Planning Beyond the Next Trial in Adaptive Experiments

Woojae Kim, Mark A. Pitt, Zhong-Lin Lu, and Jay I. Myung

Department of Psychology, Ohio State University

Columbus, OH 43210

April 6 2015

Correspondence concerning this article should be addressed to Woojae Kim, Department of Psychology, Ohio State University, 1835 Neil Avenue, Columbus, OH 43210. Tel: 1-614-292-8185, Email: kim.1124@osu.edu

Keywords: adaptive experiments, dynamic programming, backward induction, sensory threshold measurement
Abstract
Adaptive experiments maximize inferences from a limited number of observations. Current adaptive methods optimize inference on the next trial only. A lingering question has been whether additional benefit would be gained by optimizing beyond the next trial. That is, by considering the consequences of all future next trials, inference might be improved even further. Dynamic programming (DP), a tool for planning into the future, is ideally suited to this problem because it provides a means of solving the otherwise intractable computation involved in such “global” optimization. The present study presents the first demonstration of DP in adaptive behavioral experiments. Application of DP to model-based sensory threshold estimation identified some general conditions that will and will not benefit from DP, including when looking further ahead than the next trial is unnecessary.

Significance Statement
Adaptive methods are used in behavioral experiments to maximize scientific inference with the goal of improving measurement precision and accelerating knowledge acquisition. Current adaptive approaches are limited to optimizing stimulus choice on the next trial only, in part because of the seemingly insurmountable computational hurdle of looking beyond a single trial (i.e., evaluating all possible next trials). We demonstrate in simulation experiments in the context of sensory threshold estimation that dynamic programming (DP) is a solution to this longstanding problem, and identify conditions that will and will not benefit from DP, including when looking further ahead than the next trial is unnecessary. This methodological breakthrough has the potential to benefit researchers in many fields in the behavioral sciences.
Experimentation in the behavioral and neural sciences involves the process of presenting stimuli to observers and measuring their responses (e.g., categorization, response time, brain activity), with the goal of inferring the structural and functional properties of the brain. The quality of the inferences that can be drawn depends on the quality of the data collected. High-quality data, and thus better inference, can be obtained by improving the precision of measurement.

One means of improving measurement is by using adaptive data collection methods, in which the course of the experiment is tailored to each participant as a function of how that participant has responded in preceding trials. Known as adaptive measurement, adaptive estimation and active learning, methods of adaptive design optimization (ADO) maximize the efficiency of statistical inference. ADO finds an optimal stimulus for the next trial based on observations from past trials. Optimization can be applied to any type of inference, whether it be testing a hypothesis or estimating certain quantities (e.g., model parameters), and the use of ADO in experimentation is growing (1–7).

One simple form of ADO, popularized in psychophysics research, is the staircase procedure, which adjusts stimuli incrementally (e.g., increasing or decreasing stimulus intensity by predetermined units) based on the observer’s response (e.g., detecting or failing to detect a stimulus). Staircase procedures with various down-up rules have been proposed and their properties have been systematically studied (8). Other approaches to ADO have used rigorous, statistical criteria for determining when to alter the next stimulus (9), and precisely how much to alter it (1, 2). Perhaps the most advanced form of ADO is one that integrates Bayesian inference with information theory to achieve the maximum accrual of information about the unknown quantity being inferred. On each trial, each stimulus in the experimental design is evaluated on its potential informativeness given the data already collected and the objective of the experiment (e.g., parameter estimation). The design that is predicted to provide the largest gain in information is used on the next trial (2, 10, 11).

By design, adaptive methods involve sequential decision-making. Common among the above approaches is that they consider only the dependency between adjacent trials, operating under the assumption that accumulated information over the course of the entire experiment is optimized by choosing the stimulus that is expected to elicit the most information on the next trial. Intuitively, because this one-step-ahead optimization ignores the consequence of each
choice for trials beyond the next step, it is reasonable to believe that taking such consequences into account might further improve the accuracy of inference. Despite intermittent explorations (12, 13), this conjecture has not been systematically examined, making it a long-standing and unresolved issue in the field.

