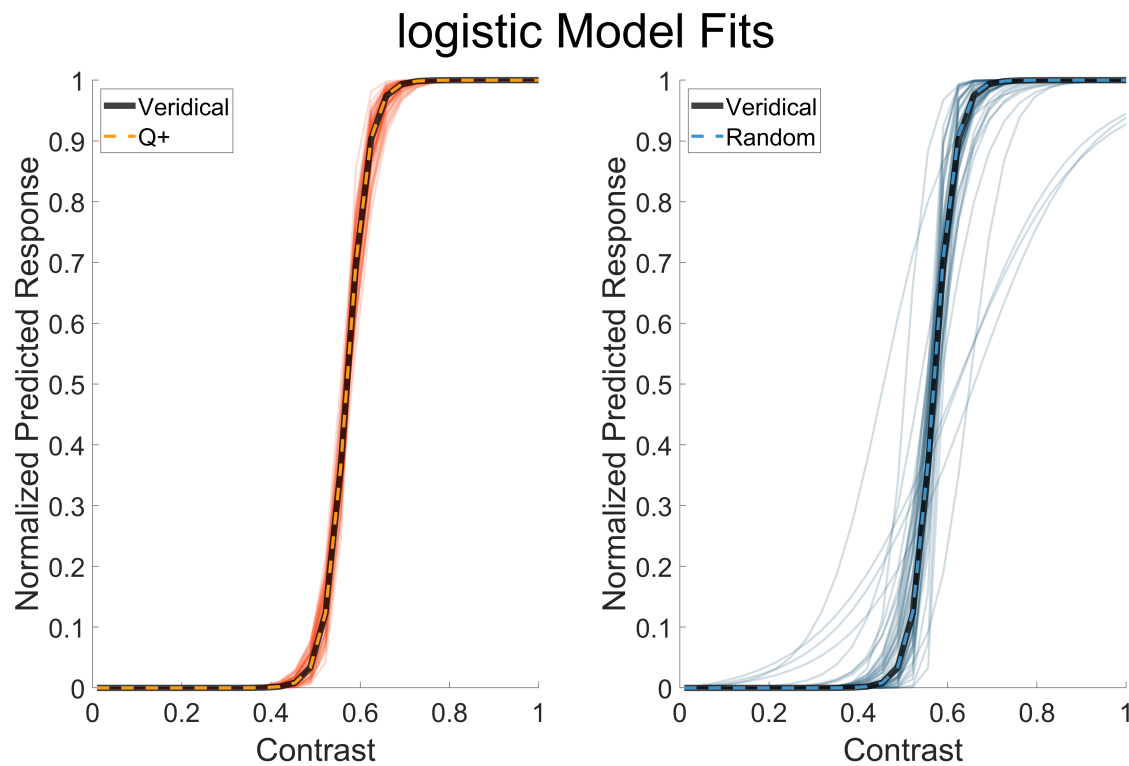
Figure 2. The first set of simulations we evaluated are plotted. Each black line represents the veridical model (slope = .41, semi-saturation = .57) normalized to be between 0 and 1. Thin orange and blue lines represent each simulation's final parameters and the thick dotted line represents the mean of all 50 simulations. Although the average fit matched the veridical parameters for both Q+ (orange, right panel) and random (blue, left panel) stimulus selection, the resulting spread of parameter fits is visibly narrower for Q+.
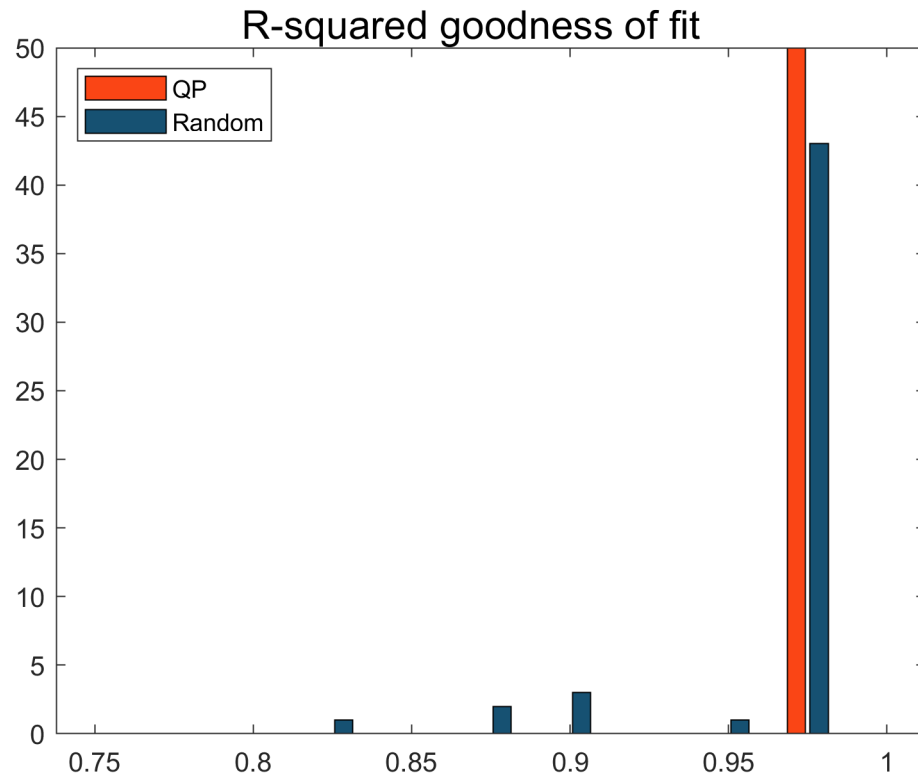
Figure 3. To evaluate how well the resulting parameters matched the veridical parameters we calculated the squared Pearson's correlation