Designing optimal behavioral experiments using machine learning

Simon Valentin[1]*†, Steven Kleinegesse[1]†, Neil R Bramley[2], Peggy Seriès[1], Michael U Gutmann[1], Christopher G Lucas[1]

[1]School of Informatics, University of Edinburgh, Edinburgh, United Kingdom; [2]Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom

Abstract Computational models are powerful tools for understanding human cognition and behavior. They let us express our theories clearly and precisely and offer predictions that can be subtle and often counter-intuitive. However, this same richness and ability to surprise means our scientific intuitions and traditional tools are ill-suited to designing experiments to test and compare these models. To avoid these pitfalls and realize the full potential of computational modeling, we require tools to design experiments that provide clear answers about what models explain human behavior and the auxiliary assumptions those models must make. Bayesian optimal experimental design (BOED) formalizes the search for optimal experimental designs by identifying experiments that are expected to yield informative data. In this work, we provide a tutorial on leveraging recent advances in BOED and machine learning to find optimal experiments for any kind of model that we can simulate data from, and show how by-products of this procedure allow for quick and straightforward evaluation of models and their parameters against real experimental data. As a case study, we consider theories of how people balance exploration and exploitation in multi-armed bandit decision-making tasks. We validate the presented approach using simulations and a real-world experiment. As compared to experimental designs commonly used in the literature, we show that our optimal designs more efficiently determine which of a set of models best account for individual human behavior, and more efficiently characterize behavior given a preferred model. At the same time, formalizing a scientific question such that it can be adequately addressed with BOED can be challenging and we discuss several potential caveats and pitfalls that practitioners should be aware of. We provide code to replicate all analyses as well as tutorial notebooks and pointers to adapt the methodology to different experimental settings.

***For correspondence:** simonvalentin@me.com

†These authors contributed equally to this work

Competing interest: The authors declare that no competing interests exist.

Funding: See page 23

Received: 17 January 2023
Preprinted: 12 May 2023
Accepted: 19 November 2023
Published: 23 January 2024

Reviewing Editor: Joshua I Gold, University of Pennsylvania, United States

Introduction

Computational modeling of behavioral phenomena is currently experiencing rapid growth, in particular with respect to methodological improvements. For instance, seminal work by ***Wilson and Collins, 2019*** has been crucial in raising methodological standards and bringing attention to how computational analyses can add value to the study of human (and animal) behavior. Meanwhile, in most instances, computational analyses are applied to data collected from experiments that were designed based on intuition and convention. That is, while experimental designs are usually motivated by scientific questions in mind, they are often chosen without explicitly and quantitatively considering how informative these data might be for the computational analyses and, finally, the scientific questions being studied. This can, in the worst case, completely undermine the research effort, especially as particularly valuable experimental designs can be counter-intuitive. Today, advancements in machine learning (ML) open up the possibility of applying computational methods when deciding how experiments should be designed in the first place, to yield data that are maximally informative with respect to the scientific question at hand.

In this work, we provide an introduction to modern BOED and a step-by-step tutorial on how these advancements can be combined and leveraged to find optimal experimental designs for any computational model that we can simulate data from, and show how by-products of this procedure allow for quick and straightforward evaluation of models and their parameters against real experimental data. Formalizing the scientific goal of an experiment such that our design aligns with our intention can be difficult. BOED forces the researcher to make explicit, and engage with, various assumptions and constraints that would otherwise remain implicit, which helps in making experimental research more rigorous but can also lead to unexpected optimal designs. We discuss several issues related to experimental design that practitioners may encounter no matter how they design their experiments. Importantly, these issues are made more salient by formalizing the process.

Why optimize experimental designs?

Experiments are the bedrock of scientific data collection, making it possible to discriminate between different theories or models, and discovering what specific commitments a model must make to align with reality by means of parameter estimation. The process of designing experiments involves making a set of complex decisions, where the space of possible experimental designs is typically navigated based on a combination of researchers' prior experience, intuitions about what designs may be informative, and convention.

This approach has been successful for centuries across various scientific fields, but it has important limitations that are amplified as theories become richer, more nuanced, and more complex. Often, a theory does not make a concrete prediction but rather has free parameters that license a variety of possible predictions of varying plausibility. For example, if we want to model human behavior, it is important to recognize that different participants in a task might have different strategies or priorities, which can be captured by parameters in a model. As we expand models to accommodate more complex effects, it becomes ever more time-consuming and difficult to design experiments that distinguish between models, or allow for effective parameter estimation. As an example from cognitive science, there are instances where models turned out to be empirically indistinguishable under certain conditions, a fact that was only recognized after a large number of experiments had already been done (***Jones and Dzhafarov, 2014***). Such problems also play an important role in the larger context of the replication crisis (e.g. ***Pashler and Wagenmakers, 2012***).

Collecting experimental data is typically a costly and time-consuming process, especially in the case of studies involving resource-intensive recording techniques, such as neuroimaging or eye-tracking (e.g. ***Dale, 1999***; ***Lorenz et al., 2016***). There is a strong economic as well as ethical case for conducting experiments that are expected to lead to maximally informative data with respect to the scientific question at hand. For instance, in computational psychiatry, researchers are often interested in estimating parameters describing people's traits and how they relate, often doing so under constrained resources in terms of the number of participants and their time. Additionally, researchers and practitioners alike require that experiments are sufficiently informative to provide confidence that they would reveal effects if they are there and also that negative results are true negatives (e.g. ***Huys et al., 2013***; ***Karvelis et al., 2018***). Consequently, as theories of natural phenomena become more realistic and complex, researchers must expand their methodological toolkit for designing scientific experiments.

Bayesian optimal experimental design

As a potential solution, *Bayesian optimal experimental design* (BOED; see ***Ryan et al., 2016***; ***Rainforth et al., 2023***, for reviews) provides a principled framework for optimizing the design of experiments. Broadly speaking, BOED rephrases the task of finding optimal experimental designs to that of solving an optimization problem. That is, the researcher specifies all controllable parameters of an experiment, also known as experimental designs, for which they wish to find optimal settings, which are then determined by maximizing a utility function. Exactly what we might want to optimize depends on our experiment, but can include any aspect of the design that we can specify, such as which stimuli to present, when or where measurements should be taken in a quasi-experiment, or, as an example, how to reward risky choices in a behavioral experiment. The utility function is selected to measure the quality of a given experimental design with respect to the scientific goal at hand, such as model discrimination, parameter estimation or prediction of future observations, with typical choices including expected information gain, uncertainty reduction, and many others (**Ryan et al., 2016**), as explained below. Importantly, BOED approaches require that theories are formalized via computational models of the underlying natural phenomena.

Computational models

Scientific theories are increasingly formalized via computational models (e.g. **Guest and Martin, 2021**). These models take many forms, but one desideratum for a model is that, given a set of experimental data, we can compare it to other models based on how well it predicts that data. To that end, many computational models are constructed in a way that allows for the analytic evaluation of the likelihood of the model given data, by means of a likelihood function. However, this imposes strong constraints on the space of models we can consider. Models that are rich enough to accurately describe complex phenomena—like human behavior—often have intractable likelihood functions (e.g. **Lintusaari et al., 2017**; **Cranmer et al., 2020**). In other words, we cannot compute likelihoods due to the prohibitive computational cost of doing so, or the inability to express the likelihood function mathematically. This has at least two implications (1) When we develop rich models, we often lack the tools to evaluate and compare them directly, and are left with coarse, qualitative proxy measures–like the "the model curves look like the human curves for some parameter settings!"; (2) when we are intent on conducting a systematic comparison of models, we are often forced to use simplified models that have tractable likelihoods, even when the simplifications lead to consequential departures from the theories we might want our models to capture.

Simulator models: What if the likelihood cannot be computed?

Many realistically complex models have the feature that we can *simulate* data from them, which allows for using simulation-based inference methods, such as approximate Bayesian computation (ABC) and many others (e.g. **Lintusaari et al., 2017**; **Cranmer et al., 2020**). There has been a growing interest in this class of models, often referred to as *generative*, *implicit,* or *simulator* models (e.g. **Marjoram et al., 2003**; **Sisson et al., 2007**; **Lintusaari et al., 2017**; **Brehmer et al., 2020**; **Cranmer et al., 2020**). We will use the term simulator model throughout this paper to refer to any model from which we can simulate data. We note that some generative models have tractable likelihood functions, but this class of models is subsumed by simulator models, which do not make any structural assumptions on the likelihood and include important cases where the likelihood is too expensive to be computed, or cannot be computed at all (simulator models also include the common case of a potentially complex deterministic simulator with an additional observation noise component, as is used in many areas of science). Simulator models have become ubiquitous in the sciences, including physics (**Schafer and Freeman, 2012**), biology (**Ross et al., 2017**), economics (**Gouriéroux et al., 2010**), epidemiology (**Corander et al., 2017**), and cognitive science (**Palestro et al., 2018**), among others.

The study of human, and animal, perception and cognition is experiencing a surge in the use of computational modeling (e.g. **Griffiths, 2015**). Here, simulator models can formalize complex theories about latent psychological processes to arrive at quantitative predictions about observable behavior. As such, simulator models appear across cognitive science, including for example, many Bayesian and connectionist models, a wide range of process models, cognitive architectures, or many reinforcement learning models and physics engines (see, e.g. **Turner and Van Zandt, 2012**; **Turner and Sederberg, 2014**; **Bitzer et al., 2014**; **Turner et al., 2018**; **Bramley et al., 2018**; **Ullman et al., 2018**; **Palestro et al., 2018**; **Kangasrääsiö et al., 2019**; **Gebhardt et al., 2021**). The applicability and usefulness of experimental design optimization for parameter estimation and model comparison has been demonstrated in many scientific disciplines, but this has often been restricted to settings with simple and tractable classes of models, for instance in cognitive science (e.g. **Myung and Pitt, 2009**; **Zhang and Lee, 2010**; **Ouyang et al., 2018**).

An overview of this tutorial

In this tutorial, we present a flexible workflow that combines recent advances in ML and BOED in order to optimize experimental designs for any computational model from which we can simulate data. A high-level overview of the underlying ML-based BOED method is shown in ***Figure 1***, including

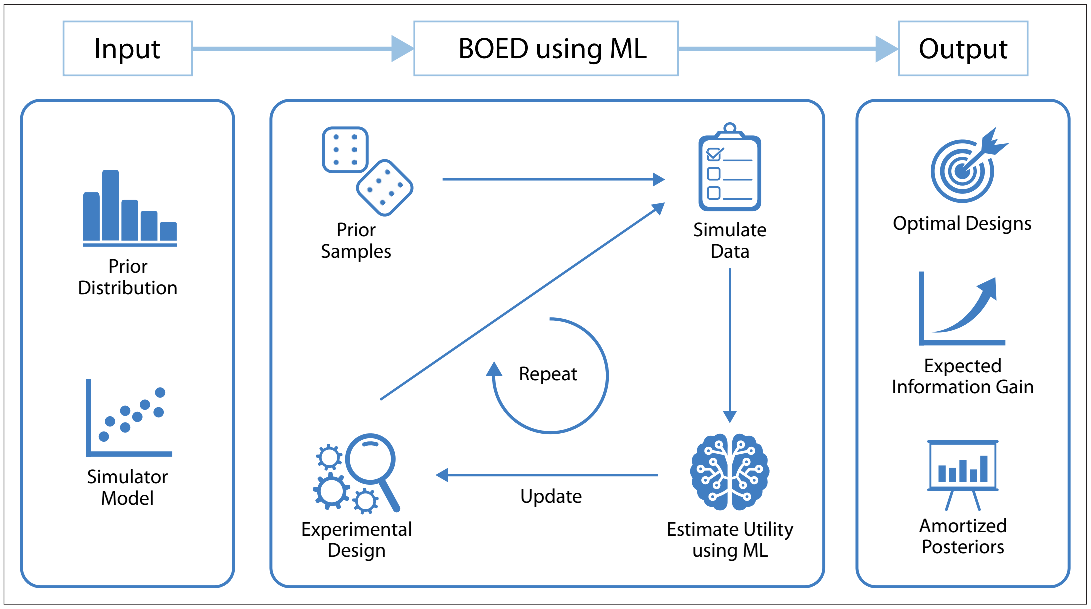


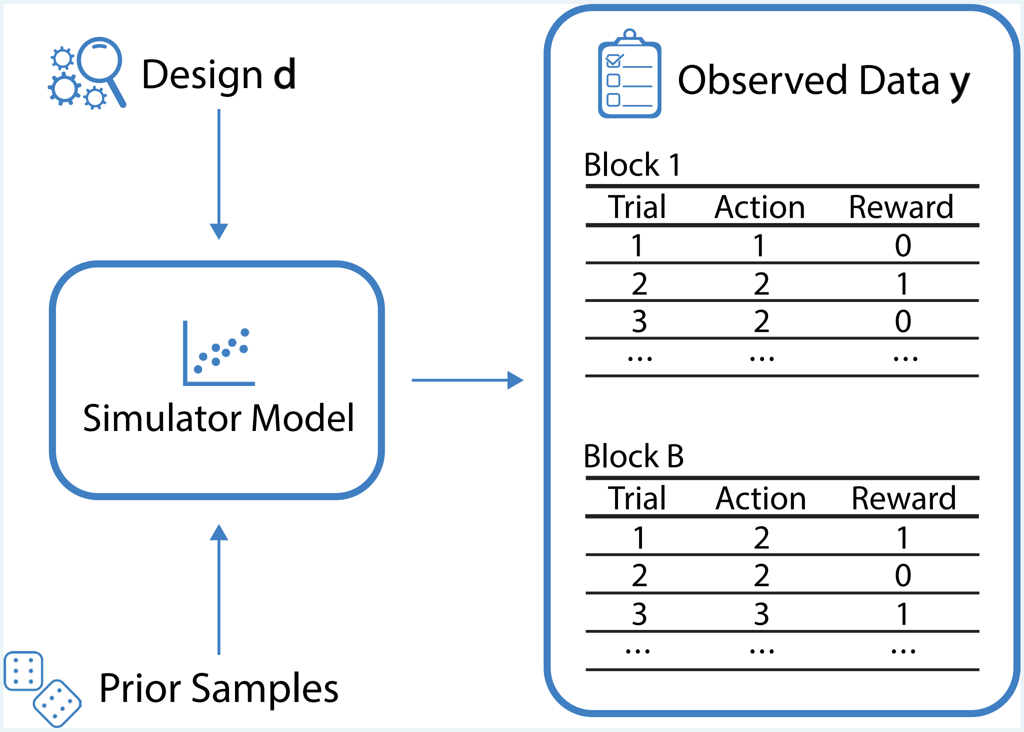
Figure 1. A high level overview of our approach that uses machine learning (ML) to optimize the design of experiments. The inputs to our method are a model from which we can simulate data and a prior distribution over the model parameters. Our method starts by drawing a set of samples from the prior and initializing an experimental design. These are used to simulate artificial data using the simulator model. We then use ML to estimate the expected information gain of those data, which is used as a metric to search over the design space. Finally, this is repeated until convergence of the expected information gain. The outputs of our method are optimal experimental designs, an estimate of the expected information gain when performing the experiment and an amortized posterior distribution, which can be used to cheaply compute approximate posterior distributions once real-world data are observed. Other useful by-products of our method include a set of approximate sufficient summary statistics and a fitted utility surface of the expected information gain.

researchers' inputs and resulting outputs. The presented approach is applicable to any model from which we can simulate data and scales well to realistic numbers of design variables, experimental trials and blocks. Furthermore, as well as optimized designs, the presented approach also results in tractable amortized posterior inference—that is, it allows researchers to use their actual experimental data to easily compute posterior distributions, which may otherwise be computationally expensive or intractable to compute. Moreover, we note that, to our knowledge, there has been no study that performs BOED for simulator models without likelihoods and then uses the optimal designs to run a real experiment. Instead, most practical applications of BOED have been limited to simple scientific theories of nature (e.g. ***Liepe et al., 2013***). This paucity of real-world applications is partly because the situations where BOED would be most useful are those where it has previously been computationally intractable to use. The ML-based approach presented in this work aims to overcome these limitations.

As with many new computational tools, however, the techniques for using BOED are not trivial to implement from scratch. In an effort to help other scientists understand these methods and use them to unlock the full potential of BOED in their own research, we next present a case study in using BOED to better understand human decision-making. Specifically, we (1) describe new simulator models that generalize earlier decision-making models that were constrained by the need to have tractable likelihoods; (2) walk through the steps to use BOED to design experiments comparing these simulator models and estimating the psychologically interpretable parameters they contain; and (3) contrast our new, optimized experiments to previous designs and illustrate their advantages and the new lessons they offer. All our steps and analyses are accompanied by code that we have commented and structured to be easily adapted by other researchers.

Box 1. Bandit tasks.

Bandit tasks constitute a minimal reinforcement learning task, in which the agent is faced with the problem of balancing exploration and exploitation. Various algorithms have been proposed for maximizing the reward in bandit tasks, and some of these algorithms have been used directly or in modified form as models of human behavior. A selection of such algorithms indeed captures important psychological mechanisms in solving bandit tasks (***Steyvers et al., 2009***). However, a richer account of human behavior in bandit tasks would seem to be required to accommodate flexibility and nuance of human behavior (e.g. see ***Dezfouli et al., 2019***; ***Lee et al., 2011***), leading to simulator models that do not necessarily permit familiar likelihood-based inference, due to, for example unobserved latent states, and which complicate the task of intuiting informative experiments. See ***Box 1—figure 1*** for a schematic of the data-generating process of multi-armed bandit tasks.



Block 1

Trial	Action	Reward
1	1	0
2	2	1
3	2	0
…	…	…

Block B

Trial	Action	Reward
1	2	1
2	2	0
3	3	1
…	…	…

Box 1—figure 1. Schematic of the data-generating process for the example of multi-armed bandit tasks. For a given design **d** (here the reward probabilities associated with the bandit arms in each individual block) and a sample from the prior over the model parameters $\boldsymbol{\theta}$, the simulator generates observed data $\boldsymbol{y}$, corresponding to the actions and rewards from B blocks.

Case study: Multi-armed bandit tasks

As a case study application, we consider the question of how people balance their pursuit of short-term reward ('exploitation') against learning how to maximize reward in the longer term ('exploration'). This question has been studied at length in psychology, neuroscience, and computer science, and one of the key frameworks for investigating it is the multi-armed bandit decision-making task (e.g. ***Steyvers et al., 2009***; ***Dezfouli et al., 2019***; ***Schulz et al., 2020***; ***Gershman, 2018***). Multi-armed bandits have a long history in statistics and machine learning (e.g. ***Robbins, 1952***; ***Sutton and Barto, 2018***) and formalize a general class of sequential decision problems—repeatedly choosing between a set of options under uncertainty with the goal of maximizing cumulative reward (for more background, see ***Box 1***).