The purpose of the present paper was to explore the application of dynamic programming (DP), a technique developed in operations research for planning far ahead into the future, to the optimal design of experiments. We compared multiple look-ahead depths (1–100 trials) as well as different experimental designs to determine the conditions in which DP improves inference. We chose sensory threshold estimation as a testbed for the investigation. It is a common procedure for assessing the health and functioning of sensory systems in clinical and research settings, and the performance of adaptive methods in threshold estimation has been of considerable interest (3).

Dynamic Programming

Since the seminal work by Richard Bellman in the 1950’s, dynamic programming has referred to a variety of methods that take advantage of recursive local structure in a given problem to find a globally optimal solution. DP has been applied to a wide range of problems that involve making decisions across varying time horizons, including operations research, automatic control, artificial intelligence, and economics (14–17).

Most DP applications can be conceptualized as shortest-path problems. Consider the network of nodes shown in Fig. 1. Nodes represent certain states that one should go through in the process of solving a problem, such as navigating between cities. Arrows connecting the nodes represent possible transitions from one state to another, made by the problem solver. At any state during the task (e.g., C1–C3), the solver needs to make a decision that determines a transition to the next state (D1–D2). The decision should be made carefully because each resulting transition is associated with a cost (e.g., travel time), represented by the numbers on the edges of the network.

Suppose that one tries to find the path from the initial state A to the goal state E that incurs the smallest total cost (e.g., elapsed time) caused by all required transitions. DP provides a solution by exploiting the problem’s recursive structure. The best starting point to identify a solution is at the last stage, state D1 or D2, where no decision needs to be made because there is
only one transition for each. At the second-to-last stage (i.e., $C_1$–$C_3$), the optimal decision that leads to the smallest total cost can be made with the knowledge of the optimal decisions at the last stage (D). By carrying this accumulated knowledge backward through the transition sequence, it is possible to establish a knowledge system for sequential decisions that provides the optimal solution to the entire problem (total cost of 19; green line in Fig 1).

The procedure of working backward to find the shortest path is known as *backward induction* (18). This form of DP succeeds because as long as a certain state is reached, the globally optimal decision at that state will be the same whatever the previous path has been. It is because of this property of induction, known as the **principle of optimality**, that global optimization can be broken down into recurring local optimization problems. Backward induction is sometimes not feasible due to the scale of the problem (e.g., state space is high-dimensional), in which case probabilistic methods (*approximate* DP) attempt to estimate the globally optimal cost for each state by sampling paths of states and decisions [see (14, 15)].

Another approach to finding the shortest path is to loop over all possible paths and map all associated total costs to single out the smallest of them. Such a brute-force method is rarely viable because the number of required operations (i.e., determining all costs) grows exponentially as the number of decision stages increases, making brute-force computation impractical. With DP, in contrast, the number of required operations increases only linearly.

How does the one-step-ahead strategy in conventional ADO fare in this example? The path generated by seeking the smallest local cost at every transition is shown in red in Fig. 1, and it does not result in the smallest possible total cost (25 vs. 19). Inference in adaptive experimentation can be implemented as a dynamic program in which the goal is to maximize the total inferential gain over all trials, or a set of trials of any length. We performed simulations to identify conditions that generate savings in sensory threshold estimation when DP is used and the length of the look-ahead horizon is varied.

**Using DP for Threshold Estimation**

Sensory thresholds are often measured by having participants make one of two response alternatives (e.g., light detection) after presentation of a stimulus at a given level (e.g., intensity). For the underlying threshold to be measured, it can be treated as the location parameter of an $S$-
shaped psychometric function that is to be estimated under Bayesian inference. For instance, the commonly used log-Weibull psychometric function (1) has the form

\[ \Psi_\theta(d) = .5 + .48 \left[ 1 - \exp \left( 10^{0.5/20(d - \theta)} \right) \right] \]  

which returns the probability of correct detection \( \Psi_\theta(d) \) with the input of stimulus intensity \( d \) given that the underlying threshold is \( \theta \). In this particular function, the upper limit (due to attention lapse), lower limit (due to guess), and slope of the function are given some assumed values (.98, .5, and 1.5, respectively). Given a stimulus in each trial, a participant’s binary response is assumed to be generated from a Bernoulli distribution with its parameter (probability of success) governed by the psychometric function. In this setup, the states of knowledge that the estimation algorithm should go through across multiple trials of the measurement are represented by the posterior distributions of the threshold after observing responses in each trial.