between the output for each simulation with the output for the veridical parameters. This histogram depicts the results of each correlation for all Q+ (orange) and random (blue) stimulus selection simulations.
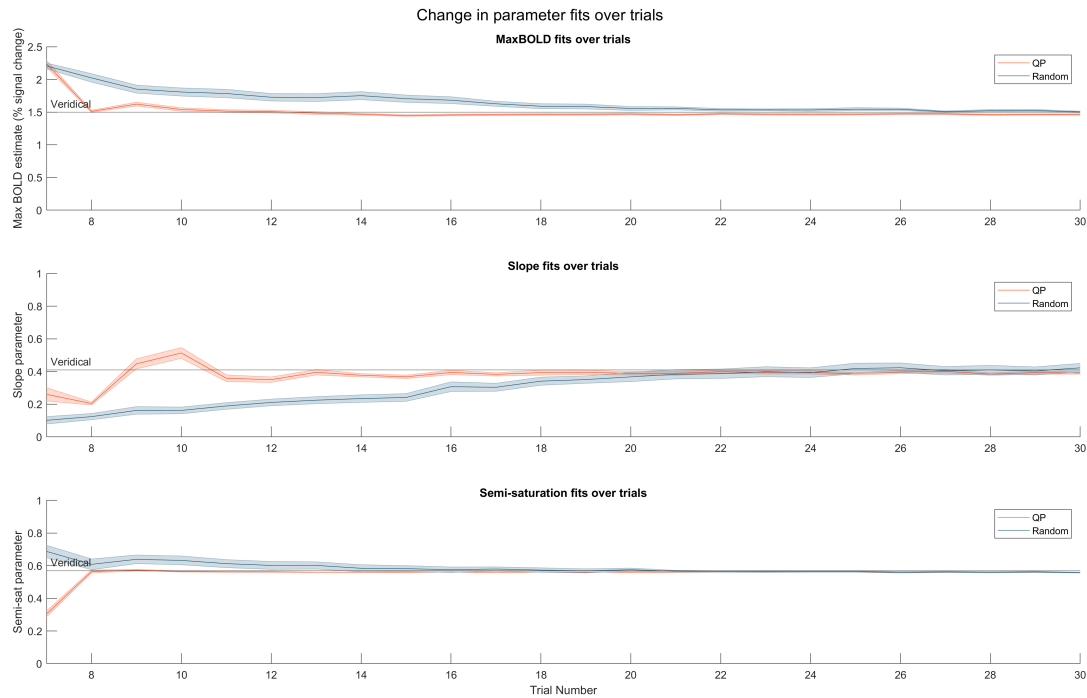
Figure 4. Excluding the first 5 trials (which were chosen to be alternating baseline and maximum BOLD trials), we plot the mean (thick line) and standard error (shaded area) estimates for the maximum BOLD estimate (top panel), slope (middle panel) and semi-saturation (bottom panel) averaged over all simulations after each trial. All three parameters are estimated more quickly and with less error when Q+ stimulus selection is used (orange) compared to random stimulus selection (blue).