We demonstrate the applicability of our method by optimizing the design of a multi-armed bandit decision-making task, considering three flexible extensions of previously proposed models of human behavior. In our experiments, we consider two general scientific tasks: model discrimination and parameter estimation. We then evaluate the optimal designs found using our method empirically with simulations and real human behavioral data collected from online experiments, and compare them to designs commonly used in the literature. We find that, as compared to commonly used designs, our optimal experimental designs yield significantly better model recovery, more informative posterior distributions and improved model parameter disentanglement.

Workflow summary

In the following sections, we describe and walk through our workflow step-by-step, using the aforementioned multi-armed bandit task setting as a concrete case study to further illustrate ML-based BOED. At a high level, our workflow comprises the following steps:

1. Defining a scientific goal, for example, which model, among a set of models, best describes a natural phenomenon.
2. Formalizing a theory, or theories, as a computational model(s) that can be sampled from.
3. Setting up the design optimization problem and deciding which aspects of the experimental design need to be optimized.
4. Constructing the required ML models and training them with simulated data.
5. Validating the obtained optimal designs in silico and, potentially, re-evaluating design and modeling choices.
6. After confirming and validating the optimal design, the real experiment can be performed. The trained ML models used for BOED can be used afterwards to easily compute posterior distributions.

We provide a more detailed description of the above workflow in the following sections. The outputs of the ML-based BOED method are optimal experimental designs, an estimate of the (maximum) expected information gain and an amortized posterior distribution, which allows for straightforward computation of posterior distributions from collections of real-world data and saves us a potentially computationally expensive (likelihood-free) inference step. Readers may find that many of these steps should be part of the design of experiments no matter whether using BOED or not. However, BOED here provides the additional advantage of making all of these steps, which are sometimes carried out heuristically, explicit and formal.

Step 1: Formulate your scientific goal

The first step in setting up the experimental design optimization is to define our scientific goal. Although there may be many different goals for an experiment, they often fall into the broad categories of model discrimination or parameter estimation.

Model discrimination

Computational models provide formal and testable implementations of theories about nature. Model discrimination (also often referred to as *model comparison*) lies at the core of many scientific questions and amounts to the problem of deciding which of a set of different models best explains some observed data. A rigorous formulation of this problem is given by the Bayesian approach: Starting from a prior belief about the plausibility of different models and their parameters, we consider how well different models can explain the observed data and use this information to update our prior beliefs. Crucially, more flexible theories are penalized automatically via the notion of the Bayesian Ockham's razor. This ensures that in cases where two theories can explain the data equally well, we should favor the simpler one. Formally, this requires computing the marginal likelihood of our different models, which may mean integrating (or for discrete parameters, summing) over a large number of model parameters, which is often intractable, and thus often tackled via information criteria, which may, however, suffer from various problems (e.g. ***Pitt and Myung, 2002***; ***Dziak et al., 2020***; ***Schad et al., 2021***). This computational complexity of computing marginal model likelihoods is exacerbated when model parameter likelihood is intractable, as is the case for many simulator models.

Box 2. Models of human behavior in bandit tasks.

We study models of human choice behavior in bandit tasks to demonstrate and validate our method. We generalize previously proposed models (see Appendix 3 for details), and consider three novel computational models for which likelihoods are intractable.

Win-Stay Lose-Thompson-Sample (WSLTS)

Posits that people tend to repeat choices that resulted in a reward, and otherwise to explore by selecting options with a probability proportional to their posterior probability of being the best (Thompson sampling). This model generalizes the Win-Stay Lose-Shift model that has been considered in previous work (***Steyvers et al., 2009***), but accommodates the possibility that people might be more likely to shift to more promising alternatives than uniformly at random.

Auto-regressive ε-Greedy (AEG)

Formalizes the idea that people might greedily choose the option with the highest estimated reward in each trial with a certain probability or otherwise explore randomly, but also permits that people may be 'sticky', that is, having some tendency to re-select whatever option was chosen on the previous trial, or 'anti-sticky', that is, preferring to switch to a new option on each trial. This generalizes the ε-greedy model studied previously (***Steyvers et al., 2009***), further accommodating the possibility that people preferentially stick to or avoid their previous selection.

Generalized Latent State (GLS)

Captures the idea that people might have an internal latent exploration-exploitation state that influences their choices when encountering situations in which they are faced with an explore-exploit dilemma. Transitions between these states may depend on previously encountered rewards and the previous latent state. This generalizes previous latent-state models that assume people shift between states independently of previous states and observed rewards ***Zhang et al., 2009***, or only once per task (***Lee et al., 2011***). As these models are novel and to illustrate the method, we do not specify strong prior beliefs about any model parameters and describe the priors distribution in Appendix 1.

With respect to our case study, we may, for instance, be interested in comparing which of the three simulator models presented in ***Box 2*** best explains human behavior in multi-armed bandit tasks.

Parameter estimation

Researchers are often interested in learning about model parameters of a particular, single model. Similar to the model discrimination approach, a Bayesian approach of learning about these parameters would involve starting with a prior distribution and then updating our belief via Bayes' rule using observed data. For instance, we may be interested in estimating an individual's (or population's) working memory capacity given a working memory model, response time distribution in a model for selective attention, choice probability of choosing between different options in a risky choice model, inclination to explore versus exploit in a reinforcement task, or associations between biases in probabilistic visual integration tasks and autistic/schizotypal traits (***Karvelis et al., 2018***) to name a few examples. Model parameters may here be continuous or discrete, which matters for how we set up up the optimization procedure in Step 4.

As an example, for the Auto-regressive ϵ-Greedy (AEG) simulator model in our case study, we are, among many other things, interested in estimating peoples' inclination to re-select previously chosen options as opposed to avoiding repeating their own past behavior, which is characterized by a specific model parameter.

Other goals

While we here focus on the problems of model discrimination and parameter estimation, we point out that the overall BOED approach via mutual information (MI) estimation is more general, and can easily be adapted to other tasks, such as improving future predictions (e.g. ***Kleinegesse and Gutmann, 2021***) and many others (see ***Ryan et al., 2016***).

How do we measure the value of an experimental design?

We formalize the quality of our experimental design via a utility function, which thereby serves as a quantitative way of measuring the value of one experimental design over any another and provides the objective function for our optimization problem. Following recent work in BOED (e.g. ***Kleinegesse and Gutmann, 2019***; ***Kleinegesse et al., 2020***; ***Kleinegesse and Gutmann, 2021***; ***Foster et al., 2019***), we use MI (***Lindley, 1956***; also known as the *expected information gain*) as our utility function $U(\mathbf{d})$ for measuring the value of an experimental design $\mathbf{d}$, that is

$$U(\mathbf{d}) = \mathbb{E}_{p(\mathbf{y}|\boldsymbol{v},\mathbf{d})p(\boldsymbol{v})}\left[\log\frac{p(\boldsymbol{v}|\mathbf{y},\mathbf{d})}{p(\boldsymbol{v})}\right], \tag{1}$$

where $\boldsymbol{v}$ is a variable we wish to estimate and $\boldsymbol{y}$ corresponds to the observed data (actions and rewards in our case study). Note that the probability density function $p(\boldsymbol{v}|\mathbf{y},\mathbf{d})$ is the posterior distribution and $p(\boldsymbol{v})$ is the prior distribution of the variable of interest.

For the model discrimination task, the variable of interest $\boldsymbol{v}$ in our utility function corresponds to a scalar model indicator variable m (where $m = 1$ corresponds to a model 1, $m = 2$ to some model 2 and so on for as many models as we are comparing). For the parameter estimation task, the variable of interest $\boldsymbol{v}$ in our utility function is the vector of model parameters $\boldsymbol{\theta}_m$ for a given model m.

MI has several appealing properties (***Paninski, 2005***) that make it a popular utility function in BOED (***Ryan et al., 2016***). Intuitively, MI quantifies the amount of information our experiment is expected to provide about the variable of interest. Furthermore, the MI utility function can easily be adapted to various scientific goals by changing the variable of interest (***Kleinegesse and Gutmann, 2021***). Optimal designs are then found by maximizing this utility function, that is $\mathbf{d}^* = \arg\max U(\mathbf{d})$. Although a useful quantity, estimating $U(\mathbf{d})$ is typically extremely difficult, especially for simulator models that have intractable likelihoods $p(\mathbf{y}|\boldsymbol{v},\mathbf{d})$ (e.g. ***Ryan et al., 2016***). Nonetheless, in the later steps of our workflow, we shall explain how we can approximate this utility function using ML-based approaches.

Step 2: Cast your theory as a computational model

After defining our scientific goal, we need to ensure that our scientific theories are, or can be, formalized as computational models. This either requires us to formalize them from the ground up, or leverage existing computational models from literature that support our theories. In the former case, this entails making all aspects of the theory explicit, which is therefore a highly useful and established practice; we refer the reader to ***Wilson and Collins, 2019*** for a more general guidance on building computational models. For the presented ML-based BOED approach, we require that our theories are cast as computational models that allow us to simulate data, that is simulator models. Further we need to decide on a general experimental paradigm that needs to be able to interact with our computational models.

Building a computational model

Formally, a computational model defines a generative model $\mathbf{y} \sim p(\mathbf{y}|\boldsymbol{\theta},\mathbf{d})$ that allows sampling of synthetic data $\mathbf{y}$, given values of its model parameters $\boldsymbol{\theta}$ and the experimental design $\mathbf{d}$. We assume that sampling from this model is feasible, but make no assumptions about whether evaluating the likelihood function $p(\mathbf{y}|\boldsymbol{\theta},\mathbf{d})$ is tractable or intractable. Even if the likelihood is tractable and simple, important quantities such as the marginal likelihood may nevertheless be intractable, since evaluating the marginal likelihood requires integrating over all free parameters. This problem is naturally exacerbated for more complex models that may have intractable likelihood functions.

See ***Box 2*** for a set of simulator models from our case study. These models describe human behavior in a multi-armed bandit setting, where the observed data $\boldsymbol{y}$ corresponds to the sequences of chosen bandit arms along with their observed reward.

What are our prior beliefs?

In addition to formalizing our theories as simulator models, we need to complete this step by defining prior distributions over the (free) model parameters, as would be required for any analysis. We point the reader to ***Wilson and Collins, 2019***; ***Schad et al., 2021***; ***Mikkola et al., 2021*** for guidance on specifying prior distributions, as this is a broad research topic and a detailed treatment is beyond the scope of this work. Generally, prior distributions should reflect what is known about the model parameters prior to running the experiment, based on either empirical data or general domain knowledge. This is crucial for the design of experiments, as we may otherwise spend experimental resources and effort on learning what is already known about our models.

Step 3: Decide which aspect of your experiment to optimize

Having formalized our scientific theories as simulator models and decided on our scientific goals, we need to think about which parts of the experiment we want to optimize. In principle, any controllable aspect of the experimental apparatus could be tuned as part of the experimental design optimization. Concretely, in order to solve an experimental design problem by means of BOED, our experimental designs $\mathbf{d}$ need to be formalized as part of the simulator model $p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{d})$. Instead of optimizing all controllable aspects of an experiment, it may then help to leverage domain knowledge and prior work to choose a sufficient experimental design set, in order to reduce the dimensionality of the design problem.

In our case study, for instance, we have chosen the reward probabilities of different bandit arms to be the experimental designs to optimize over. This choice posits that certain reward environments yield more informative data than others, in the context of model discrimination and parameter estimation. Our task is then to find the optimal reward environment that yields maximally informative data.

There may be other experimental choices that we will have to make, in particular to limit variability in the observed data, to control for confounding effects, and to reduce non-stationarity, among many other reasons. In our multi-armed bandit setting, for example, we may choose to limit the number of trials in the experiment to avoid participant fatigue, which is usually not accounted for in computational models of behavioral phenomena. Additionally, we may limit the number of available choices in a discrete choice task or impose constraints on any other aspects suggested by domain knowledge, or where we want to ensure comparability with prior research. For example, we may have reason to believe that certain designs are more ecologically valid than others—we discuss such considerations in the *Caveats* Section.

Static and adaptive designs

We consider the common case of *static* BOED, with the possibility of having multiple blocks of trials. That is, we find a set of optimal designs prior to actually performing the experiment. In some cases, however, a scientist may wish to optimize the designs sequentially as the participant progresses through the experiment. This is called *sequential*, or *adaptive*, BOED (see ***Ryan et al., 2016***; ***Rainforth et al., 2023***) and has been explored previously for tractable models of human behavior (***Myung et al., 2013***), or in the domain of adaptive testing (e.g. ***Weiss and Kingsbury, 1984***). More recently, sequential BOED has also been extended to deal with implicit models (***Kleinegesse et al., 2020***; ***Ivanova et al., 2021***). At this point, sequential BOED is an active area of research and may be a consideration for researchers in the future. Generally, sequential designs can be more efficient, as they allow to adapt the experiment 'on-the-fly'. However, sequential designs also make building the actual experiment more complicated, because the experiment needs to be updated in almost real-time to provide a seamless experience. Further, in some settings, participants may notice that the experiment adapts to their actions, which may inadvertently alter their behavior. As we discuss below in the *Caveats* Section, the effects of model misspecification may be especially severe with sequential designs, contributing to our focus on static designs. For future applications, which approach to follow is likely going to be a choice that uniquely depends on the experiment and research questions at hand, while the overall workflow outlined in this tutorial will be relevant to both (***Box 3***).

Step 4: Use machine learning to design the experiments

As previously explained, MI shown in ***Equation 1*** is an effective means of measuring the value (i.e. the *utility*) of an experimental design setting $\mathbf{d}$ in BOED (***Paninski, 2005***; ***Ryan et al., 2016***). Unfortunately, it is generally an expensive, or intractable, quantity to compute, especially for simulator models. The reason for this is twofold: Firstly, the MI is defined via a high-dimensional integral. Secondly, the MI

Box 3. Experimental settings in the case study.

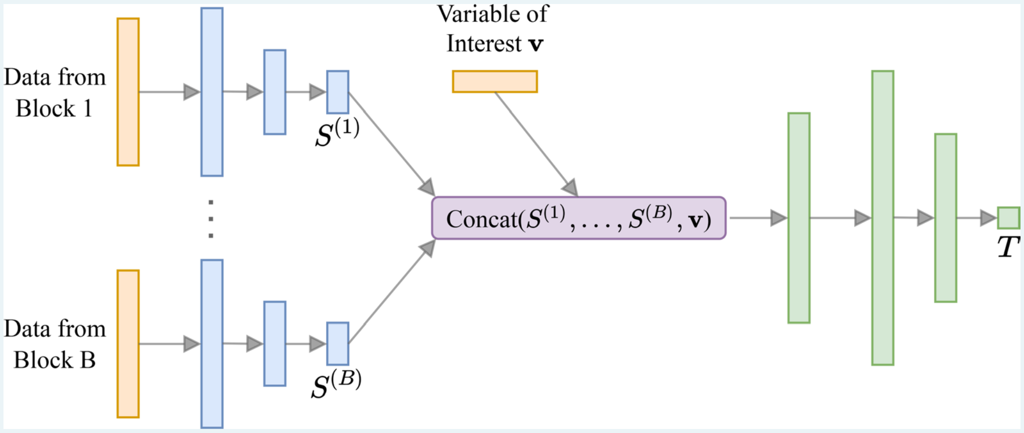
Behavioral experiments are often set up with several experimental blocks. We here make the common assumption of exchangeability with respect to the experimental blocks in our analyses and randomize block order in all experiments involving human participants. In our experiments, the bandit task in each experimental block has 3 choice options (bandit arms) and 30 trials (sequential choices). For each block of trials, the data $\mathbf{y}$ consists of 30 actions and 30 rewards, leading to a dimensionality of 60 per block. For a given block, the experimental design $\mathbf{d}$ we wish to optimize are the reward probabilities of the multi-armed bandit, which, in our setting, are three-dimensional (corresponding to independent scalar Bernoulli reward probabilities for each bandit arm). The generative process for models of human behavior in bandit tasks is illustrated in ***Box 2***. We demonstrate the optimization of reward probabilities for multi-armed bandit tasks, with the scientific goals of *model discrimination* and *parameter estimation*. We consider five blocks of 30 trials per participant, yielding a fairly short but still informative experiment. We divide this into two blocks for model discrimination followed by three blocks for parameter estimation, where the parameter estimation blocks are conditional on the winning model for the participant in the first two blocks. In other words, for the parameter estimation task, a different experimental design (which corresponds to a particular setting of the reward probabilities associated with each bandit arm) is presented to a participant depending on the simulator model that best described them in the first two blocks. Since we consider three-armed bandits, the design space is constituted by the possible combinations of reward probabilities. This is six-dimensional for model discrimination and nine-dimensional for PE. Similarly, the data for a single block is 60-dimensional, so the dimensionality of the data is 120 for model discrimination and 180 for PE. Our method applies to any prior distribution over model parameters or the model indicator, but for demonstration purposes, we assume uninformative prior distributions on all model parameters (see Appendix 1 for details). As our baseline designs, we sample all reward probabilities from a $\mathrm{Beta}(2,2)$ distribution, following a large and well-known experiment on bandit problems with 451 participants (***Steyvers et al., 2009***). More information about the experimental settings and detailed descriptions of the algorithms can be found in the Appendix.

requires density evaluations of either the posterior distribution or the marginal likelihood, both of which are expensive to compute for general statistical models, and intractable for simulator models. Recent advances in BOED for simulator models thus advocate the use of cheaper MI lower bounds instead (e.g. ***Foster et al., 2019***; ***Kleinegesse et al., 2020***, ***Kleinegesse and Gutmann, 2021***; ***Ivanova et al., 2021***).

Many recent innovations in BOED rely on ML to accurately and efficiently estimate functions. ML is currently revolutionizing large parts of the natural sciences, with applications ranging from understanding the spread of viruses (***Currie et al., 2020***) to discovering new molecules (***Jumper et al., 2021***), forecasting climate change (***Runge et al., 2019***), and developing new theories of human decision-making (***Peterson et al., 2021***). Due to its versatility, ML can be integrated effectively into a wide range of scientific workflows, facilitating theory building, data modeling and analysis. In particular, ML allows us to automatically discover patterns in high-dimensional data sets. Many recent state-of-the-art methods in natural science, therefore, deeply integrate ML into their data analysis pipelines (e.g. ***Blei and Smyth, 2017***). However, there has been little focus on applying ML to improve the data collection process, which ultimately determines the quality of data, and thus the efficacy of downstream analyses and inferences. In this work, we use ML methods as part of our optimization scheme, where they serve as flexible function approximators.