Within information theory, the amount of knowledge one can expect to gain upon the next trial by using a certain stimulus is quantified by the expected information gain associated with that stimulus choice. One-trial-ahead optimization finds the stimulus that maximizes this quantity [e.g., Psi method in (2)]. Formally, the expected information gain after the next trial is

\[
H_t(\Theta) - H_{t+1}(\Theta | Y_{t+1}, d_{t+1})
\]

where \( H_t(\Theta) \) denotes the entropy of the state of knowledge \( \Theta \) (posterior distribution of thresholds) upon trial \( t \), and the second term \( H_{t+1}(\Theta | Y_{t+1}, d_{t+1}) = \mathbb{E}_{Y_{t+1}, d_{t+1}}[H_{t+1}(\Theta)] \) is the conditional entropy of the next state of knowledge given potential observation \( Y_{t+1} \) made with stimulus \( d_{t+1} \) in trial \( t+1 \). Here, the prediction of \( Y_{t+1} \) is made by the posterior predictive distribution given all previous observations up to the current trial \( t \).

---

* In information theory, this measure of expected information gain is regarded as the mutual information between the variable being inferred and the data to be added (19). In the present context, maximizing the mutual information is equivalent to minimizing the conditional entropy of the inferred variable.
To go beyond one-trial-ahead optimization, one must consider expected information gain after presenting a sequence of multiple stimuli ahead of the current trial. The expected information gain after \( k \) additional trials is expressed as

\[
H_t(\Theta) - H_{r+k}(\Theta | Y_{r+1}, \ldots, Y_{r+k}, d_{r+1}, \ldots, d_{r+k})
\]

where the second term, defined as \( \mathbb{E}_{Y_{r+1}, \ldots, Y_{r+k} | d_{r+1}, \ldots, d_{r+k}}[H_{r+k}(\Theta)] \), is the conditional entropy of \( \Theta \) given a sequence of potential observations \( Y_{r+1}, \ldots, Y_{r+k} \) ahead made with stimulus choices \( d_{r+1}, \ldots, d_{r+k} \). Expecting that the measurement session will be finished after \( k \) trials, \( k \)-trial-ahead optimization prescribes that one should select the stimulus \( d_{t+1}^* \) for the next trial which comes first in the sequence \( d_{t+1}^*, \ldots, d_{r+k}^* \) that maximizes the expected information gain defined above.

To find a solution using DP, it is useful to observe correspondences between the above optimization problem and the shortest-path problem described earlier. Just as the cost of transitioning from one state to another depends on the result of each decision-making stage in the shortest-path problem, the information gain expected from each future state transition (e.g., from \( \Theta | Y_{r+1}, d_{r+1} \) to \( \Theta | Y_{r+1}, Y_{r+2}, d_{r+1}, d_{r+2} \)) depends on a stimulus choice in each trial (e.g., \( d_{r+2} \)). What matters, however, is the total information gain after all scheduled future trials.† As with the shortest-path problem, we can work backward to maximize the expected information gain over multiple trials. For each of all possible states in the second-to-last trial, or trial \( t + k - 1 \), the optimal stimulus \( d_{t+k}^* \) for the last trial can be found using one-trial-ahead optimization. Repeat this for all possible states in trial \( t + k - 1 \) and store the corresponding optimal stimuli and expected information gain in a table. Then, going one trial backward, for each of possible states in trial \( t + k - 2 \), find the optimal stimulus \( d_{t+k-1}^* \) using the fact that, whatever transition to the next state is made, the maximized expected information gain from \( that \) state on can be looked up from the table previously filled. This means that the required computation is no more expensive than one-trial-ahead optimization because the algorithm needs to loop over stimulus choices only for the next trial and choices in the future are already reflected in the results in the look-up table.