Varying control parameters

Q+ improves model fit over random stimulus selection for one set of control parameters, but how does Q+ stimulus selection perform over a range of control parameters? We varied two properties of the experimental design: time to repetition (TR; 800 or 1600 ms) and how long each trial was (8 or 12 seconds). We also varied one property of the reverse model: the number of outcomes categories or bins (7 or 15). Finally, we explored three levels of noise in the response (.15, .45, .75 SDs). These levels could roughly correspond to low, medium, and high noise. (FIGURE XX). The purpose of these set of simulations was to see both how robust the Q+ stimulus selection advantage was and whether any experimental design choices may affect the accuracy of the results.

Table 1 presents the overall findings averaged over each set of 50 simulations.

Table 1
Results by simulation parameter sets.

Parameters				Results					
Noise	Bins	Trial Length (s)	TR	qpMean	qpSD	randMean	randSD	Z-score	p-value
0.1500	7	8	800	0.9952	0.0192	0.9821	0.0498	4.8636	0.0000
0.1500	7	8	1600	0.9979	0.0021	0.9849	0.0277	3.9191	0.0001
0.1500	7	12	800	0.9986	0.0012	0.9817	0.0348	5.1256	0.0000
0.1500	7	12	1600	0.9972	0.0030	0.9819	0.0451	1.6097	0.1075
0.1500	15	8	800	0.9976	0.0023	0.9836	0.0363	4.4224	0.0000
0.1500	15	8	1600	0.9978	0.0023	0.9825	0.0297	6.0631	0.0000
0.1500	15	12	800	0.9971	0.0044	0.9801	0.0628	3.4159	0.0006
0.1500	15	12	1600	0.9981	0.0020	0.9853	0.0273	5.1462	0.0000
0.4500	7	8	800	0.9829	0.0258	0.9609	0.0787	2.4508	0.0143
0.4500	7	8	1600	0.9728	0.0391	0.9471	0.0681	1.4718	0.1411
0.4500	7	12	800	0.9767	0.0396	0.9587	0.0504	2.9471	0.0032
0.4500	7	12	1600	0.9832	0.0297	0.9460	0.0623	4.1260	0.0000
0.4500	15	8	800	0.9880	0.0163	0.9409	0.0700	4.9394	0.0000
0.4500	15	8	1600	0.9770	0.0462	0.9290	0.0811	4.2018	0.0000
0.4500	15	12	800	0.9886	0.0145	0.9445	0.0759	5.3668	0.0000

Table 1 continued

	Parameters			Results					
Noise	Bins	Trial Length (s)	TR	qpMean	qpSD	randMean	randSD	Z-score	p-value
0.4500	15	12	1600	0.9874	0.0096	0.9517	0.0578	3.1746	0.0015
0.7500	7	8	800	0.9084	0.0794	0.8591	0.1446	1.5821	0.1136
0.7500	7	8	1600	0.8479	0.2067	0.8214	0.1661	1.4925	0.1356
0.7500	7	12	800	0.9143	0.1026	0.8851	0.0894	2.0578	0.0396
0.7500	7	12	1600	0.8910	0.1960	0.8355	0.1850	3.0850	0.0020
0.7500	15	8	800	0.8474	0.2254	0.8454	0.1540	1.5992	0.1098
0.7500	15	8	1600	0.8421	0.1743	0.7961	0.1830	1.8303	0.0672
0.7500	15	12	800	0.9044	0.1453	0.8698	0.1392	1.9199	0.0549
0.7500	15	12	1600	0.9044	0.1596	0.8393	0.1784	2.7541	0.0059

Note. Non-parametric t-tests (Wilcoxon signed-rank) comparing each set of simulations for parameter values. TR = Time to repetition.

Table 2

Results by simulation parameter sets.

Predictor	Sum of Squares	df	Mean.Sq.	F	p	Partial eta squared
TR	0.0586	1	0.0586	0.7693	0.3805	0.0003
fMRI Noise	5.9027	2	2.9514	38.7407	0.0000	0.0303
Trial Length	0.1732	1	0.1732	2.2741	0.1317	0.0009
TR*fMRI Noise	0.1723	2	0.0862	1.1308	0.3229	0.0009
TR*Trial Length	0.5499	1	0.5499	7.2181	0.0073	0.0029
fMRI Noise*Trial Length	0.0856	2	0.0428	0.5616	0.5704	0.0005
Error	182.0005	2389	0.0762	NA	NA	0.4906
Total	188.9453	2398	NA	NA	NA	0.5000

Note. ANOVA

Table 2 presents the results of ANOVA comparing the three levels.

Show an example simulated experiment with the convergence of the adaptive procedure upon the parameter(s) of the response function

Explore choices for the trial duration and TR to find optimal experimental design for rapid and accurate parameter characterization. Examples: How finely to sample the stimulus space; scan length, number of trials, stimulus duration; Forcing occasional "baseline" trials to account for 1/f drift in the BOLD fMRI signal

Comparison of method of constant stimuli to Bayesian adaptive fMRI for time to achieve a given degree of confidence in an experimental parameter.

Discussion

Limitations

This is really only useful if you already have a strong prior on the possible shape of the response function.

255   A great use case is when the distribution and/or bounds on the parameter

256 values for a population is known, but the investigator now wishes to estimate

257 the parameter value for a particular individual under study.

258 **Extensions**

259   Model parameters that account for variation across cortical space. Could

260 have a model that takes as input multiple time-series and then has a parameter

261 that describes systematic variation in responses across space (e.g., retinotopic

262 mapping).

263   Could model not just parameters for the neural response, but physiologic

264 parameters as well. E.g., the parameters that define the shape of the

265 hemodynamic response, or saturating non-linearities in the conversion of

266 neural activity to BOLD signal.

267   **Results**

268   **Discussion**

References

David H. Brainard. (2017). mQUESTPlus: A Matlab implementation of QUEST+.

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive
  psychometric method. *Journal of Vision*, *17*(3), 10–10.
  https://doi.org/10.1167/17.3.10

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method.
  Perception & Psychophysics, *33*(2), 113–120. https://doi.org/10.3758/BF03202828