The step of training a good ML model currently requires some previous expertise with (or willingness to learn about) applying ML methods. Fortunately, this step is less experiment-specific than it may first appear. Many different experiments generate similar data, and the architecture we provide can deal with multiple blocks and automatically learns summary statistics, as we explain above. Many excellent resources on the basics of ML exist, and a detailed treatment is beyond the scope of this

Box 4. Bespoke neural network architecture for behavioral experiments.


Data from Block 1
$S^{(1)}$
Data from Block B
$S^{(B)}$
Variable of Interest **v**
Concat($S^{(1)}, \ldots, S^{(B)}, \mathbf{v}$)
T

Box 4—figure 1. Neural network architecture for behavioral experiments. For each block of data we have a small sub-network (shown in blue) that outputs summary statistics S. These are concatenated with the variable of interest $\boldsymbol{v}$ (e.g. corresponding to a model indicator for model discrimination tasks) and passed to a larger neural network (shown in green). ***Box 4—figure 1*** has been adapted from **Figure 1** in ***Valentin et al., 2021***.

tutorial. However, we provide pointers to several aspects in the code repository: https://github.com/simonvalentin/boed-tutorial (copy archived at ***Valentin, 2023***).

Approximating mutual information with machine learning

In this work, we leverage the MINEBED method (***Kleinegesse et al., 2020***), which allows us to efficiently estimate the MI using ML. Specifically, for a particular design $\mathbf{d}$, we first simulate data from the computational models under consideration, (shown in Box ***Box 2***). The approach works by constructing a lower bound on the MI that is parameterized via an ML model. In our case study, we use a neural network with an architecture customized for behavioral experiments (shown in ***Box 4***), but any other ML model would work in place. We then train the ML model by using the simulated data as input and the lower bound as an objective function, which gradually tightens the lower bound.

More concretely, this method works by training a neural network $T_\psi(\boldsymbol{v}, \mathbf{y})$, or any other ML model, where ψ are the neural network parameters and the data $\mathbf{y}$ is simulated at design $\mathbf{d}$ with samples from the prior $p(\boldsymbol{v})$. We train this neural network by maximizing the following objective function, which is a *lower bound* of the MI,

$$U(\mathbf{d}; \psi) = \mathbb{E}_{p(\mathbf{y}|\boldsymbol{v},\mathbf{d})p(\boldsymbol{v})}\left[T_\psi(\boldsymbol{v}, \mathbf{y})\right] - \mathbb{E}_{p(\mathbf{y}|\mathbf{d})p(\boldsymbol{v})}\left[e^{T_\psi(\boldsymbol{v},\mathbf{y})-1}\right] \quad (2)$$

The above lower bound is also known as the NWJ lower bound and has several appealing bias-variance properties (***Poole et al., 2019***; we note that alternative bounds on MI can be used in the place of the NWJ bound as our loss function, as discussed by ***Kleinegesse and Gutmann, 2021***). The expectations in ***Equation 2*** are usually approximated using (Monte-Carlo) sample-averages.

Once we have trained the neural network $T_\psi(\boldsymbol{v}, \mathbf{y})$ on ψ, we estimate the MI at $\mathbf{d}$ by computing ***Equation 2*** using a held-out test set of data simulated from the computational model(s).

Dealing with high-dimensional observations

We can efficiently deal with high-dimensional data and do not have to construct (approximately sufficient) summary statistics based on domain expertise, as is commonly done (***Myung and Pitt, 2009***; ***Zhang and Lee, 2010***; ***Ouyang et al., 2018***), and which can be prohibitively difficult for complex simulator models (*Chen et al., 2021*). In fact, the previously described method can allow us to automatically learn summary statistics as a by-product, as explained below.

Critical to the performance of our method in our case study, as with any application of neural networks, is the choice of architecture (see *Elsken et al., 2019* and *Ren et al., 2021* for recent reviews of the challenges and solutions in neural architecture search). In order to develop an effective neural network architecture, we leverage the method of neural approximate sufficient statistics (*Chen et al., 2021*) and propose them as a novel and practical alternative to hand-crafted summary statistics in the context of BOED. Behavioral data are often collected in several blocks that can have different designs but should be analyzed jointly, which can be challenging due to the high dimensionality of the joint data, which typically comprises numerous choices, judgments or other measurements (including high-resolution and/or high-frequency neuroimaging data). In order to deal with a multi-block data structure effectively, we propose an architecture devised specifically for application to behavioral experiments. Our architecture, visualized in *Box 4—figure 1*, incorporates a sub-network for each block of an experiment. The outputs of these sub-networks are then concatenated and passed as an input to a larger neural network. In doing so, besides being able to reduce the dimensionality of the input data, each sub-network conveniently learns to approximate sufficient summary statistics of the data from each block in the experiment (*Chen et al., 2021*).

In this work, we develop a bespoke feed-forward neural network architecture, which allows us to efficiently deal with multiple blocks of data. The code we are providing is straightforwardly adaptable to other problems 'out of the box', as the architecture can scale to multiple blocks of data and realistic complexity (dimensionality) of the observed data. However, the overall approach is not restricted to this particular ML method; in fact, any sufficiently flexible ML model may be used, as long as the loss function can be specified. In some settings, researchers may be interested in using ML models that only require little tuning, such as tree-based ensemble methods like Random Forests (*Breiman, 2001*) or gradient boosted trees (*Friedman, 2001*). On the other hand, when dealing with high-dimensional but structured data, such as in eye-tracking or neuroimaging (or combinations of different types of data, see *Turner et al., 2017*), one may use other appropriate architectures, such as convolutional (*LeCun et al., 2015*), recurrent (*Rumelhart et al., 1986*), transformers, (*Vaswani et al., 2017*), or graph neural networks (*Zhou et al., 2020*). For further pointers to practical recommendations, see the code repository: https://github.com/simonvalentin/boed-tutorial, (copy archived at *Valentin, 2023*).

Search over the space of possible experimental designs

Once we have obtained an estimate of $U(\mathbf{d})$ at a candidate design $\mathbf{d}$, we need to update the design appropriately when searching the design space to solve the experimental design problem $\mathbf{d}* = \arg\max_{\mathbf{d}} U(\mathbf{d})$. Unfortunately, in the setting of discrete observed data $\mathbf{y}$ (such as choices between a set of options), gradients with respect to designs are generally unavailable, requiring the use of gradient-free optimization techniques. We here optimize the design $\mathbf{d}$ by means of Bayesian optimization (BO; see *Shahriari et al., 2015*, for a review), which has been used effectively in the experimental design context before (e.g. *Martinez-Cantin, 2014*; *Kleinegesse et al., 2020*). We specifically use a Gaussian Process (GP) as our probabilistic surrogate model with a Matérn-5/2 kernel and Expected Improvement as the acquisition function (these are standard choices). A formal summary of our BOED approach is shown in Appendix 4.

For higher dimensional continuous design spaces, this search may become difficult and necessitate the use of more scalable BO variants (e.g. *Overstall and Woods, 2017*; *Oh et al., 2018*; *Eduardo and Gutmann, 2023*). For discrete design spaces or combinations of discrete and continuous design spaces, one can straightforwardly parallelize the optimization for each level of the discrete variable. If the observed data are continuous, such as participants' response times, and one is able to compute gradients with respect to the designs (i.e. the simulator is differentiable), the design space may also be explored with gradient-based optimization (*Kleinegesse and Gutmann, 2021*). In some settings, we may have some non-trivial constraints on permissible experimental designs, and these can be enforced as part of the search, making sure that impermissible designs are rejected and/or not explored. Thus, if we want our designs to have certain properties, we need to encode these constraints. For example, in our bandit tasks, we could enforce a certain minimum degree of randomness associated with each bandit arm's reward probability (although we do not do so in the present work), such that each reward probability is within in a specified range.

Step 5: Validate the optimal experimental design in silico

Once we have converged in our search over experimental designs, we can simulate the behavior of our theories under our optimal experimental design(s). Irrespective of the final experiment that will be run, model simulations can already provide useful information for theory building, for example by noticing that the simulator models generate implausible data for certain experimental designs. This means that we can use simulations to assess whether we can recover our models, or individual model parameters, given the experimental design, as we did in our simulation study. We highlight this again as a critical step, since issues at this stage warrant revising the models under consideration and, importantly, any corresponding real-world experiment would be expected to yield uninformative data and waste resources (***Wilson and Collins, 2019***).

As we discuss in the *Caveats* Section, BOED may sometimes lead to unusual yet effective experimental designs, and we suggest that surprising optimal designs may deserve careful consideration, as they may reveal subtle differences in predicted behaviors or model misspecifications. One appeal of optimal designs is that some of the ways of obtaining information about which models is correct correspond to finding situations in which one of them makes extreme predictions. BOED can thereby be seen as facilitating good prior predictive checking. We have illustrated this step using our case study in ***Box 5*** and ***Box 6***, where we present simulation study results for model discrimination and parameter estimation, respectively.

Step 6: Run the real experiment

Once we are satisfied with the found optimal experimental design, we can run the actual experiment. Here, the ML models we used to approximate MI allow us to cheaply perform posterior inference. After running the real-world experiment, we can use our trained ML models to cheaply compute approximate posterior distributions over models and their parameters. This procedure provides a valuable addition to a Bayesian modeling workflow ***Schad et al., 2021***; ***Wilson and Collins, 2019***, as computing posterior distributions is a computationally expensive step, which is especially difficult for models with intractable likelihoods. We explain theoretical details below, and showcase an analysis of a real experiment with our case study in ***Box 7***, ***Box 8***, ***Box 9***.

Posterior estimation

Once the neural network is trained, we can conveniently compute an approximate posterior distribution with a single forward-pass (see ***Kleinegesse et al., 2020***, for a derivation and further explanations). The NWJ lower bound is tight when $T_{\psi^*}(\boldsymbol{v}, \mathbf{y}) = 1 + \log p(\boldsymbol{v}|\mathbf{y}, \mathbf{d}^*)/p(\boldsymbol{v})$. By rearranging this, we can thus use our trained neural network $T_{\psi^*}(\boldsymbol{v}, \mathbf{y})$ to compute a (normalized) estimate of the posterior distribution, This can be done with the following equation,

$$p(\boldsymbol{v}|,\ ^*) = p(\boldsymbol{v})e^{T_*(\boldsymbol{v},)-1}. \tag{3}$$

Neural networks are notorious for large variations in performance, mainly due to the random initialization of network parameters. The quality of the posterior distribution estimated using ***Equation 3*** is therefore quite sensitive to the initialization of the neural network parameters. To overcome this limitation and to obtain a more robust estimate of the posterior distribution, we suggest to use *ensemble learning* to compute ***Equation 3***. This is done by training several neural networks (50 in our case study) and then averaging estimates of the posteriors obtained from each trained neural network.

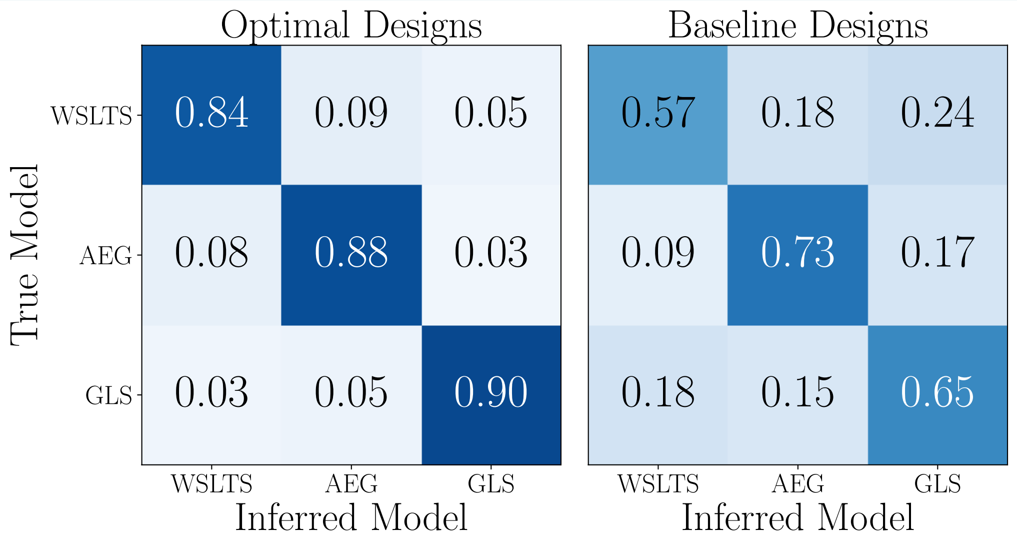
Summary statistics

Additionally, as another component of our approach, we also obtain automatically-learned (approximate) sufficient summary statistics for data observed with the optimal experimental designs. Summary statistics are often crafted manually by domain experts, which tends to be a time-consuming and difficult task. In fact, hand-crafted summary statistics are becoming less effective and informative as our models naturally become more complex as science advances (***Chen et al., 2021***). Our learned summary statistics could be used for various downstream tasks, for example as an input for approximate inference techniques (such as ABC) or as means to interpret and analyze the data more effectively, for example as a lower dimensional representation used to visualize the data.

Box 5. Model discrimination.

Simulation study Our method reveals that the optimal reward probabilities for the model discrimination task are, approximately, $[0, 0, 0.6]$ for one block of trials and $[1, 1, 0]$ for the other block. These optimal designs stand in stark contrast to the usual reward probabilities used in such behavioral experiments, which characteristically use less extreme values (***Steyvers et al., 2009***). Contrary to our initial intuitions, and, presumably those of previous experiment designers, the extreme probabilities in these optimal designs are effective at distinguishing between particular theoretical commitments of our different models. For instance, the AEG model (see ***Box 2***) can break ties between options with equivalent (observed) reward rates via the 'stickiness' parameter. To illustrate this effect, consider the bandit with $[1, 1, 0]$ reward probabilities, as found above. Here, switching between the two options that always produce a reward is different from switching to the option that never produces a reward. Specifically, switching between the winning options can be seen as 'anti-stickiness', motivated for example, out of boredom or epistemic curiosity about subtle differences in the true reward rates of the winning options. This type of strategy is less apparent when the bandit arms are stochastic. As such, these mechanisms can be investigated most effectively by not introducing additional variability due to stochastic rewards and instead isolating their effects. Hence, we can interpret the, perhaps counter-intuitive, optimal experimental designs as effective choices for isolating distinctive mechanisms of our behavioral models or their parameters under the assumed priors.

Using the neural network that was trained at that optimal design, we can cheaply compute approximate posterior distributions of the model indicator (as described in Materials and methods). This yields the confusion matrices in ***Box 5—figure 1***, which show that our optimal designs result in considerably better model recovery than the baseline designs.



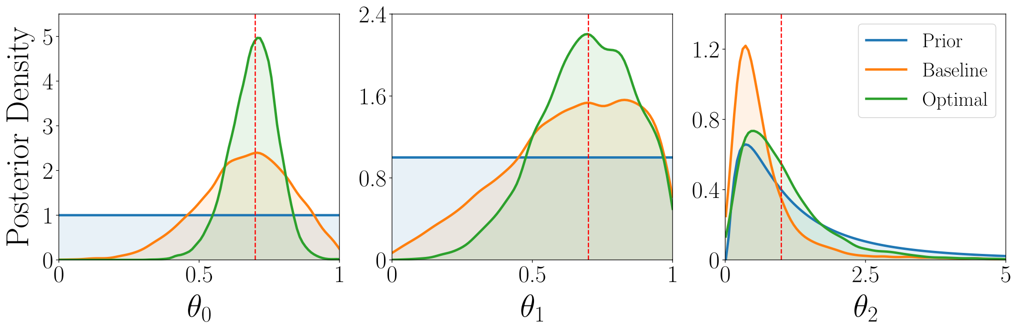
Box 5—figure 1. Simulation study results for the model discrimination task, showing the confusion matrices of the inferred behavioral models, for optimal (left) and baseline (right) designs. ***Box 5—figure 1*** has been adapted from **Figure 2** in ***Valentin et al., 2021***.

Model checking

As also emphasized by previous works on computational modeling, a model-based data analysis (such as via the amortized posterior distributions discussed above) should go hand-in-hand with model checking, also referred to as validation (e.g. ***Wilson and Collins, 2019***; ***Palminteri et al., 2017***). A detailed treatment of methods and techniques for model checking is beyond the scope of this paper.

Box 6. Parameter estimation: Simulation study.

We next discuss the parameter estimation results, focusing on the WSLTS model; the results for the AEG and GLS model can be found in Appendix 5. We find that the optimal reward probabilities for the WSLTS model are [0, 1, 0], [0, 1, 1] and [1, 0, 1] for the three blocks. Similar to the model discrimination task, these optimal designs take extreme values, unlike commonly used reward probabilities in the literature. In ***Box 6—figure 1*** we show posterior distributions of the WSLTS model parameters for optimal and baseline designs. We find that our optimal designs yield data that result in considerably improved parameter recovery, as compared to baseline designs.



Box 6—figure 1. Simulation study results for the parameter estimation task of the WSLTS model, showing the marginal posterior distributions of the three WSLTS model parameters for optimal (green) and baseline (orange) designs, averaged over 1,000 simulated observations. The ground-truth parameter values were chosen in accordance with previous work on simpler versions of the WSLTS model (***Zhang and Lee, 2010***) and as plausible population values for the posterior reshaping parameter.

However, it should be noted that model checking is just as important with data collected through optimized experimental designs as through hand-crafted designs.

Caveats

We now turn to more high-level considerations and discuss several categories of potential issues and caveats related to BOED. We first discuss how the optimality of an experimental design is always defined with respect to a specific goal. Second, we consider how posterior inferences always depend on the experimental design and how this impacts BOED considerations. Third, we bring attention to the important issue of model misspecification in the context of BOED. Fourth, we discuss the topic of ecological validity and generalizability of our inferences with BOED. Finally, we cover the interpretability of our optimal designs. While we discuss these issues in light of BOED, they are just as important with hand-crafted experiments. Importantly, however, BOED makes these issues more salient, by turning the design of experiments into a more explicit process. Therefore, these points should be relevant to any researcher collecting or analyzing experimental data using computational models.