† The total expected information gain in Eq. 3 can be written as the sum of expected information gains from each of consecutive state transitions in the future.
By repeating this procedure until the current trial \( t \) is reached, the optimal stimulus \( d_{t+1}^* \) in trial \( t+1 \) can be found.

In fact, the idea of applying DP by means of backward induction to parameter estimation of statistical models was introduced early in statistics \((20)\), but actual implementations of the procedure have been limited to shallow look-ahead depth (maximum 4) for simple models. A distinctive challenge has been that, unlike the shortest-path or other sequential decision-making problems, the space of possible states (i.e., posterior distributions) is not preset to a manageable size but increases exponentially as observations accumulate over trials. If a low-dimensional sufficient statistic exists to summarize posterior distributions, the state space may be discretized and represented on a grid \((21)\). However, sufficient statistics do not always exist, which is also the case of threshold estimation in this study.

The current implementation adopted the constrained backward induction \((22)\) in which the state space is approximated by a low-dimensional statistic that can adequately describe posterior distributions. While the basic concept of this approach is solid, it is not in the form of a straightforward recipe so its implementation entails a problem-specific solution. A key characteristic of the posterior distributions arising from model-based threshold estimation is their highly asymmetric tail thicknesses, which are caused by different lower asymptotes of the likelihood functions for success and failure responses. We parameterized possible shapes of the posterior distribution, which were not amenable to standard probability distributions, by the multiplication of a normal distribution and a sigmoid function of the form

\[
p(\theta | \mu, \sigma, \delta) = \frac{1}{c} \exp\left(-\frac{(\theta - \mu)^2}{2\sigma^2}\right) \left[\frac{1}{1 + \exp\left(\frac{\pi(\theta - \mu)}{\sigma\sqrt{3}}\right)} + \frac{1}{\delta}\right]
\]

where \( c \) is a normalizing constant, and the three parameters \( \mu, \sigma, \) and \( \delta \) determine the location, dispersion, and asymmetry of the distribution, respectively. The space of these parameters was represented on a grid and treated as the state space for backward induction. Adequacy of the approximation was assessed in multiple ways, including checking its visual fit to posterior distributions generated from simulated experiments, confirming the convergence of total expected information gain (Eq. 3) with an increase in grid resolution, and by comparing the accuracy of threshold estimation between two implementations of backward induction with and
without approximation (this was possible only up to three-trial-ahead optimization). We also performed a sensitivity analysis of sampling densities on parameter (threshold) and stimulus spaces, and found that convergence of estimation performance occurs before reaching the densities employed to generate the results presented in this paper.

Simulation Experiments

Different look-ahead depths for sensory threshold estimation were evaluated in two conditions. The first, unconstrained, condition assumed a simple, conventional procedure in which any of the preset spectrum of stimuli on the studied physical dimension (e.g., visual contrast, luminance intensity) could be selected in each measurement trial with no restriction on their order of presentation. In the second, order-constrained, condition, dependencies on stimulus choice were imposed across the experiment. Once a stimulus was presented from one of ten equally spaced bins, future stimuli could be chosen only from the same bin or bins lower on the continuum (i.e., weaker stimuli) in subsequent trials. Creation of this condition was motivated by a consideration of the dependencies involved in planning multiple trials ahead. The requirement on stimulus choice in the order-constrained condition should create a strong sequential dependency across decision-making trials, possibly accentuating the effect of further-ahead optimization. Our hope was that comparison of this with the unconstrained condition could help determine the principle behind the performance of optimizing beyond the next trial.

Although such an order constraint is rarely used when measuring auditory and visual thresholds, its use is commonplace in taste and odor studies (23, 24).

In each of the two, unconstrained and order-constrained conditions, simulated threshold estimation sessions were conducted in which the look-ahead depth of stimulus optimization was manipulated. First, conventional, one-trial-ahead optimization served as a baseline against which algorithms with longer look-ahead distances were compared.‡ Second, assuming that a measurement session is scheduled to finish in a fixed number of trials, full-horizon optimization was implemented. Using the procedure of $k$-trial-ahead optimization described earlier, the

‡ One-trial-ahead optimization was implemented using the Psi method (2) with threshold and stimulus spaces sampled in approximately 0.5 dB steps.
algorithm starts with the look-ahead depth \( k \) made equal to the total number of trials in the experiment (100), and then, over the course of the session, decreases the depth incrementally to match the remaining number of trials. Last, algorithms for intermediate-horizon optimization were included to assess the relationship between horizon depth and any improvement in inference. Given a look-ahead depth \( k \) less than the total number of trials, this algorithm starts with \( k \)-trial-ahead optimization and maintains it until the number of remaining trials is reduced to \( k \). After that point, the look-ahead depth is decreased incrementally to match the number of remaining trials. Intermediate values of \( k \) were 5, 20, and 50.

In all simulations, the data-generating structure was the log-Weibull psychometric function shown in Eq. 1. In each measurement session, the true, underlying threshold was sampled from a uniform prior distribution of log intensities ranging from -60 to -0.9 dB. In each trial of a session, an optimal stimulus was determined by the algorithm of the given look-ahead depth from the range of log intensities -66 to -0.2 dB discretized in approximately 0.5 dB steps. Then, a response to it was generated by Bernoulli sampling from the assumed psychometric function value at the optimal stimulus. In each simulation condition, 100,000 independent replications of measurement sessions, each consisting of 100 trials, were made with underlying thresholds drawn from the uniform distribution.

**Results**

To examine how efficiently the threshold is estimated over measurement trials, estimation error was quantified by the root mean squared error (RMSE) defined by

\[
\text{RMSE}(\hat{\theta}_t) = \sqrt{\mathbb{E}[(\hat{\theta}_t - \theta_{\text{true}})^2]}
\]

where \( \hat{\theta}_t \) is the estimate of the threshold for a simulated session, which was obtained as the posterior mean after observing trial \( t \)’s outcome, \( \theta_{\text{true}} \) is the true, underlying threshold for that session. The expectation is assumed to be over all underlying thresholds and replicated sessions, and hence was replaced by the sample mean over 100,000 simulated sessions. Smaller values indicate greater accuracy of threshold estimation.

Fig. 2 shows the error in threshold estimation observed under the two constraint conditions (left and right panels) with different optimization depths (curves in each panel). When
stimulus selection was unconstrained, no benefit was provided by looking further ahead than one trial. The gray curve in Fig. 2A shows the RMSEs (y-axis) over 100 trials (x-axis) incurred by one-trial-ahead optimization. The estimation error drops quickly with observations in the first 20 to 30 trials, but after that, the rate of error reduction slows greatly as more data are collected. If the longer look-ahead horizons in the other conditions improved inference, those curves should be below this one. That just the opposite is found over the majority of trials for all but the 5-trial-ahead condition indicates that inference is actually worse, except in the last few trials where the curves converge to values close to each other.

The reason for the banana-shaped pattern, where the curves diverge in the center but converge at the endpoints, is due to the different goals of the one-trial versus multi-trial-ahead optimization. In DP applications with long look-ahead horizons, a drop in performance during the early stages is not a surprise but a common phenomenon. This is because the algorithm is designed to seek longer-term gains rather than the highest immediate gains. Hence, the performance of farther-ahead optimization must be duly evaluated after many decision stages have passed. In particular, in the case of a finite-horizon problem like the current one (i.e., with a fixed number of decision stages), it must be assessed when the session is close to its end and the longer-term gains are harvested by decreasing the look-ahead depth. This means that even if looking further ahead than one trial were not necessary in the unconstrained condition, or one-trial-ahead optimization were sufficient to achieve the highest possible reduction of error, longer-horizon optimization should exhibit the same level of performance at the end of a session, but not necessarily in the middle of it. Indeed, the gap between their RMSEs and those of the one-trial-ahead algorithm is negligible when the session ends. §

The reason why this gap is not completely reduced to zero is due to approximation error in the DP algorithms. Recursive computing over many stages with approximate terms (e.g., approximate states in the current implementation), which is often the nature of DP applications,