No single design to rule them all

The process of making all of our assumptions explicit and leaving the process of proposing concrete designs to the optimization procedure can provide us with assurance that our design is optimal with respect to our assumptions. Therefore, this process may also reduce bias in the design of experiments, as the optimization process can easily be replicated. We believe that these properties will help tackle the notorious *replication crisis* (***Ioannidis, 2005***; ***Pashler and Wagenmakers, 2012***; ***Baker, 2016***) and potentially setting new standards.

However, it is important to point out that the designs found using our optimization procedure are not assured to be optimal for all scenarios. That is, there is no 'free lunch' (***Wolpert, 1996***): No

Box 7. Human participant study: Methodology.

To validate our method on empirical data, we collected a sample of $N = 326$ participants from Prolific (https://www.prolific.com/; all experiments were certified according to the University of Edinburgh Informatics Research Ethics Process, RT number 2019/58792). Participants were randomly allocated to our optimal designs or to one of ten baseline designs (the same as in the simulation study); we refer to the former as the *optimal* design group and to the latter as the *baseline* group. The first two blocks correspond to the model discrimination task, that is they were used to identify the model that matches a participant's behavior best. This was done by computing the posterior distribution over the model indicator m using the trained neural networks (see Appendix 1) and then selecting the model with the highest posterior probability, that is the *maximum* a posteriori estimate. This inference process was done online without a noticeable interruption to the experiment. The following three blocks were then used for the parameter estimation task. Participants in the optimal design group were allocated the optimal design according to their inferred model (as provided by the model discrimination task). Participants in the baseline group were again allocated to baseline designs. This data collection process facilitates gathering real-world data for both the model discrimination and parameter estimation tasks, allowing us to effectively compare our optimal designs to baseline designs. See Appendix 2 for more details about the human participant study setup.

Here, we follow the rationale that more useful experimental designs are those that lead to more information (or lower posterior uncertainty) about the variable of interest. The belief about our variable of interest is provided by its posterior distribution, which we can estimate using the neural network output. We assess the quality of these posterior distributions by estimating their entropy, which quantifies the amount of *information* they encode (***Shannon, 1948***). This is an effective and prominent metric to evaluate distributions, as it directly relates to the uncertainty about the variable of interest. For the model discrimination task, we specifically consider the Shannon entropy as a metric, since the model indicator m is a discrete random variable. As the continuous analogue of Shannon entropy, we use differential entropy for measuring the entropy of posterior distributions for the parameter estimation task, since the model parameters are continuous random variables. This allows us to quantitatively evaluate our optimal designs and compare their efficacy to that of baseline designs.

design suits all situations, as the usefulness of experimental designs depends on the theories under consideration, prior beliefs, and utility function. For instance, if we already have strong prior beliefs about a particular model parameter (such as people being strongly inclined to explore instead of exploit), there may be little to learn about this parameter and the BOED procedure may instead focus on parameters we have weaker prior beliefs about, which may yield different optimal designs. More generally, different scientific goals can lead to different optimal designs.

As an illustrative example, we consider the design of a survey item for political orientation. For simplicity, we assume a one-dimensional spectrum with extremes to either side of the scale being equally likely in the population, with most people falling in the middle of the scale. In one scenario the goal may be to estimate a given person's political belief along the spectrum. Here, an informative question is likely going to be moderate, since extreme items on either side of the spectrum ignore the other side, have a low prior probability and are therefore less useful. However, if the goal is to distinguish 'extremist' participants on one side of the spectrum from the rest of the population, or if the scientific question implies that a participant can only belong to either side of the one-dimensional political spectrum then polarizing items are likely more informative than moderate items that do not reveal such extremist beliefs. This emphasizes the importance of clearly defining the scientific goal, making sure that the computational models reflect the theories under consideration, as described in Steps 1 and 2.

Box 8. Human participant study: Results.

We here briefly present the results of our human participant study. For the model discrimination tasks, we find that our optimal designs yield posterior distributions that tend to have lower Shannon entropy than the baseline designs, as shown in ***Box 8—figure 1***. Our ML-based approach thus allows us to obtain more informative beliefs about which model matches real individual human behavior best, as compared to baselines.

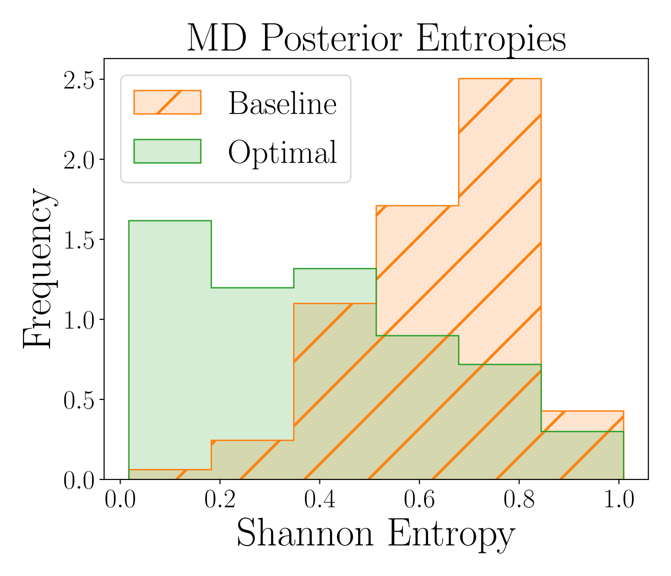
Having determined the best model to explain a given participant's behavior, we turn to the parameter estimation task, focusing our efforts on estimating the model parameters that explain their specific strategy. Similar to the model discrimination task, we here assess the quality of joint posterior distributions using entropy, where smaller values directly correspond to smaller uncertainty. ***Box 8—figure 2*** shows the distribution of differential posterior entropies across participants that were assigned to the AEG simulator model; we show corresponding plots for the WSLTS and GLS models in Appendix 5. Similar to the model discrimination task, we find that our optimal designs provide more informative data, that is they result in joint posterior distributions that tend to have lower differential entropy, as compared to the ones resulting from baseline designs, as shown in the figure.

Moreover, the proportions of participants allocated to the WSLTS, AEG, and GLS models for the optimal design were 62(37%), 75(45%) and 29(18%), respectively. For the baseline designs, the proportions were 57(36%), 22(14%) and 81(51%), for the WSLTS, AEG, and GLS models, respectively. The finding that the largest proportion of participants for the optimal designs were best described by the AEG model differs from prior work, where most participants were best described by Win-Stay Lose-Shift (a special case of our WSLTS model) ***Steyvers et al., 2009***. This suggests that a reinterpretation of previous results is required, where the human tendency to stick with previous choices is less about myopically repeating the last successful choice, and more about balancing reward maximization, exploration, and a drive to be consistent over recent runs of choices. Interestingly, we find different distributions over models for the optimal as compared to the baseline designs (A chi-square test revealed a significant difference between the distributions of WSLTS, AEG, and GLS in the Optimal and Baseline groups, $\chi^2(2) = 53.66$, p < 0.0001.). In particular, while, descriptively, the proportions of participants allocated to the WSLTS model are very close (37%, and 36% for the optimal and baseline designs, respectively), the proportions allocated to the AEG and GLS model differ more substantially. This may suggest that people's strategy depends on the experimental design, potentially with some participants favoring strategies similar to the AEG model with more reliable reward environments, and the GLS in more unreliable (baseline) environments. The models we have considered in our case study do not account for the possibility of transferring knowledge across experimental blocks. However, learning about the reward probabilities in one (or several) experimental blocks may influence people's expectations about the reward distribution of future block(s). For instance, having inferred (correctly or not), that at least one arm bandit arm always pays a reward, a learner would do well to quickly identify this arm in a future block (see, e.g. ***Wang et al., 2018***, for a related line of research). Our space of computational models does not include such strategy switching mechanisms, and we discuss the impact of such potential misspecifications below in more detail. While exploring these findings in detail from a cognitive science perspective is beyond the scope of the present work, we view this as a fruitful avenue for future work.

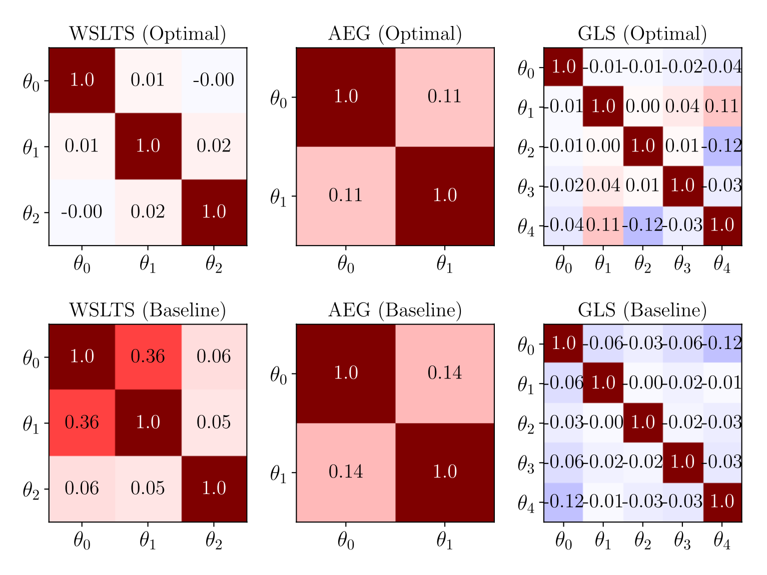
The extreme designs here may be surprising, considering the typically used stochastic rewards in bandit problems. However, these designs allow the experiment to focus on particular aspects of people's behavior, which turns out to be most informative for distinguishing between the computational models we consider. For instance, while our AEG model performs uniformly random exploration, the WSLTS model performs uncertainty-guided exploration. Meanwhile, our optimal designs raise the question of ecological validity, as it is unclear which kinds of 'reward environments' are more reflective of real-world situations. On the one hand,

continued on next page

research in causal reasoning suggests that people expect relationships to be deterministic and reliable (e.g. ***Schulz and Sommerville, 2006***). On the other hand, some environments are known to be random and more akin to a classic, stochastic bandit problem, such as the stock market. We view the question of the ecological validity of stochastic rewards as an important direction for future work, and discuss this more below in the *Caveats* Section.



Box 8—figure 1. Human-participant study results for the model discrimination (MD) task, showing thedistribution of posterior Shannon entropies obtained for optimal (green) and baseline (orange)designs (lower is better).

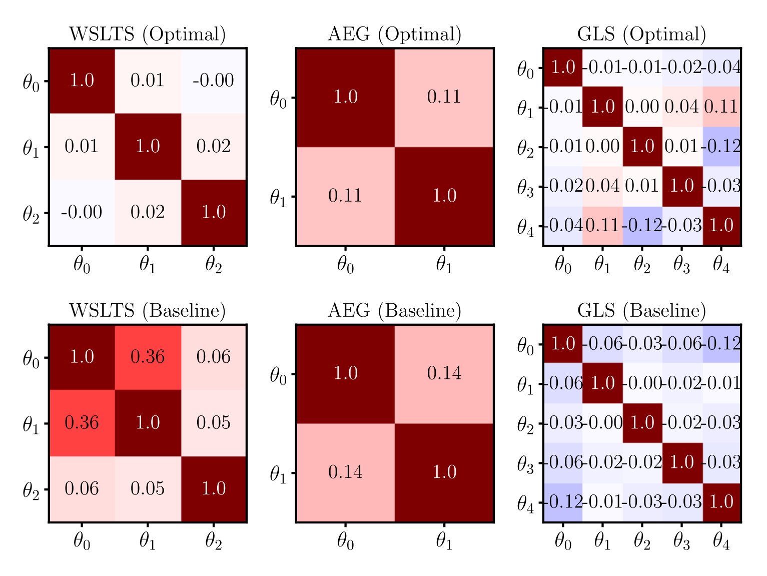


Box 8—figure 2. Human-participant study results for the parameter estimation (PE) task of the AEG model,showing the distribution of posterior differential entropies obtained for optimal (green) and baseline(orange) designs (lower is better).

Box 9. Exploring model parameter disentanglement.

In addition to measuring the efficacy of our optimal experiments via the differential entropy of posterior distributions, we can ask how well they isolate, or *disentangle*, the effects of individual model parameters. Specifically, we measure this by means of the (Pearson) correlation between model parameters in the joint posterior distribution, which we present in ***Box 9—figure 1*** for all simulator models and for both optimal designs and baseline designs. Here, a correlation of $r = 0$ means that there is no (linear) relationship between the respective pair of inferred model parameters across individuals, thereby providing evidence that the data under this design allow us to separate the natural mechanisms corresponding to those parameters. On the other hand, correlations closer to $r = -1$ or $r = 1$ would imply that the model parameters essentially encode the same mechanism. As shown descriptively in ***Box 9—figure 1***, our optimal designs are indeed able to isolate the mechanisms of individual model parameters more effectively than baseline designs. Note that these are not confusion matrices, as displayed in ***Box 5—figure 1***, but rather average correlation matrices of the model parameter posterior distributions.

To assess this observation quantitatively, we first z-transform (***Fisher, 1915***) the correlation coefficients (i.e. all entries below the main diagonal in the correlation matrices) and compute the average absolute z-score over the model parameters within each participant. The resulting average of absolute z-scores is an expression for the average magnitude of linear dependency between the posterior model parameters for each participant. Next, we assess whether the average absolute z-scores are significantly different between those participants allocated to our optimal designs and those allocated to the baseline designs. We compute this statistical significance by conducting two-sided Welch's t-tests (***Welch, 1947***). We find that our optimal designs significantly outperform baseline designs for all three models, that is for the WSLTS model we find a p-value of $p < 0.001$ ($t(60.12) = 13.89$), for the AEG model we have $p = 0.004$ ($t(27.30) = 3.12$) and for the GLS model we have $p < 0.001$ ($t(91.01) = 8.44$). This provides additional evidence that our optimal designs are better able to disentangle the effects of individual model parameters than baseline designs. As explained earlier, we argue that this isolation of effects is facilitated by the extreme values of our optimal designs.



Box 9—figure 1. Human-participant study average linear correlations in the posterior distribution of model parameters for all models in the parameter estimation task, and for both optimal (top) and baseline (bottom) designs.

Inferences are implicitly conditioned on the experimental design

An important fact that is easy to forget is that all parameter inferences, be they Bayesian or frequentist, are inherently conditioned on the experimental designs used to generate the observed data. That is, the inferred belief about some parameter θ is determined by $p(\theta|\mathbf{y},\mathbf{d})$ and not $p(\theta|\mathbf{y})$, where $\mathbf{d}$ is the design with which the data $\mathbf{y}$ was observed, even though this is usually not reflected in the notation explicitly, but rather kept implicit.

As we have seen in our case study, different experimental designs may lead to different posterior inferences. For example, for the task of model discrimination, we found different proportions of participants best explained by our optimal designs and our baseline designs. The question of when and why different experimental designs might support different models or theories is an important one that should concern anyone doing empirical research, but contrasting qualitatively different designs, as we do, makes this especially salient. We would expect the simplest and most compelling evidence for some design $\mathbf{d}_a$ being 'better' than another ($\mathbf{d}_b$) to be posterior distributions that are in qualitative agreement having conducted experiments with both designs, but with lower posterior variances given design $\mathbf{d}_a$. This raises the question of what any differences in inferences may imply. In extreme cases, one might observe two different experiments leading to dramatically different conclusions—whatever the provenance of the two experimental designs. These differences may go so far as two experimental designs, $\mathbf{d}_a$ and $\mathbf{d}_b$ supporting different theories, such that $\mathbf{d}_a$ supports theory A, but falsifies theory B and $\mathbf{d}_b$ supports theory B, but falsifies theory A. Such cases might plausibly be the result of serious model misspecification, where no model under consideration gives a good account of human behavior across different designs. For example, in the present case study, if every model under consideration supposes that people use fixed strategies, but in reality people adapt to the structure or statistics of the specific task at hand, then different designs would support different models. It seems plausible that some participants in bandit tasks might perform such strategy selection (e.g. ***Lieder and Griffiths, 2017***): Having engaged with a block of deterministic bandit arms, some participants may approach the second block differently than if they had just engaged with a more stochastic design, or may even switch strategies within the same block. The problem of model misspecification is thus brought to light very clearly by BOED, as can be seen in the case study.

More generally, in cognitive science, we typically aspire to estimate parameters $\boldsymbol{\theta}$ that describe some stable (with respect to time, but also context) cognitive attribute or process. If two designs do not support the same estimated parameters, then $\boldsymbol{\theta}$ likely does not describe the stable trait that we would like it to describe. Sometimes, this may suggest adopting a higher level model that decides when $\boldsymbol{\theta}$ should have an effect (such as a strategy switching model, as discussed above), such that our new $\boldsymbol{\theta}'$ describes a policy. Alternatively, we can explicitly limit the context (or domain) in which $\boldsymbol{\theta}$ should operate. For example, in the context of our case study, instead of generalizing our theory by proposing strategy-switching mechanisms, we may say that a theory on behavior in bandit problems should only cover stochastic reward environments, where no bandit arm has deterministic (0 or 1) rewards. These considerations depend very much on the concrete phenomena being studied and the theories used to describe them.

Model misspecification and informativeness

Despite best efforts, computational models will suffer from some degree of model misspecification, failing to capture some aspects of people's behavior. While the misspecification may only be subtle in some cases, it can also be more severe, which poses a very real and important problem for computational modeling. BOED takes the perspective that the scientific question and models should come first and thus puts the model more into central focus than traditional approaches, thereby providing a more formal way to analyze the impact of modeling choices on the utility of different designs and expected results. For instance, computational models may sometimes only provide reasonable predictions in limited regions of the space of possible experimental designs. We can inspect this in Step 5 of our proposed workflow and surprising predictions may motivate iterating on our theory/model building, or if we are committed to the model, can be exploited by BOED to perform powerful tests of competing theories.

While the issue of model misspecification is not specific to BOED, optimizing experimental designs based on misspecified models can have negative consequences. In particular, if the space of models is too narrow, our data may be less useful for later analyses with well-specified models. Two simple ways of lessening the impact of model misspecification are to enforce diversity constraints and to select qualitatively different high-utility designs to cover important behavioral phenomena (***Palminteri et al., 2017***). As discussed, we can include constraints in our search over experimental designs, such that our designs are not too narrowly focused on a single region of the parameter space. We could, for example, enforce this on a block-level, to ensure that every participant is presented with a diverse set of experimental designs. Alternatively, we can perform a more manual selection and choose designs that have high utility, but are likely to elicit qualitatively different phenomena. We present exemplary ways of engaging with the estimated utility function in Appendix 6.