§ This observation is not because of the number of trials (100) being large enough for the error to converge to a certain level. We performed the same simulation with the total scheduled trials of 50, 150 and 200, and found that different optimization depths exhibit the same banana-shaped patterns of error reduction.
causes approximation error to propagate continually over the stages, resulting in performance degradation below the theoretically highest level (25). For this reason, using DP with the deepest possible horizon may not automatically improve inference, and this is evident in our application of DP in the unconstrained condition. As can be seen at the last trial in the results graph, error propagation increases as the look-ahead horizon (20, 50, 100 trials) increases. On the other hand, shorter horizons make the algorithms converge to the one-trial-ahead optimization, as shown by the one- and five-trial-ahead algorithms being nearly indistinguishable.

One may wonder whether the efficiency of threshold estimation might be improved if the approximation errors in multi-trial-ahead optimization were eliminated. If there were truly some benefit, it might be exhibited even in two- or three-trial-ahead optimization though the effect could be small. To address this suspicion, we performed the unrestricted, exact form of backward induction in which no approximation to the state space was made up to depths of three trials. The results were indistinguishable from the one-trial-ahead optimization. In addition, comparisons were also made with the two- and three-trial-ahead algorithms based on approximated states (i.e., current DP implementation). All RMSE curves were on top of one another. The fact that the performance of the currently implemented DP algorithm was identical to that of the exact DP algorithm also served as an independent check of the adequacy of implementation.

The above results demonstrate that, contrary to intuition, looking further ahead in stimulus optimization provides no benefit for threshold estimation in the unconstrained condition, meaning that the commonly used one-trial-ahead optimization is not only sufficient, but optimal. The same conclusion, however, was not drawn when a strong sequential dependency was introduced to stimulus selection. Fig. 2B shows the RMSEs of the five different, look-ahead algorithms under the measurement procedure that constrained the order of stimulus presentation, and as can be seen, differences in final error levels are evident. The one-trial-ahead curve drops rapidly for the first 20 trials, but then struggles to decrease thereafter. By contrast, the algorithms that look further ahead start with greater error during the initial 15 trials, trailing behind those algorithms with shorter horizons. Soon after this point (no later than trial 30), the benefits of looking further ahead start to accrue, with the curves of the longer-horizon conditions dipping below the 1-trial condition. These benefits continue to increase across trials for the longer-horizon conditions (20, 50, 100), so that by the end of the experiment, RMSE has been reduced to about 50% (2 dB) of that in the 1-trial condition. There seems to be a limit to the benefit of
looking further ahead, as indicated by negligible difference between the 20- and 50-trial-ahead and full-horizon algorithms at the last trial.

Overall, as the look-ahead depth increases, the RMSE reduction is slower initially, but persists longer, resulting in considerably lower error by trial 100. This pay-off in seeking long-term gains by holding back from immediate gains is a clear distinction from what was observed in the unconstrained condition, and comes about by a shift in the stimulus selection strategy. Under the order constraint, selecting stimuli that provide immediate gains without considering the consequences of that decision for stimulus choices in future trials quickly leaves the algorithm with a severely limited range of stimuli to choose from. In contrast, algorithms with greater look-ahead vision avoid falling into this trap by being more conservative in selecting stimuli, with look-ahead depth modulating how conservative the algorithm behaves; the longer the horizon, the more conservative the choice of stimuli. Note that none of these strategies, including the 100-trial, yields a final RMSE as low as that in the unconstrained condition in the left graph (1.74 vs. 1.36 dB). This is because the constraint itself imposes a cost in estimation by restricting stimulus choice.

Discussion

Adaptive methods are designed to improve inference in experimentation. To date, they have achieved this by optimizing inference on the next trial. A lingering question in the discipline has been whether additional benefit would be gained by optimizing beyond the next trial. That is, by considering the consequences of all future next trials, can inference be improved even further? Dynamic programming, a tool for planning into the future, is ideally suited to this problem because it provides a means of solving the otherwise intractable computation involved in multi-trial-ahead optimization. In short, it is a “smart” means of assessing the quality of each next trial by considering the quality of possible subsequent trials.