Ecological validity and generalizability

We have seen that BOED may sometimes lead to extreme or atypical designs, as the present results illustrate, and in some cases these designs may be deemed undesirable. Similarly, we may sometimes want to use not only one experimental design for all units of observation, but for example, slightly vary the design across participants. As discussed by ***Yarkoni, 2020***, there is a common policy in experimental research of ignoring stimulus sampling variability, also referred to as the 'fixed-effect fallacy' ***Clark, 1973***. This policy can lead to problems with generalizing inferences to broader classes of stimuli, especially when the phenomenon under question is likely sensitive to contextual effects (***Van Bavel et al., 2016***). Fortunately, the BOED workflow is inherently designed to handle such cases.

First, the optimization procedure over designs can include constraints (e.g. at most one deterministic bandit arm), which may be appropriate depending on the research question and can be viewed as inserting expert knowledge. Second, the approximation of the learned utility surface, allows for exploring the utility of different designs systematically. Rather than simply picking the design with maximal utility, practitioners can select from this surface to capture qualitatively different phenomena, or simply include slight variations of similar designs. For instance, researchers can pick the top-k designs, or adopt a more sophisticated policy to ensure that the designs being used cover important parts of the design space to ensure robustness and generalizability of results, as argued for, for example, by ***Yarkoni, 2020***. If the goal is to provide more varied designs, instead of the present approach of optimizing the reward probabilities directly, it would be straightforward to parametrize the design space differently and optimize a design policy. For instance, in the context of our case study, we could optimize the parameters (α and β) of a $\mathrm{Beta}(\alpha, \beta)$ distribution from which the reward probabilities for the bandit arms of a given block are drawn. This is clearly only one option, and different ways of parametrizing design policies are possible, as an alternative or complement to choosing designs from the estimated utility surface. The BOED procedure is agnostic to such choices, and decisions should be based on the concrete problem being studied. Researchers may here consider the trade-off between generalizabiltiy and variance control, which can be studied by simulating data from the model and performing parameter recovery analyses. We illustrate how the utility surface can be explored in Appendix 6, with code that shows how this can be accomplished in the related GitHub repository: https://github.com/simonvalentin/boed-tutorial, (copy archived at ***Valentin, 2023***).

Interpretability

Simulating data from the computational models under consideration is of key importance in all stages of the experimental design process. In this work, we treat all models as simulators and, as argued, for example, by ***Palminteri et al., 2017***, simulated data are crucial in assessing and potentially falsifying specific model commitments. Prior predictive analyses are crucial for refining models and their priors, including detecting potential problems that may already be obvious from simulations alone, without having to run a real experiment. Beyond comparing models' posterior probability, posterior predictive analyses can provide insights into where models are 'right' and where they may depart from human behavior, referring back to the previous theme of model misspecification. As proposed by ***Palminteri et al., 2017***, model comparison (through posterior probabilities) and falsification through simulation of behavioral effects thus serve complementary roles. Such simulations can serve as rigorous tests of theories, as a failure of a model to generate an effect of interest can be used as a criterion for rejecting the model (or a particular mechanism encoded in the model) in that form. These simulations should be considered a key part of the workflow presented in this work, beyond just inspecting posterior distributions (over models or their parameters).

Conclusion and outlook

This tutorial provides a flexible step-by-step workflow to finding optimal experimental designs, by leveraging recent advances in ML and BOED. As a first step, the proposed workflow suggests to define a scientific goal, such as model discrimination or parameter estimation, thereby expressing the value of a particular design as a measurable quantity using a utility function. Secondly, the scientific theories under investigation need to be cast as a computational model from which we can sample synthetic data. The methodologies discussed in this tutorial make no assumption, however, on whether or not the likelihood function of the computational model needs be tractable, which opens up the space of scientific theories that can be tested considerably. In the third step, the design optimization problem needs to be set up, which includes deciding which experimental design variables need to be optimized and whether there are any additional constraints that should be included. The fourth step of our proposed workflow then involves using ML to estimate and optimize the utility function. In doing so, we discussed a recent approach that leverages neural networks to learn a mapping between experimental designs and expected information gain, that is MI, and provided a novel extension that efficiently deals with behavioral data. In the fifth step, the discovered optimal designs are then validated in silico using synthetic data. This ensures that the optimal designs found are sound and that no experimental resources are wasted, possibly also if it is found that the computational models need to be revisited after this step. Lastly, in the sixth step, the real experiment can be performed using the discovered optimal design. Using the discussed methodologies, the trained neural networks can then conveniently be used to compute posterior distributions and summary statistics, immediately after the data collection.

This tutorial and the accompanying case study demonstrate the usefulness of modern ML and BOED methods for improving the way in which we design experiments and thus collect empirical data. Furthermore, the methodologies discussed in this tutorial optimize experimental designs for models of cognition without requiring computable likelihoods, or marginal likelihoods. This is critical to provide the methodological support required for scientific theories and models of human behavior that become increasingly realistic and complex. As part of our case study, using simulations and real-world data, we showed that the proposed workflow and the discussed methodologies yield optimal designs that outperform human-crafted designs found in the literature, offering more informative data for model discrimination as well as parameter estimation. We showed that adopting more powerful methodological tools allows us to study more realistic theories of human behavior, and our results provide empirical support for the expressiveness of the simulator models we studied in our case study. BOED is currently an active area of research and future work will likely make some steps easier and more computationally efficient, such as automatic tuning of the ML methods used to estimate MI, variations of the loss function, improved ways of searching over the design space, new ways of dealing with model misspecification or computationally efficient sequential BOED (***Rainforth et al., 2023***). What will remain despite such technical advancements, however, is the need to formulate sound scientific questions and actively engaging with tools that help to automate designs. We thus believe that the steps and considerations outlined in the workflow presented in this work will stay relevant. More broadly, ML has seen success in behavioral research regarding the analysis of large datasets and discovering new theories with promising results (e.g. ***Peterson et al., 2021***) but its potential for data collection and in particular experimental design has been explored remarkably little. We view the present work as an important step in this direction.

Acknowledgements

We thank Fiona Knäussel for her help in creating the schematics used in this paper, Maximilian Harkotte for valuable feedback during preparation of the paper and three anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. SV was supported by a Principal’s Career Development Scholarship, awarded by the University of Edinburgh. SK was supported in part by the EPSRC Centre for Doctoral Training in Data Science, funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016427/1) and the University of Edinburgh.

Additional information

Funding

Funder	Grant reference number	Author
University of Edinburgh	Principal's Career Development Scholarship	Simon Valentin
EPSRC Centre for Doctoral Training in Data Science	EP/L016427/1	Steven Kleinegesse

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Simon Valentin, Conceptualization, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing; Steven Kleinegesse, Conceptualization, Data curation, Software, Formal analysis, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing; Neil R Bramley, Michael U Gutmann, Christopher G Lucas, Conceptualization, Formal analysis, Supervision, Methodology, Writing – original draft, Writing – review and editing; Peggy Seriès, Supervision, Writing – review and editing

Author ORCIDs

Simon Valentin http://orcid.org/0000-0003-2039-8928

References

Baker M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* **533**:452–454. DOI: https://doi.org/10.1038/533452a

Bitzer S, Park H, Blankenburg F, Kiebel SJ. 2014. Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Frontiers in Human Neuroscience* **8**:102. DOI: https://doi.org/10.3389/fnhum.2014.00102, PMID: 24616689

Blei DM, Smyth P. 2017. Science and data science. *PNAS* **114**:8689–8692. DOI: https://doi.org/10.1073/pnas.1702076114

Bramley NR, Gerstenberg T, Tenenbaum JB, Gureckis TM. 2018. Intuitive experimentation in the physical world. *Cognitive Psychology* **105**:9–38. DOI: https://doi.org/10.1016/j.cogpsych.2018.05.001, PMID: 29885534

Brehmer J, Louppe G, Pavez J, Cranmer K. 2020. Mining gold from implicit models to improve likelihood-free inference. *PNAS* **117**:5242–5249. DOI: https://doi.org/10.1073/pnas.1915980117

Breiman L. 2001. Random forests. *Machine Learning* **45**:5–32. DOI: https://doi.org/10.1023/A:1010933404324

Chen Y, Zhang D, Gutmann MU, Courville A, Zhu Z. 2021. Neural approximate sufficient statistics for implicit models. In International Conference on Learning Representations (ICLR. .

Clark HH. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* **12**:335–359. DOI: https://doi.org/10.1016/S0022-5371(73)80014-3

Corander J, Fraser C, Gutmann MU, Arnold B, Hanage WP, Bentley SD, Lipsitch M, Croucher NJ. 2017. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nature Ecology & Evolution* **1**:1950–1960. DOI: https://doi.org/10.1038/s41559-017-0337-x, PMID: 29038424

Cranmer K, Brehmer J, Louppe G. 2020. The frontier of simulation-based inference. *PNAS* **117**:30055–30062. DOI: https://doi.org/10.1073/pnas.1912789117, PMID: 32471948

Currie CSM, Fowler JW, Kotiadis K, Monks T, Onggo BS, Robertson DA, Tako AA. 2020. How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation* **14**:83–97. DOI: https://doi.org/10.1080/17477778.2020.1751570

Dale AM. 1999. Optimal experimental design for event-related fMRI. *Human Brain Mapping* **8**:109–114. DOI: https://doi.org/10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W, PMID: 10524601

Dezfouli A, Griffiths K, Ramos F, Dayan P, Balleine BW. 2019. Models that learn how humans learn: The case of decision-making and its disorders. *PLOS Computational Biology* **15**:e1006903. DOI: https://doi.org/10.1371/journal.pcbi.1006903, PMID: 31185008

Dziak JJ, Coffman DL, Lanza ST, Li R, Jermiin LS. 2020. Sensitivity and specificity of information criteria. *Briefings in Bioinformatics* **21**:553–565. DOI: https://doi.org/10.1093/bib/bbz016

Eduardo A, Gutmann MU. 2023. Bayesian optimization with informative covariance. *Transactions on Machine Learning Research* 1–37.

Elsken T, Metzen JH, Hutter F. 2019. Neural architecture search: A survey. *Journal of Machine Learning Research* **20**:1–21. DOI: https://doi.org/10.1007/978-3-030-05318-5

Fisher RA. 1915. Frequency distribution of the values of the correlation coeffients in samples from an indefinitely large popu;ation. *Biometrika* **10**:507–521. DOI: https://doi.org/10.1093/biomet/10.4.507

Foster A, Jankowiak M, Bingham E, Horsfall P, Teh YW, Rainforth T, Goodman N. 2019. Variational Bayesian optimal experimental design. Advances in Neural Information Processing Systems.

Friedman JH. 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* **29**:1232. DOI: https://doi.org/10.1214/aos/1013203451

Gebhardt C, Oulasvirta A, Hilliges O. 2021. Hierarchical reinforcement learning explains task interleaving behavior. *Computational Brain & Behavior* **4**:284–304. DOI: https://doi.org/10.1007/s42113-020-00093-9

Gershman SJ. 2018. Deconstructing the human algorithms for exploration. *Cognition* **173**:34–42. DOI: https://doi.org/10.1016/j.cognition.2017.12.014, PMID: 29289795

Gershman SJ. 2020. Origin of perseveration in the trade-off between reward and complexity. *Cognition* **204**:104394. DOI: https://doi.org/10.1016/j.cognition.2020.104394, PMID: 32679270

Gouriéroux C, Phillips PCB, Yu J. 2010. Indirect inference for dynamic panel models. *Journal of Econometrics* **157**:68–77. DOI: https://doi.org/10.1016/j.jeconom.2009.10.024

Griffiths TL. 2015. Manifesto for a new (computational) cognitive revolution. *Cognition* **135**:21–23. DOI: https://doi.org/10.1016/j.cognition.2014.11.026, PMID: 25497482

Guest O, Martin AE. 2021. How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science* **16**:789–802. DOI: https://doi.org/10.1177/1745691620970585, PMID: 33482070

Huys QJ, Pizzagalli DA, Bogdan R, Dayan P. 2013. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of Mood & Anxiety Disorders* **3**:12. DOI: https://doi.org/10.1186/2045-5380-3-12, PMID: 23782813

Ioannidis JPA. 2005. Why most published research findings are false. *PLOS Medicine* **2**:e124. DOI: https://doi.org/10.1371/journal.pmed.0020124, PMID: 16060722

Ivanova DR, Foster A, Kleinegesse S, Gutmann MU, Rainforth T. 2021. Implicit Deep Adaptive Design: Policy-Based Experimental Design without Likelihoods. 35th Conference on Neural Information Processing Systems. .

Jones M, Dzhafarov EN. 2014. Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review* **121**:1–32. DOI: https://doi.org/10.1037/a0034190, PMID: 24079307

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. DOI: https://doi.org/10.1038/s41586-021-03819-2, PMID: 34265844

Kangasrääsiö A, Jokinen JPP, Oulasvirta A, Howes A, Kaski S. 2019. Parameter inference for computational cognitive models with approximate bayesian computation. *Cognitive Science* **43**:e12738. DOI: https://doi.org/10.1111/cogs.12738, PMID: 31204797

Karvelis P, Seitz AR, Lawrie SM, Seriès P. 2018. Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *eLife* **7**:e34115. DOI: https://doi.org/10.7554/eLife.34115, PMID: 29757142

Kleinegesse S, Gutmann MU. 2019. Efficient Bayesian experimental design for implicit models. In Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. .

Kleinegesse S, Drovandi C, Gutmann MU. 2020. Sequential bayesian experimental design for implicit models via mutual information. [arXiv]. https://arxiv.org/abs/2003.09379

Kleinegesse S, Gutmann MU. 2021. Gradient-based Bayesian experimental design for implicit models using mutual information lower bounds. *arXiv*. https://arxiv.org/abs/2105.04379DOI: https://doi.org/10.1214/20-BA1225

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* **521**:436–444. DOI: https://doi.org/10.1038/nature14539, PMID: 26017442

Lee MD, Zhang S, Munro M, Steyvers M. 2011. Psychological models of human and optimal performance in bandit problems. *Cognitive Systems Research* **12**:164–174. DOI: https://doi.org/10.1016/j.cogsys.2010.07.007

Lieder F, Griffiths TL. 2017. Strategy selection as rational metareasoning. *Psychological Review* **124**:762–794. DOI: https://doi.org/10.1037/rev0000075, PMID: 29106268

Liepe J, Filippi S, Komorowski M, Stumpf MPH. 2013. Maximizing the information content of experiments in systems biology. *PLOS Computational Biology* **9**:e1002888. DOI: https://doi.org/10.1371/journal.pcbi.1002888, PMID: 23382663

Lindley DV. 1956. On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics* **27**:986–1005. DOI: https://doi.org/10.1214/aoms/1177728069

Lintusaari J, Gutmann MU, Dutta R, Kaski S, Corander J. 2017. Fundamentals and recent developments in approximate bayesian computation. *SYSTEMATIC BIOLOGY* **66**:e66–e82. DOI: https://doi.org/10.1093/sysbio/syw077, PMID: 28175922

Lorenz R, Monti RP, Violante IR, Anagnostopoulos C, Faisal AA, Montana G, Leech R. 2016. The Automatic Neuroscientist: A framework for optimizing experimental design with closed-loop real-time fMRI. *NeuroImage* **129**:320–334. DOI: https://doi.org/10.1016/j.neuroimage.2016.01.032

Marjoram P, Molitor J, Plagnol V, Tavaré S. 2003. Markov chain Monte Carlo without likelihoods. *PNAS* **100**:15324–15328. DOI: https://doi.org/10.1073/pnas.0306899100

Martinez-Cantin R. 2014. Bayesopt: a bayesian optimization library for nonlinear optimization, experimental design and bandits. *Journal of Machine Learning Research: JMLR* **15**:3735–3739.

Mikkola P, Martin OA, Chandramouli S, Hartmann M, Pla OA, Thomas O, Pesonen H, Corander J, Vehtari A, Kaski S. 2021. Prior Knowledge Elicitation: The Past, Present, and Future. [arXiv]. https://arxiv.org/abs/2112.01380

Müller P. 1999. Simulation-based optimal design. *Bayesian Statistics* **6**:459–474. DOI: https://doi.org/10.1093/oso/9780198504856.001.0001

Myung JI, Pitt MA. 2009. Optimal experimental design for model discrimination. *Psychological Review* **116**:499–518. DOI: https://doi.org/10.1037/a0016104

Myung JI, Cavagnaro DR, Pitt MA. 2013. A tutorial on adaptive design optimization. *Journal of Mathematical Psychology* **57**:53–67. DOI: https://doi.org/10.1016/j.jmp.2013.05.005, PMID: 23997275

Oh C, Gavves E, Welling M. 2018. BOCK: Bayesian optimization with cylindrical kernels. In Proceedings of the 35th International Conference on Machine Learning. .

Ouyang L, Tessler MH, Ly D, Goodman ND. 2018. webppl-oed: A practical optimal experiment design system. In Proceedings of the annual meeting of the cognitive science society. .

Overstall AM, Woods DC. 2017. Bayesian design of experiments using approximate coordinate exchange. *Technometrics* **59**:458–470. DOI: https://doi.org/10.1080/00401706.2016.1251495

Palestro JJ, Sederberg PB, Osth AF, Van Zandt T, Turner BM. 2018. *Likelihood-Free Methods for Cognitive Science* Springer. DOI: https://doi.org/10.1007/978-3-319-72425-6

Palminteri S, Wyart V, Koechlin E. 2017. The Importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences* **21**:425–433. DOI: https://doi.org/10.1016/j.tics.2017.03.011

Paninski L. 2005. Asymptotic theory of information-theoretic experimental design. *Neural Computation* **17**:1480–1507. DOI: https://doi.org/10.1162/0899766053723032, PMID: 15901405

Pashler H, Wagenmakers E-J. 2012. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspectives on Psychological Science* **7**:528–530. DOI: https://doi.org/10.1177/1745691612465253, PMID: 26168108

Peterson JC, Bourgin DD, Agrawal M, Reichman D, Griffiths TL. 2021. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* **372**:1209–1214. DOI: https://doi.org/10.1126/science.abe2629, PMID: 34112693

Pitt MA, Myung IJ. 2002. When a good fit can be bad. *Trends in Cognitive Sciences* **6**:421–425. DOI: https://doi.org/10.1016/S1364-6613(02)01964-2

Poole B, Ozair S, Den Oord A, Alemi A, Tucker G. 2019. On variational bounds of mutual information. In Proceedings of the 36th International Conference on Machine Learning. .