A technique that enables one to look into the future when conducting an experiment is truly appealing. However, given the complexity of its implementation, one would rightly wonder whether it is really necessary. It is entirely possible that one-trial-ahead, greedy optimization is sufficient for the given problem, or looking further ahead, even far into the future, might offer no appreciable benefit to justify using DP. Because it is generally not possible in most real-world optimization problems to know the answer in advance [e.g., analytic proof about a need for
global optimization; for exceptions, see minimum spanning trees and Huffman trees problems (26), researchers are faced with the dilemma of not knowing which to use unless the two are compared.

Our demonstration of multi-trial-ahead optimization’s performance in the unconstrained condition resolves this issue in the domain of sensory threshold estimation. Application of DP showed that looking further ahead than one trial in selecting stimuli has no additional benefit when stimulus selection is unconstrained. This outcome has far-reaching implications. Until now, practitioners measuring thresholds in the clinic or laboratory had to acknowledge the unresolved issue surrounding the optimality of their adaptive method (2, 13, 27). At an extreme, they can be confronted by a bold statement that the myopic, one-trial-ahead strategy in use might bring out merely a marginal portion of the total improvement that must be possible with full-horizon optimization. The present results enable practitioners to be confident that the conventional, one-trial-ahead method is indeed optimal.

The impact of the current study extends beyond sensory threshold estimation, to most model-based experimental settings. The generality of the contribution can be grasped when it is framed as a theoretical explanation of why such results are produced. Pragmatically, this insight provides guidance on determining when DP should be used. The statistical assumption made for the unconstrained threshold estimation is

$$p(y_t | \theta, d_t, \{y_{t-1}, d_{t-1}, \ldots, y_1, d_1\}) = p(y_t | \theta, d_t)$$

where $\theta$ denotes the model’s parameter(s), $d_t$ represents the stimulus, independent variable(s), or more generally, design used in trial $t$, and $y_t$ is the observation in trial $t$. This assumption translates that data generation in a certain trial depends only on the design chosen in that trial and the underlying parameter setting, and is thus independent of the trajectory of previous design choices and resulting outcomes. Under this assumption, it can be shown that one-trial-ahead optimization is asymptotically optimal, meaning that it can be as efficient as methods of any look-ahead abilities under a schedule of sufficiently many trials (28, 29).

Most of statistical models in psychology require a relatively large number of observations for them to be inferred with meaningful precision, due to high sampling error in data. Hence, it can be said that if a model assumes independence as stated in Eq. 6, like the psychometric model for threshold estimation, using the one-trial-ahead method will most likely be sufficient.
irrespective of specifics of the model or the design (e.g., a model with two parameters instead of one and stimuli of two dimensions rather than one). Nevertheless, a formal proof about the greedy method’s sufficiency is generally difficult to attain in the case of finite-horizon DP problems. Essentially, what our simulation study provides is a concrete demonstration, whose results cannot be predicted by the asymptotic optimality alone, that the conventional method is indeed sufficient for an inference problem of practical significance. We believe that the results have generality because many modeling problems in behavioral sciences are mechanically similar (e.g., high sampling error, simple parametric forms, relatively low-dimensional designs).

To investigate further the question of what is a sufficiently large number of trials in order for the above explanation to hold for unconstrained threshold estimation, we performed yet another simulation in which full-horizon optimization is used when the total scheduled trials are 2 to 10. Although such short lengths of a session have no practical implications, the simulation helped answer the question. We found that the full-horizon method has an infinitesimal edge over the one-trial-ahead optimization but it does only up to a session length of 5. Beyond that length, whether it being due to the convergence of the two methods or the approximation error in DP, full-horizon optimization never outperformed one-trial-ahead optimization.

In contrast to the unconstrained condition, the order constraint represents dependency of a type that violates the independence assumption in Eq. 6 in a particular way. Whenever the constraint is not observed on design $d_i$, the probability $p(y_i | \theta, d_i, (y_{i-1}, d_{i-1}), \ldots, (y_{1}, d_1))$ is undefined, or defined to be zero, not being able to equal $p(y_i | \theta)$. This leads to zero expected information gain from such $d_i$, which would not be the case without the constraint. This observation should make it clear that there exists a bad choice of a stimulus, the quality of which cannot be seen under one-trial-ahead optimization, but can cause an adverse effect on estimation in the long run because it severely restricts the range of stimulus choices later on.