Rainforth T, Foster A, Ivanova DR, Smith FB. 2023. Modern bayesian experimental design. [arXiv]. https://arxiv.org/abs/2302.14545

Ren P, Xiao Y, Chang X, Huang P, Li Z, Chen X, Wang X. 2021. A comprehensive survey of neural architecture search: Challenges and solutions. *ACM Computing Surveys* **54**:3447582. DOI: https://doi.org/10.1145/3447582

Robbins H. 1952. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society* **58**:527–535. DOI: https://doi.org/10.1090/S0002-9904-1952-09620-8

Ross RJH, Baker RE, Parker A, Ford MJ, Mort RL, Yates CA. 2017. Using approximate Bayesian computation to quantify cell-cell adhesion parameters in a cell migratory process. *NPJ Systems Biology and Applications* **3**:9. DOI: https://doi.org/10.1038/s41540-017-0010-7, PMID: 28649436

Rumelhart DE, Hinton GE, Williams RJ. 1986. Learning representations by back-propagating errors. *Nature* **323**:533–536. DOI: https://doi.org/10.1038/323533a0

Runge J, Bathiany S, Bollt E, Camps-Valls G, Coumou D, Deyle E, Glymour C, Kretschmer M, Mahecha MD, Muñoz-Marí J, van Nes EH, Peters J, Quax R, Reichstein M, Scheffer M, Schölkopf B, Spirtes P, Sugihara G, Sun J, Zhang K, et al. 2019. Inferring causation from time series in Earth system sciences. *Nature Communications* **10**:2553. DOI: https://doi.org/10.1038/s41467-019-10105-3, PMID: 31201306

Ryan EG, Drovandi CC, McGree JM, Pettitt AN. 2016. A review of modern computational algorithms for bayesian optimal design. *International Statistical Review* **84**:128–154. DOI: https://doi.org/10.1111/insr.12107

Schad DJ, Betancourt M, Vasishth S. 2021. Toward a principled Bayesian workflow in cognitive science. *Psychological Methods* **26**:103–126. DOI: https://doi.org/10.1037/met0000275, PMID: 32551748

Schafer CM, Freeman PE. 2012. Likelihood-free inference in cosmology: potential for the estimation of luminosity functions. Feigelson E, Babu G (Eds). *Statistical Challenges in Modern Astronomy V* New York: Springer. p. 3–19. DOI: https://doi.org/10.1007/978-1-4614-3520-4

Schulz LE, Sommerville J. 2006. God does not play dice: causal determinism and preschoolers' causal inferences. *Child Development* **77**:427–442. DOI: https://doi.org/10.1111/j.1467-8624.2006.00880.x, PMID: 16611182

Schulz E, Franklin NT, Gershman SJ. 2020. Finding structure in multi-armed bandits. *Cognitive Psychology* **119**:101261. DOI: https://doi.org/10.1016/j.cogpsych.2019.101261, PMID: 32059133

Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N. 2015. Taking the human out of the loop: a review of bayesian optimization. *Proceedings of the IEEE* **104**:148–175. DOI: https://doi.org/10.1109/JPROC.2015.2494218

Shannon CE. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**:379–423. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sisson SA, Fan Y, Tanaka MM. 2007. Sequential monte carlo without likelihoods. *PNAS* **104**:1760–1765. DOI: https://doi.org/10.1073/pnas.0607208104, PMID: 17264216

Steyvers M, Lee MD, Wagenmakers EJ. 2009. A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology* **53**:168–179. DOI: https://doi.org/10.1016/j.jmp.2008.11.002

Sutton RS, Barto AG. 2018. *Reinforcement Learning: An Introduction. Adaptive Computation and Machine Learning Series* The MIT Press.

Thompson WR. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25**:285–294. DOI: https://doi.org/10.1093/biomet/25.3-4.285

Turner BM, Van Zandt T. 2012. A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology* **56**:69–85. DOI: https://doi.org/10.1016/j.jmp.2012.02.005

Turner BM, Sederberg PB. 2014. A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review* **21**:227–250. DOI: https://doi.org/10.3758/s13423-013-0530-0, PMID: 24258272

Turner BM, Forstmann BU, Love BC, Palmeri TJ, Van Maanen L. 2017. Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology* **76**:65–79. DOI: https://doi.org/10.1016/j.jmp.2016.01.001, PMID: 31745373

Turner BM, Schley DR, Muller C, Tsetsos K. 2018. Competing theories of multialternative, multiattribute preferential choice. *Psychological Review* **125**:329–362. DOI: https://doi.org/10.1037/rev0000089, PMID: 29265855

Ullman TD, Stuhlmüller A, Goodman ND, Tenenbaum JB. 2018. Learning physical parameters from dynamic scenes. *Cognitive Psychology* **104**:57–82. DOI: https://doi.org/10.1016/j.cogpsych.2017.05.006, PMID: 29653395

Valentin S, Kleinegesse S, Bramley NR, Gutmann MU, Lucas CG. 2021. Bayesian optimal experimental design for simulator models of cognition. *arXiv*. https://arxiv.org/abs/2110.15632

Valentin S. 2023. Boed-Tutorial. swh:1:rev:94f32a04693763ec6af5d31247caf129bfe99142. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:18d09063fd478a0d6b69251334b52ffe9c7f66d8;origin=https://github.com/simonvalentin/boed-tutorial;visit=swh:1:snp:f53dfc8f8da5a38fe0e90465366f408bb1489f01;anchor=swh:1:rev:94f32a04693763ec6af5d31247caf129bfe99142

Van Bavel JJ, Mende-Siedlecki P, Brady WJ, Reinero DA. 2016. Contextual sensitivity in scientific reproducibility. *PNAS* **113**:6454–6459. DOI: https://doi.org/10.1073/pnas.1521897113

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017. Attention is all you need. Advances in Neural Information Processing Systems. .

Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M. 2018. Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience* **21**:860–868. DOI: https://doi.org/10.1038/s41593-018-0147-8, PMID: 29760527

Weiss DJ, Kingsbury GG. 1984. Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement* **21**:361–375. DOI: https://doi.org/10.1111/j.1745-3984.1984.tb01040.x

Welch BL. 1947. The generalisation of student's problems when several different population variances are involved. *Biometrika* **34**:28–35. DOI: https://doi.org/10.1093/biomet/34.1-2.28, PMID: 20287819

Wilson RC, Collins AG. 2019. Ten simple rules for the computational modeling of behavioral data. *eLife* **8**:e49547. DOI: https://doi.org/10.7554/eLife.49547, PMID: 31769410

Wolpert DH. 1996. The lack of a priori distinctions between learning algorithms. *Neural Computation* **8**:1341–1390. DOI: https://doi.org/10.1162/neco.1996.8.7.1341

Yarkoni T. 2020. The generalizability crisis. *The Behavioral and Brain Sciences* **45**:e1. DOI: https://doi.org/10.1017/S0140525X20001685, PMID: 33342451

Zhang S, Lee MD, Munro M. 2009. Human and optimal exploration and exploitation in bandit problems. *Ratio* **13**:14.

Zhang S, Lee MD. 2010. Optimal experimental design for a class of bandit problems. *Journal of Mathematical Psychology* **54**:499–508. DOI: https://doi.org/10.1016/j.jmp.2010.08.002

Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M. 2020. Graph neural networks: A review of methods and applications. *AI Open* **1**:57–81. DOI: https://doi.org/10.1016/j.aiopen.2021.01.001

Appendix 1

Experiments

Priors

We use a uniform categorical prior over the model indicator m, that is $p(m) = \mathcal{U}(\{1, 2, 3\})$. We generally use uninformative priors $\mathcal{U}(0, 1)$ for all model parameters, except for the temperature parameter of the WSLTS model that has a $\mathrm{LogNorm}(0, 1)$ prior, as it acts as an exponent in reshaping the posterior. We generate 50,000 samples from the prior and then simulate corresponding synthetic data $\mathbf{y}|\boldsymbol{\theta}, \mathbf{d}$ at every design $\mathbf{d}$.

Sub-networks

For all our experiments, we use sub-networks $\mathbf{S}_{\psi}(\boldsymbol{v}, \mathbf{y})$ that consist of two hidden layers with 64 and 32 hidden units, respectively, and ReLU activation functions. The number of sufficient statistics we wish to learn for each block of behavioral data is given by the number of dimensions in the output layer of the sub-networks. These are 6, 8, 6 and 8 units for the model discrimination, parameter estimation (WSLTS), parameter estimation (AEG) and parameter estimation (GLS) experiments, respectively. The flexibility of a sub-network is naturally improved when increasing the number of desired summary statistics, but the computational cost increases accordingly. When the number of summary statistics is too low, the summary statistics we learn may not be sufficient. We have found the above numbers of summary statistics to be effective middle-grounds and refer to ***Chen et al., 2021*** for more detailed guidance on how to select the number of summary statistics.

Main network

The main network $T_{\psi}(\boldsymbol{v}, \mathbf{y})$ consists of the concatenated outputs of the sub-networks for each block of behavioral data and the variable of interest. This is then followed by two fully-connected layers with ReLU activation functions. For the model discrimination experiment, we use 32 hidden units for the two hidden layers, while we use 64 and 32 hidden units for the parameter estimation experiments. See ***Box 4—figure 1*** for a visualization of this bespoke neural network architecture.

Training

We use the Adam optimizer to maximize the lower bound shown in ***Equation 2***, with a learning rate of 10^{-3} and a weight decay of 10^{-3} (except for the parameter estimation (WSLTS) experiments where we use a weight decay of 10^{-4}). We additionally use a plateau learning rate scheduler with a decay factor of 05 and a patience of 25 epochs. We train the neural network for 200, 400, 300 and 300 epochs for the model discrimination, parameter estimation (WSLTS), parameter estimation (AEG) and parameter estimation (GLS) experiments, respectively. At every design we simulate 50,000 samples from the data-generating distribution (one for every prior sample) and randomly hold out 10,000 of those as a validation set, which are then used to compute an estimate of the mutual information via ***Equation 2***. During the BO procedure we select an experimental budget of 400 evaluations (80 of which were initial evaluations), which is more than double needed to converge.

Appendix 2

Human participant study

Task

Participants completed the multi-armed bandit tasks in online experiments. After going through instructions on the interface and setup of the task, participants were required to pass a series of five comprehension questions (in a true-false format). If any of the comprehension questions were answered incorrectly, the participant was sent back to the instructions and could only progress to the task once they answered all questions correctly. An example screen-shot of the task interface can be found in Appendix 5.

Two-phase design

To allocate participants to the parameter estimation designs for the respective model that best matched their behavior in the model discrimination blocks, we implemented a simple API that uses the ensemble of trained neural networks and performs approximate MAP inference over the model indicator (see the previous sub-section on posterior estimation). As we have obtained amortized posterior distributions as by-products in the BOED procedure, this inference can be done efficiently: Forward-passes through the neural networks are computationally cheap, only taking a fraction of a second, which means that there was no noticeable delay for the allocation of the optimal model for each participant.

Participants

$N = 326$ adults (154 female, mean age $M = 35.61$) participated in the experiment in return for a basic payment of £0.60 and performance related bonuses of up to £1.00. Participants took, on average, 8.8 ($SD = 2.89$) minutes to complete the task.

Appendix 3

Computational models

We here provide further details about the computational models of human behavior in bandit tasks, which we briefly presented in the main text. We consider the multi-armed bandit setting where, at each trial $t = 1, \dots, T$, participants have to make an action $a^{(t)} \in \{1, \dots, K\}$, which consists of choosing any of the K bandit arms, and subsequently observe a reward $r^{(t)} \in \{0, 1\}$. After a participant has gone through all T trials, we summarize their behavior by a set of actions $\boldsymbol{a} = (a^{(1)}, \dots, a^{(T)})$ and observed rewards $\boldsymbol{r} = (r^{(1)}, \dots, r^{(T)})$.

Whether a participant observes a reward of 0 or 1 when making a particular choice $a^{(t)}$ depends on the specified reward probability. We here assume that each bandit arm $k \in \{1, \dots, K\}$ is associated with a Bernoulli reward distribution. Each of these reward distributions has a (potentially) different reward probability, which is given by the corresponding entry in the design vector $\boldsymbol{d} \in [0, 1]^K$. In other words, the reward $r^{(t)}$ a participant receives when making a particular choice $a^{(t)} = k$ is sampled via $r^{(t)} \sim \text{Bernoulli}(d_k)$, where d_k is the k-th element of $\mathbf{d}$.

Computational models of human behavior in bandit tasks only differ in how they model the choices of a participant depending on the previous actions and rewards. To help us describe the mechanisms of such computational models, we here define the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which store the number of observed 0 and 1 rewards for all K arms. That is, α_k refers to the number of times the k-th arm was selected and subsequently generated a reward. Similarly, $\boldsymbol{\beta}_k$ refers to the number of times a participant did not observe a reward when selecting the k-th bandit arm. Below, we describe each of the computational models in more detail and include corresponding pseudo-code.

Win-Stay Lose-Thompson-Sample (WSLTS)

Here, we propose Win-Stay Lose-Thompson-Sample (WSLTS) as an amalgamation of Win-Stay Lose-Shift (WSLS; ***Robbins, 1952***) and Thompson Sampling (***Thompson, 1933***). WSLTS has three model parameters $(\theta_0, \theta_1, \theta_2)$ that regulate different aspects of exploration and exploitation behavior. Specifically, θ_0 corresponds to the probability of staying after winning, that is the probability of re-selecting the previous arm $a^{(t-1)}$ after having observed $r^{(t-1)} = 1$. However, with probability $1 - \theta_0$ the agent may decide to choose a different bandit arm even after having observed a reward of 1 at the previous trial. In this case, the agent performs Thompson Sampling, using a temperature parameter θ_2, to decide which bandit arm to select (see ***Equation 4***). Conversely, θ_1 corresponds to the probability of switching to another arm when observing a loss in the previous trial, that is when $r^{(t-1)} = 0$, in which case the agent again performs Thompson Sampling to select another bandit arm. However, the agent may also re-select the previous bandit arm with probability $1 - \theta_1$ even after having observed a loss.

During the exploration phases mentioned above, the WSLTS agent performs Thompson Sampling from a reshaped posterior, which is controlled via a temperature parameter $\boldsymbol{\theta}_2$. The corresponding choice of bandit arm is then given by

$$a^{(t)} = \arg\max_k \omega_k, \quad \omega_k \sim \begin{cases} \text{Beta}\left(\alpha_k^{1/\boldsymbol{\theta}_2}, \beta_k^{1/\boldsymbol{\theta}_2}\right) & \text{if } k \neq a^{(t-1)} \\ 0 & \text{if } k = a^{(t-1)}, \end{cases} \tag{4}$$

where α_k and β_k correspond to the number of times a participant observed a reward of 1 and 0, respectively, for a bandit arm k. This stands in contrast to standard WSLS, where the agent shifts to another arm uniformly at random. We note that, for θ_2 values close to 1 we recover standard Thompson-Sampling (excluding the previous arm), while for $\theta_2 \to \infty$ we recover standard WSLS, and for $\theta_2 \to 0$ we obtain a greedy policy. We provide pseudo-code for the WSLTS computational model in Algorithm 1.

Algorithm 1 **Win-Stay Lose-Thompson-Sample (WSLTS)**

```
Input: Parameter $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$, design $\boldsymbol{d} = (d_1, \ldots, d_K)$
Output: Actions $\boldsymbol{a} = (a^{(1)}, \ldots, a^{(T)})$, rewards $\boldsymbol{r} = (r^{(1)}, \ldots, r^{(T)})$
Initialize pseudo-counts $\alpha_k$ and $\beta_k$ to 1 for all bandit arms $k$.
for t = 1, …, T do
    Sample $u \sim \mathcal{U}(0, 1)$.
    if t=1 then
        Select the first bandit arm $a^{(1)}$ uniformly at random.
    else if $r^{(t-1)} = 1$ then
        if $u < \theta_0$ then
            Re-select the previous bandit arm $a^{(t-1)}$.
        else
            Thomson-Sample according to Equation 4.
    else
        if $u < \theta_1$ then
            Thomson-Sample according to equation 4.
        else
            Re-select the previous previous bandit arm $a^{(t-1)}$.
    Sample the reward $r^{(t)} \sim \text{Bernoulli}(d_{a^{(t)}})$.
    Increment $\alpha_{a^{(t)}} \leftarrow \alpha_{a^{(t-1)}} + 1$ if $r^{(t)} = 1$ and $\beta_{a_t} \leftarrow \beta_{a^{(t-1)}} + 1$ if $r^{(t)} = 0$.
```

Autoregressive ε-Greedy (AEG)

Standard ε-Greedy (e.g. ***Sutton and Barto, 2018***) is a ubiquitous method in reinforcement learning, where the agent selects the arm with the highest expected reward with probability $1 - \varepsilon$, where ε is a model parameter. Conversely, with probability ε, the agent performs exploration by uniformly selecting a bandit arm. Here, we propose Auto-regressive ε-Greedy (AEG) as a generalization of ε-Greedy, which allows for modeling people's tendency towards auto-regressive behavior (e.g. ***Gershman, 2020***).

Specifically, the AEG model has two model parameters (θ_0, θ_1), where θ_0 corresponds to the same ε parameter as in ε-Greedy. However, as opposed to randomly selecting any bandit arm, the probability of selecting the previous arm, in order to break ties between options with the same expected reward, is specifically controlled via the second model parameter θ_1. We provide pseudo-code for the AEG computational model in Algorithm 2.

Algorithm 2 **Autoregressive ε-Greedy (AEG)**

```
Input: Parameter $\boldsymbol{\theta} = (\theta_0, \theta_1)$, design $\boldsymbol{d} = (d_1, \ldots, d_K)$
Output: Actions $\boldsymbol{a} = (a^{(1)}, \ldots, a^{(T)})$, rewards $\boldsymbol{r} = (r^{(1)}, \ldots, r^{(T)})$
Initialize pseudo-counts $\alpha_k$ and $\beta_k$ to 1 for all bandit arms $k$.
for t=1, …, T do
    Sample $u \sim \mathcal{U}(0, 1)$ and $v \sim \mathcal{U}(0, 1)$.
    Update the estimated reward probability for all $K$ bandit arms as $\frac{\alpha_k}{\alpha_k + \beta_k}$.
    Let $M$ be the set of bandit arms with maximal expected reward.
    if t=1 then
        Select the first bandit arm $a^{(1)}$ uniformly at random.
    else if $u < \theta_0$ then
        if $v < \boldsymbol{\theta}_1 + \left(1 - \boldsymbol{\theta}_1\right)/K$ then
            Re-select the previous bandit arm $a^{(t-1)}$.
        else
            Randomly select another bandit arm.
    else
        if $a^{(t-1)} \in M$ and $v < \boldsymbol{\theta}_1 + \left(1 - \boldsymbol{\theta}_1\right)/|M|$ then
            Re-select the previous bandit arm $a^{(t-1)}$.
        else
            Randomly select another bandit arm among the set $M$.
    Sample the reward $r^{(t)} \sim \text{Bernoulli}(d_{a^{(t)}})$.
    Increment $\alpha_{a^{(t)}} \leftarrow \alpha_{a^{(t-1)}} + 1$ if $r^{(t)} = 1$ and $\beta_{a_t} \leftarrow \beta_{a^{(t-1)}} + 1$ if $r^{(t)} = 0$.
```

Generalized Latent State (GLS)

Lee et al., 2011 proposed a latent state model for bandit tasks whereby a learner can be in either an *explore* or an *exploit* state and switch between these as they go through the task. Here, we propose the Generalized Latent State (GLS) model, which unifies and extends latent-state and latent-switching models, previously studied in ***Lee et al., 2011***, allowing for more flexible and structured transitions (we thank Patrick Laverty for contributing towards this). The transition distribution over whether the agent is in an exploit state is specified by four parameters, each of which specify the probability of being in an exploit state at trial t. This transition distribution is denoted by $\pi(l, r_{t-1})$, where l is the previous latent state and r_{t-1} the reward observed in the previous trial. We here refer to trials where a bandit arm was selected but failed to produce a reward as *failures*. We provide pseudo-code for our GLS computational model in Algorithm 3.

Algorithm 3 **Generalized Latent State (GLS)**

```
Input: Parameter $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$, design $\boldsymbol{d} = (d_1, \ldots, d_K)$
Output: Actions $\boldsymbol{a} = (a^{(1)}, \ldots, a^{(T)})$, rewards $\boldsymbol{r} = (r^{(1)}, \ldots, r^{(T)})$
Initialize pseudo-counts $\alpha_k$ and $\beta_k$ to 1 for all bandit arms $k$.
Sample the initial latent state $l \sim \mathrm{Bernoulli}(0.5)$.
for t=1, … do
    Let $R \leftarrow$ all $k$ such that $k = \arg\max_k \boldsymbol{\alpha}$ (arms with the maximal number of rewards).
    Let $F \leftarrow$ all $k$ such that $k = \arg\min_k \boldsymbol{\beta}$ (arms with the minimal number of failures).
    Let $S \leftarrow R \cap F$.
    Sample $u \sim \mathcal{U}(0, 1)$ and $v \sim \mathcal{U}(0, 1)$.
    if t=1 then
        Select the first bandit arm $a^{(1)}$ uniformly at random.
    else if $|S| > 1$ then
        Randomly select a bandit arm $a^{(t)}$ from the set $S$.
    else if $|S| = 1$ then
        if $u < \theta_0$ then
            Select the bandit arm $a^{(t)} \in S$
        else
            Randomly select another bandit arm.
    else if $v < \pi(l, r_{t-1})$ then
        $l \leftarrow 1$                                    ▷ Latent state is *exploit*
        if $u < \theta_0$ then
            Let $R_{\min}$ be the set of bandit arms in $R$ with the minimal number of failures.
            Randomly select a bandit arm $a^{(t)}$ from the set $R_{\min}$.
        else
            Randomly select another bandit arm.
    else
        $l \leftarrow 0$                                    ▷ Latent state is *explore*
        if $u < \theta_0$ then
            Let $F_{\max}$ be the set of bandit arms in $F$ with the maximal number of rewards.
          Randomly select a bandit arm $a^{(t)}$ from the set $F_{\max}$.
     else
          Randomly select another bandit arm.

          Sample the reward $r^{(t)} \sim \mathrm{Bernoulli}(d_{a^{(t)}})$.
          Increment $\alpha_{a^{(t)}} \leftarrow \alpha_{a^{(t-1)}} + 1$ if $r^{(t)} = 1$ and $\beta_{a_t} \leftarrow \beta_{a^{(t-1)}} + 1$ if $r^{(t)} = 0$.
```

Appendix 4

BOED Algorithm

Algorithm 4 describes the BOED procedure used in our work to determine optimal experiments in bandit tasks.

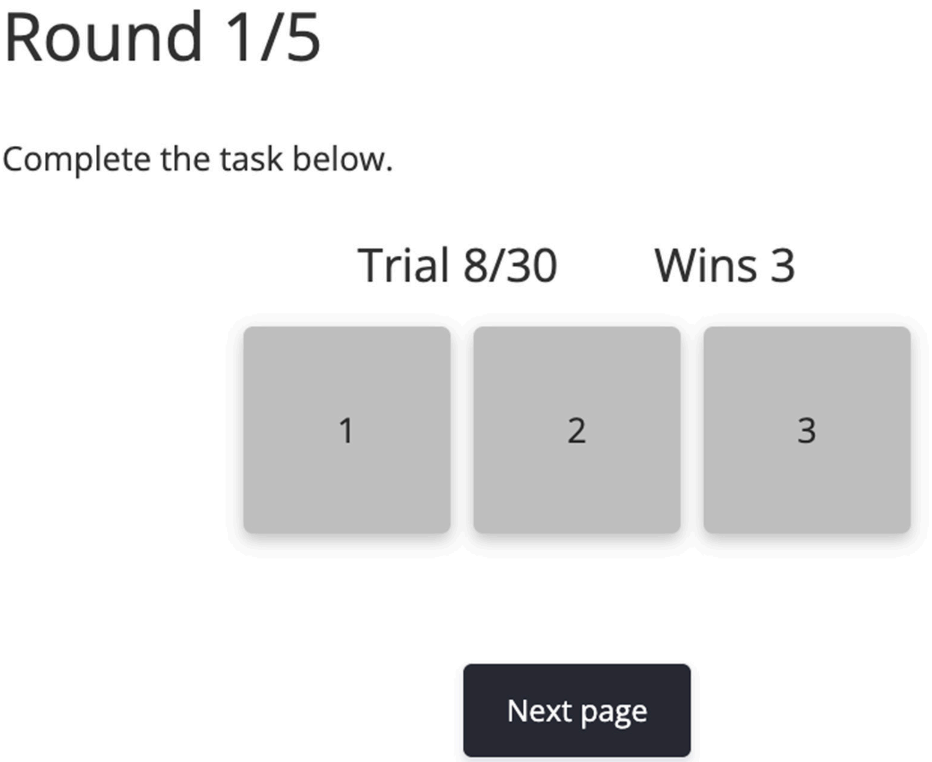
Algorithm 4 **BOED**

```
Input: Simulator model $\mathbf{y} \sim p(\mathbf{y}|\boldsymbol{v}, \mathbf{d})$, prior distribution $p(\boldsymbol{v})$, neural network (NN) architecture for $T_{\boldsymbol{\psi}}(\boldsymbol{v}, \mathbf{y})$
Output: Optimal design $\mathbf{d}^*$, trained NN $T_{\boldsymbol{\psi}^*}(\boldsymbol{v}, \mathbf{y})$ at $\mathbf{d}^*$
Randomly initialize the experimental designs $\mathbf{d}_n \leftarrow \mathbf{d}_0$
Initialize the Gaussian Process for Bayesian optimization (BO)
while $U(\mathbf{d}_n)$ not converged do
    Sample from the prior: $\boldsymbol{v}^{(i)} \sim p(\boldsymbol{v})$ for $i = 1, \ldots, M$
    Sample from the simulator: $\mathbf{y}^{(i)} \sim p(\mathbf{y}|\boldsymbol{v}^{(i)}, \mathbf{d})$ for $i = 1, \ldots, M$
    Randomly initialize the NN parameters $\boldsymbol{\psi}_n \leftarrow \boldsymbol{\psi}_0$
    while $U(\mathbf{d}_n; \boldsymbol{\psi}_n)$ with fixed $\mathbf{d}_n$ not converged do
        Compute a sample average of the lower bound (see the main text)
        Estimate gradients of the sample average with respect to $\boldsymbol{\psi}_n$
        Update $\boldsymbol{\psi}_n$ using any gradient-based optimizer
    Use $\mathbf{d}_n$ and $U(\mathbf{d}_n)$ to update the Gaussian Process
    Use BO to determine which $\mathbf{d}_{n+1}$ to evaluate next
```

Appendix 5

Additional figures and results

In this section, we provide additional figures and results that supplement our main text. ***Appendix 5—figure 1*** shows a screenshot of the bandit task that participants were presented with in our human experiment. ***Appendix 5—figures 2 and 3*** show the parameter estimation results for the AEG and GLS simulator models, respectively, of our simulation study. Specifically, these figures show marginal posterior distributions of the model parameters, averaged over several (simulated) observations, for baseline designs and our optimal designs. With regards to the real experiment with human participants, ***Appendix 5—figures 4 and 5*** show distributions of the posterior entropies in the parameter estimation task for the WSLTS and GLS simulator models, respectively. These distributions compare the differential entropy of the posteriors obtained using our optimal designs with those obtained using the baseline designs. Lastly, ***Appendix 5—figures 6 and 7*** show example posteriors of human participants assigned to the WSLTS model, while ***Appendix 5—figures 8 and 9*** show example posteriors of human participants assigned to the GLS model.

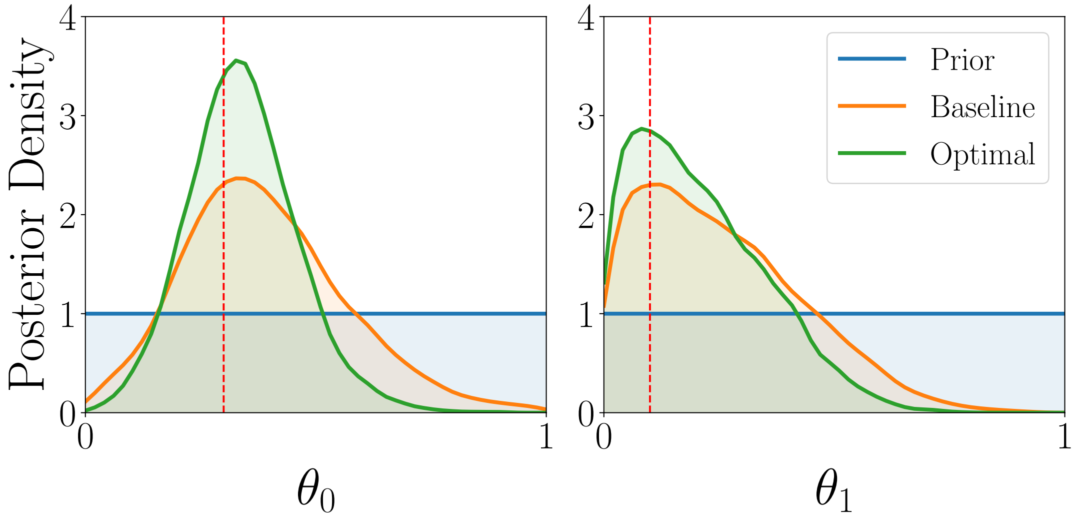


Round 1/5

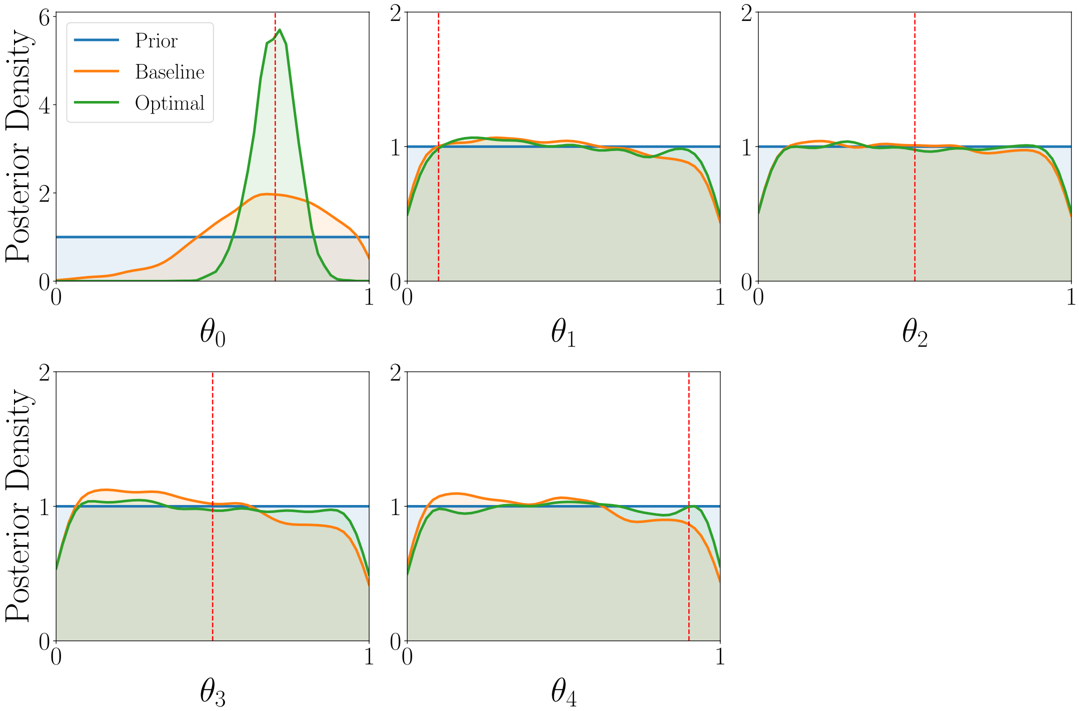
Complete the task below.

Trial 8/30 Wins 3

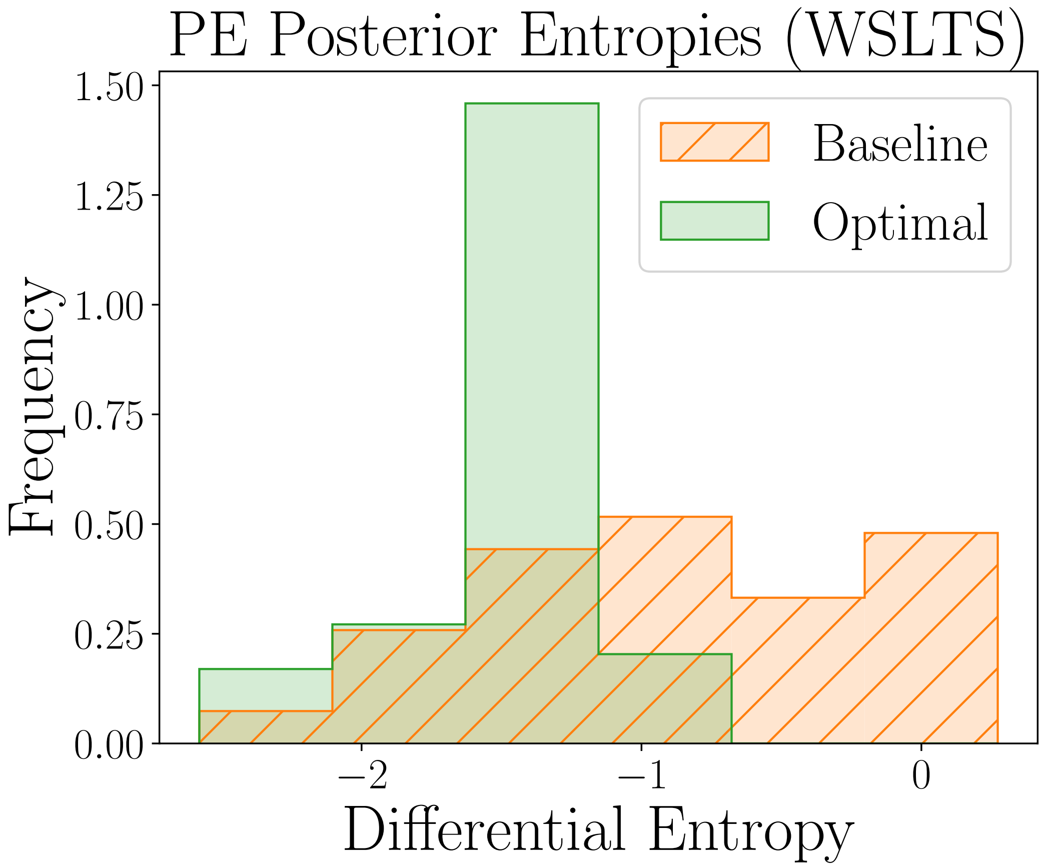
Appendix 5—figure 1. Screenshot of bandit task that participants completed in online experiments.



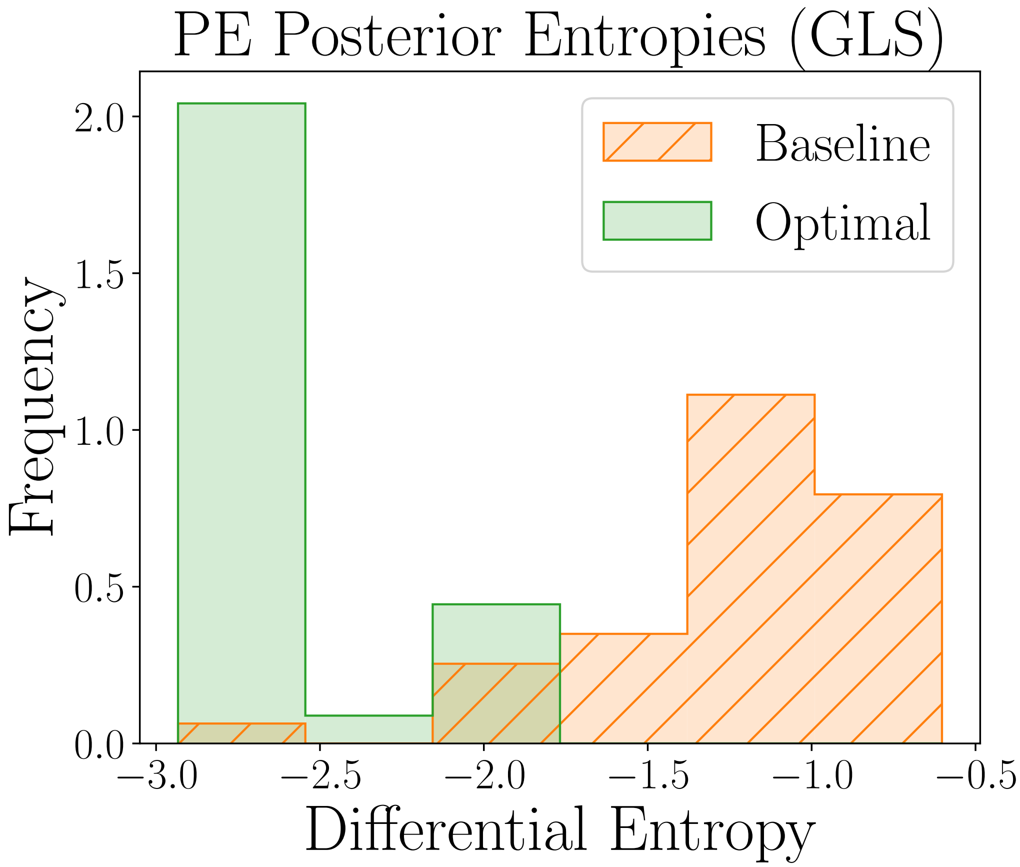
Appendix 5—figure 2. Results for the parameter estimation task of the AEG model, showing the marginal posterior distributions of the three AEG model parameters for optimal (green) and baseline (orange) designs, averaged over 1000 observations.