The application of DP under sequential dependency is not limited to a strict definition of decision trajectories like the order constraint used in our study. Likely examples of models that will benefit from DP include models of learning and dynamic decision making processes, in which the dependency of participant responses across trials itself is the subject of study and built into the models. Processing models in these areas are usually first-order Markovian models in which a participant’s latent state keeps being updated with each new observation and the model’s
prediction depends on the most recent state \((30-33)\). In these models, design choices and participant responses across previous trials influence the predicted probability of data in the next trial, violating the independence assumption in Eq. 6, and thus it is likely advantageous to look far ahead in optimizing the designs sequentially.

Of primary consideration, then, is whether the stimulus has a lasting impact on the participant. If the stimulus imparts a short-term change in the individual (brief sensory registration or perceptual identification), one-trial-ahead optimization should be optimal. To the extent that stimulus influences are longer-lasting, and thus lead to possible sequential dependencies in responding to stimuli across trials, DP could improve inference with a model that captures such dependencies. Concerning threshold estimation, for example, one may suspect that the underlying psychometric function adapts over time in real experiments, the behavior of which may depend on the trajectory of previous stimuli and outcomes. This non-stationary behavior violates the independence assumption, and may be specifically modeled. The timeframe of such carry-over effects will differ across the domains of study, from less than a few minutes when measuring sensory adaptation, to permanent changes in the individual when studying learning, whether it being perceptual learning in vision or higher-order learning as found in skill acquisition. Although the use of full-horizon DP would ensure optimality in all these cases, if one knows the extent of the dependency across trials, shorter-horizon optimization is an option.

Finally, because our main interest in this study was the effect of look-ahead depth on the performance of adaptive experiments, we made the simplifying assumption that the data-generating process is identical to the theoretical model. The behavior of multi-trial-ahead optimization under a misspecified model may need to be investigated to understand the consequences when used in empirical settings (e.g., a model of non-stationary adaptation simulates the generating process). Ultimately, considering that all models are approximations, what needs to be decided is whether the benefits obtained by the use of sequentially global optimization overshadow any losses incurred by the use of an approximate model or the approximation error in the method. This will be a domain-specific problem.

To the best of our knowledge, this study represents the first demonstration of DP in adaptive behavioral experiments, albeit simulated. In addition to showing that one-trial-ahead optimization is sufficient for sensory threshold estimation, the results provide insight into more general conditions that will and will not benefit from DP. Although there is still much to be
learned about the methodology, including overcoming technical challenges, it is a promising tool for improving inference in the behavioral and brain sciences.

**Acknowledgments**

This research is supported by National Institute of Health Grant R01-MH093838 to J.I.M and M.A.P.

**References**


**Figure Legends**

**Fig. 1.** Network depicting the shortest-path problem. Nodes represent locations (e.g, cities), and the numbers adjacent to the arrows connecting the nodes represent the cost (distance) of moving from one node to another. The red line from A to E is the path (total cost of 25) taken when using one-step-ahead optimization, a strategy in which the next node that is chosen is always the least costly. The green line denotes the shortest path from A to E (total cost of 19), and can be identified through backward induction by looking beyond the next node to consider the costs of future transitions between nodes.

**Fig. 2.** Performance of the five different look-ahead horizons when stimulus choice was unconstrained (*A*) and constrained (*B*) across trials. Each graph shows threshold estimation error (RMSE) as a function of trial number in the experiment. The depth of the horizon had no effect on final error levels in the unconstrained condition, whereas estimates were reduced by up to 2
dB in the constrained condition when the look-ahead horizon was the length of the experiment (100 trials). Precision bands are not included in the graphs because standard errors of the values are infinitesimal due to the large sample size (100,000). RMSE = root mean square error.

Figure 1

Figure 2