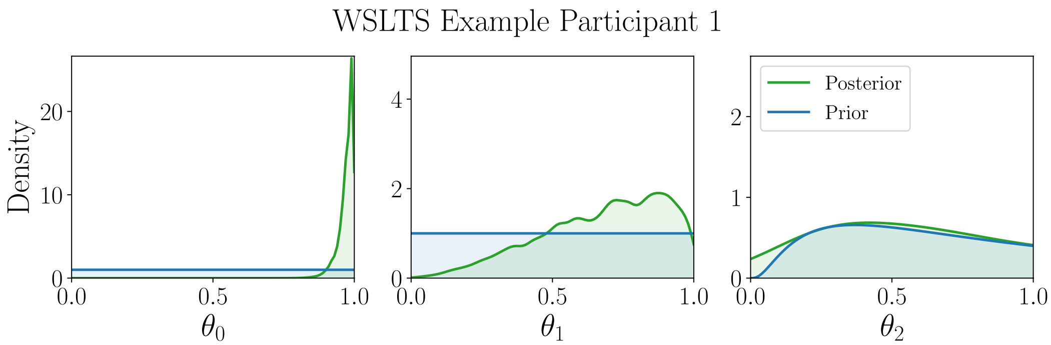
Appendix 5—figure 3. Results for the parameter estimation task of the GLS model, showing the marginal posterior distributions of the three GLS model parameters for optimal (green) and baseline (orange) designs, averaged over 1000 observations.



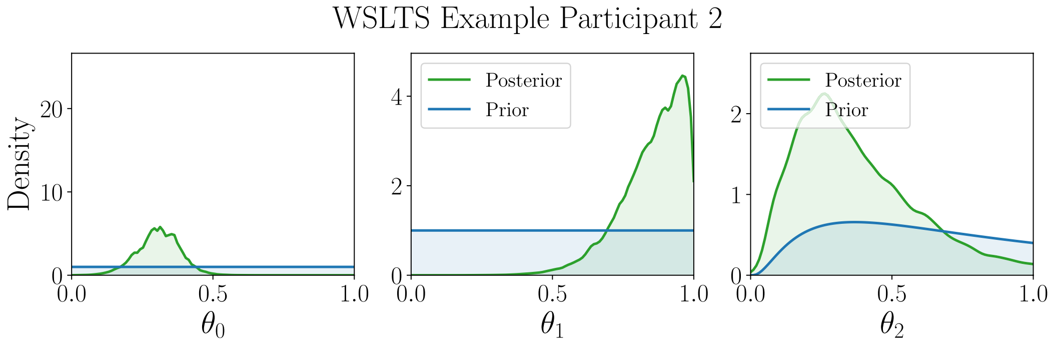
Appendix 5—figure 4. Results for the parameter estimation task of the WSLTS model, showing the distribution of posterior differential entropies obtained for optimal (green) and baseline (orange) designs (lower is better).



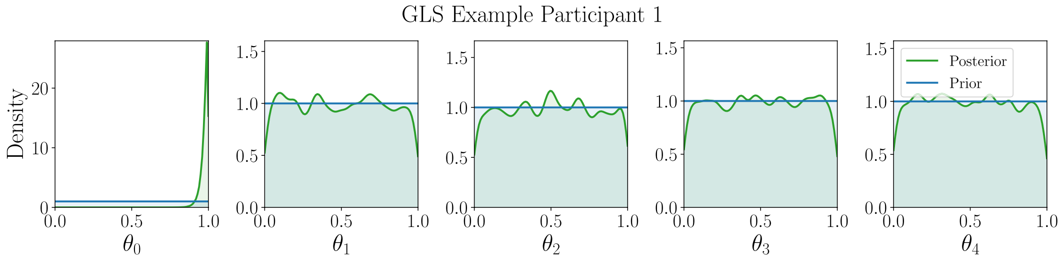
Appendix 5—figure 5. Results for the parameter estimation task of the GLS model, showing the distribution of posterior differential entropies obtained for optimal (green) and baseline (orange) designs (lower is better).



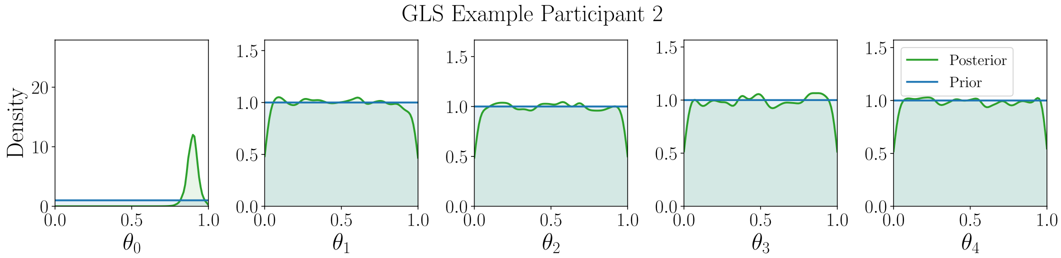
Appendix 5—figure 6. Marginal posterior distributions (green) of the WSLTS model parameters for example participant 1.



Appendix 5—figure 7. Marginal posterior distributions (green) of the WSLTS model parameters for example participant 2.



Appendix 5—figure 8. Marginal posterior distributions (green) of the GLS model parameters for example participant 1.

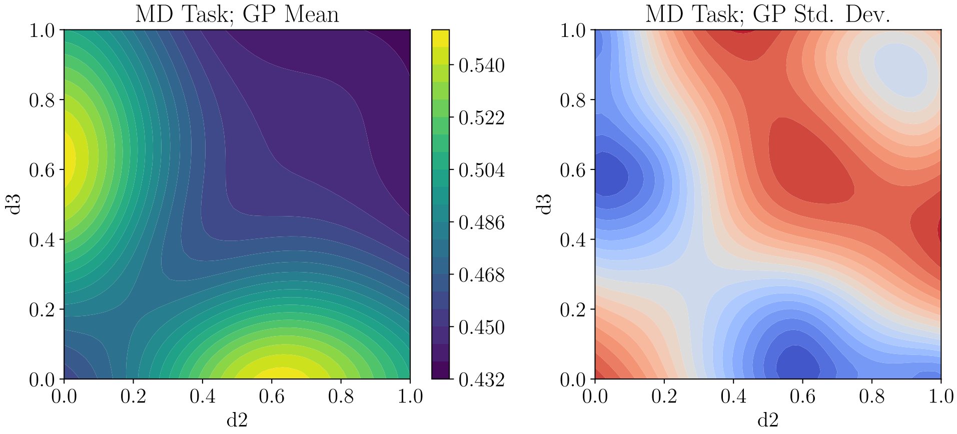


Appendix 5—figure 9. Marginal posterior distributions (green) of the GLS model parameters for example participant 2.

Appendix 6

Exploring the utility surface and locally optimal designs

In this section, we provide an example of how our framework allows for exploring the surface of the utility function and (locally) optimal designs. To do so, we consider the example of the model discrimination task where the optimal design was found to be around $[0, 0, 0.6]$ for the first bandit arm and $[1, 1, 0]$ for the second bandit arm. Our framework maximizes a probabilistic surrogate model, for example a Gaussian process (GP), over the estimated mutual information (MI) values in order to find optimal designs. It is straightforward to extract the learned GP from the final round of training, which allows us to utilize it for post-training analysis and exploration. Since the design space is six-dimensional, it is difficult to visualize the entirety of the utility function, provided by the mean function of the learned GP, at once. However, it is possible to explore the utility function by means of slicing. An example of visualizing such a slice is shown in ***Appendix 6—figure 1***. Furthermore, it is possible to systematically search for local optima of the utility function by running stochastic gradient ascent (SGA) over the mean function of the GP, with several restarts. We have done this for 20 restarts for the above example and summarized the (unique) local optima of the utility function in ***Appendix 6—table 1***. Note that instead of using SGA, we could also treat the utility as an unnormalized density and sample designs proportional to this density (see, e.g. ***Müller, 1999***).



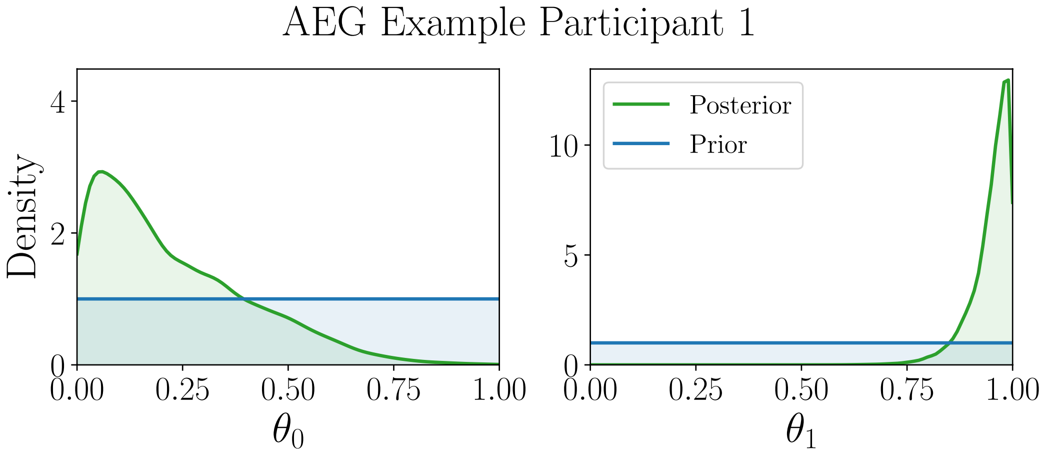
Appendix 6—figure 1. Example of how to explore the high-dimensional utility function for the model discrimination (MD) task, by slicing the Gaussian process (GP) learned during the design optimization step. The left plot shows the mean of the GP and the right plot shows the standard deviation of the GP. Shown is the slice corresponding to the design,$\mathbf{d} = [[0, d_2, d_3], [1, 1, 0]]$ which contains the global optimum.

Appendix 6—table 1. A ranking of local design optima for the model discrimination task.
Shown are the rank, the mutual information (MI) estimate computed via the Gaussian process (GP) mean, and the locally optimal designs. Note that, for this task, the designs for the first bandit arm were $[d_1, d_2, d_3]$ and $[d_4, d_5, d_6]$ for the second bandit arm. The 5 unique optima were obtained by running stochastic gradient ascent on the GP mean with 20 restarts.

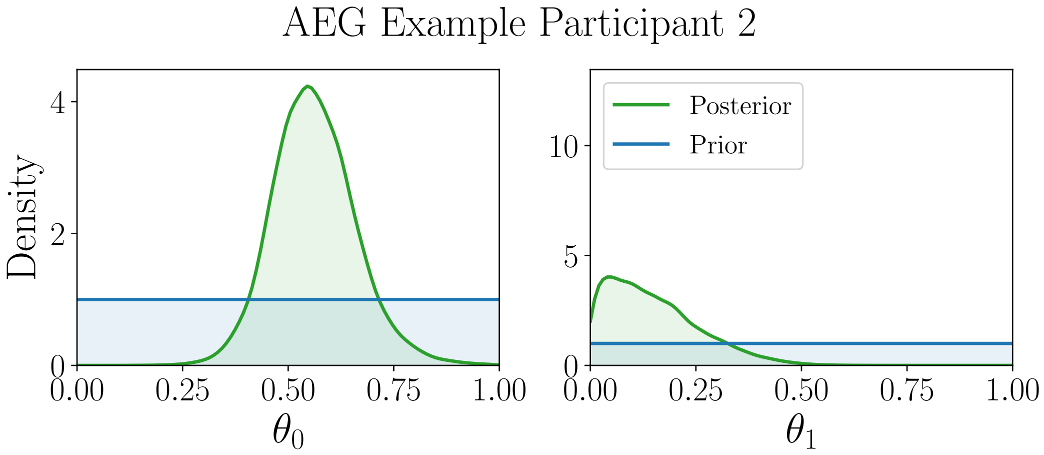
Rank of Optimum	MI	d_1	d_2	d_3	d_4	d_5	d_6
1	0.672	0.000	0.000	0.607	1.000	1.000	0.000
2	0.526	1.000	0.006	1.000	0.140	0.266	0.528
3	0.485	0.000	1.000	0.774	0.275	0.399	0.630
4	0.479	0.683	0.775	0.000	0.216	0.717	0.128
5	0.475	0.009	0.710	0.583	0.842	0.000	0.109

Appendix 7

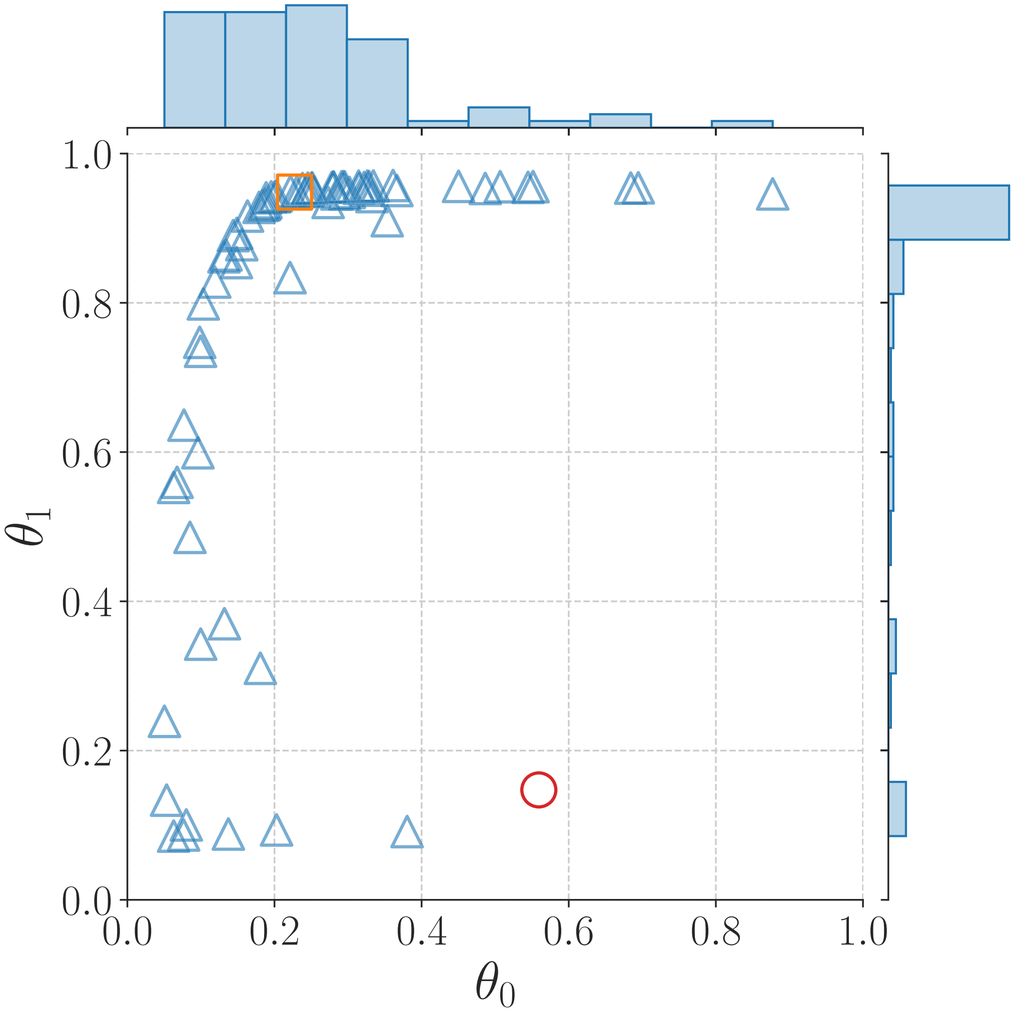
Example participants



Appendix 7—figure 1. Inferred marginal posterior model parameters for human participants best described by the AEG model. Marginal posterior distributions (green) of the AEG model parameters for example participant 1.



Appendix 7—figure 2. Posterior distribution (green) of the AEG model parameters for example participant 2.



Appendix 7—figure 3. Mean values of the parameter posterior distributions for each of the 75 participants best described by the AEG model; the square and circle markers represent example participants 1 and 2, respectively.

We now explore the data from our optimal design group in the human participant study, specifically focusing on participants that were best described by the AEG model as an example. We qualitatively investigate their behavior by visualizing the (marginal) posterior distribution of the AEG model, as shown in ***Appendix 7—figure 1*** and ***Appendix 7—figure 2*** and for two example participants. The posterior distribution for the first participant (***Appendix 7—figure 1***) indicates small values for the first parameter 0 of the AEG model, that is the ε parameter, which implies that the participant tends to select the arm with the highest probability of receiving a reward as opposed to making random exploration decisions. Larger values of the second parameter 1 of the AEG model, that is the 'stickiness' parameter, imply that the participant is biased towards re-selecting the previously chosen bandit arm. The second example participant (***Appendix 7—figure 1***) shows markedly different behavior. As indicated by the marginal posterior of the 0 parameter, this participant engages in considerably more random choices. Moreover, this participant shows 'anti-sticky' behavior, as implied by low values of the second parameter 1. Lastly, we compute the posterior mean for each participant best described by the AEG model and visualize them in a scatter plot in ***Appendix 7—figure 3***. We find that most participants best described by the AEG model show behavior somewhat similar to the first example participant, but as expected, there is inter-individual variation, for example, with some participants displaying 'anti-sticky' behavior.