

A Bayesian Approach for Quantifying Data Scarcity when Modeling Human Behavior via Inverse Reinforcement Learning

TAHERA HOSSAIN*, University of Michigan and Kyushu Institute of Technology

WANGGANG SHEN*, University of Michigan

ANINDYA DAS ANTAR, University of Michigan

SNEHAL PRABHUDESAI, University of Michigan

SOZO INOUE, Kyushu Institute of Technology

XUN HUAN, University of Michigan

NIKOLA BANOVIC, University of Michigan

Computational models that formalize complex human behaviors enable study and understanding of such behaviors. However, collecting behavior data required to estimate the parameters of such models is often tedious and resource intensive. Thus, estimating dataset size as part of data collection planning (also known as Sample Size Determination) is important to reduce the time and effort of behavior data collection while maintaining an accurate estimate of model parameters. In this paper, we present a sample size determination method based on Uncertainty Quantification (UQ) for a specific Inverse Reinforcement Learning (IRL) model of human behavior, in two cases: 1) *pre-hoc* experiment design—conducted in the planning stage before any data is collected, to guide the estimation of how many samples to collect; and 2) *post-hoc* dataset analysis—performed after data is collected, to decide if the existing dataset has sufficient samples and whether more data is needed. We validate our approach in experiments with a realistic model of behaviors of people with Multiple Sclerosis (MS) and illustrate how to pick a reasonable sample size target. Our work enables model designers to perform a deeper, principled investigation of effects of dataset size on IRL model parameters.

CCS Concepts: • **Human-centered computing** → **HCI theory, concepts and models**.

Additional Key Words and Phrases: Sample size determination, behavior modeling, Inverse Reinforcement Learning, Bayesian inference.

ACM Reference Format:

Tahera Hossain, Wanggang Shen, Anindya Das Antar, Snehal Prabhudesai, Sozo Inoue, Xun Huan, and Nikola Banovic. 2022. A Bayesian Approach for Quantifying Data Scarcity when Modeling Human Behavior via Inverse Reinforcement Learning. *ACM Trans. Comput.-Hum. Interact.* 1, 1 (2022), 27 pages. <https://doi.org/10.1145/3551388>

*Both authors contributed equally to this research.

Authors' addresses: Tahera Hossain, taherah@umich.edu, University of Michigan, Ann Arbor, Michigan, 48109; Kyushu Institute of Technology, Kitakyushu, Fukuoka, 804-8550; Wanggang Shen, wgshen@umich.edu, University of Michigan, Ann Arbor, Michigan, 48109; Anindya Das Antar, adantar@umich.edu, University of Michigan, Ann Arbor, Michigan, 48109; Snehal Prabhudesai, snehalbp@umich.edu, University of Michigan, Ann Arbor, Michigan, 48109; Sozo Inoue, sozo@brain.kyutech.ac.jp, Kyushu Institute of Technology, Kitakyushu, Fukuoka, 808-0196; Xun Huan, xhuan@umich.edu, University of Michigan, Ann Arbor, Michigan, 48109; Nikola Banovic, nbanovic@umich.edu, University of Michigan, Ann Arbor, Michigan, 48109.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

1073-0516/2022/-ART \$15.00

<https://doi.org/10.1145/3551388>

1 INTRODUCTION

The ability to formally express human behaviors as a computational model enables study and understanding of such behaviors to inform design and implementation of future behavior-aware interfaces [7]. Human behavior is most often purposeful—people perform actions in situations they find themselves in to open up opportunities that allow them to accomplish their goals [6]. People’s enacted behaviors result in behavior instances: sequences of situations people find themselves in and actions they perform in those situations.

Computational models can formalize, classify, and predict complex human behaviors and environments in which the behavior is situated. Computational modeling [67] approaches then enable exploration of such models by simulating situations that people find themselves in and predicting the actions that they take in those situations. In this work, we focus on methods that train models on empirically collected data of behavior instances to predict enacted behaviors (e.g., physical actions) in a given situation. Although we assume that human behavior can be estimated as a computationally rational policy [35, 60], we do not focus on cognitive frameworks [50, 60] and architectures [52] that model the cognitive processes (e.g., task planning) preceding those actions and that do not explicitly train on behavior instances data.

While researchers have traditionally used commodity supervised and semi-supervised Machine Learning (ML) methods to classify and predict human behavior [20, 28, 44, 65, 90], recently, Inverse Reinforcement Learning (IRL) [69] has emerged as a viable alternative computational modeling approach for capturing, exploring, and predicting such behaviors [4, 8]. A form of supervised ML itself, existing IRL approaches have already shown value to the Human-Computer Interaction (HCI) community for modeling human behavior [4, 49, 93, 94, 96] and supporting behavior-aware interfaces (e.g., coaching aggressive drivers [8]). In contrast to Reinforcement Learning (RL) algorithms [48] that compute an optimal policy given *existing, known* rewards, IRL approaches to modeling human behavior [4, 8, 96] leverage historical behavior data (e.g., previously collected behavior instances) to estimate an *unknown* reward function (i.e., the preference that people have for different situations and actions) that could have led to their behavior.

Although large amounts of high-quality training data will almost always result in an accurate IRL model of human behavior, it is not clear how to decide how many behavior instances to collect (especially when collecting such data is challenging or resource intensive). For example, unlike in the driving safety domain where datasets could easily have tens of thousands of behavior instances or more [8], in the healthcare domain participant data pools are often small [43], in particular when modeling behaviors of people with rare conditions and when data collection places significant burden on the participants (e.g., people with Multiple Sclerosis [97]). Not collecting enough data (under-collection) will inevitably results in high uncertainty of model parameters and predictions. Collecting more than necessary amount of data (over-collection) places undue burden on the participants and wastes precious resources and time.

Thus, having a well-informed idea of *how much data to collect* is important for both conservation of resources as well as obtaining an accurate estimate of model parameters. Sample Size Determination (SSD) [2] (i.e., estimating the effects of dataset sizes on the model parameters), which has traditionally focused on power analysis in Null Hypothesis Statistical Testing (NHST) [22, 59], has seen increased application for determining accuracy of parameter estimation [66]. However, despite the HCI community’s interest in quantifying uncertainty for its models (e.g., [91]), no direct applications of existing (mostly frequentist) SSD methods to the domain of modeling human behavior *via* IRL exist. Bayesian IRL methods (e.g., [18]) have potential for estimating the effects of dataset size on the uncertainty of model parameters, but there is no immediately obvious application to SSD. Note that knowing how much data to collect is different from methods for training models from scarce datasets [30, 79, 87, 92], as such methods still give no indication about how much data to collect or how such collected data impacts the uncertainty in the model parameters.

95 In this paper, we seek to quantitatively approach the question of *how much data to collect* through concepts
96 of Uncertainty Quantification (UQ) [36]. UQ is the field concerning the characterization and computation of
97 uncertainty and confidence for models and data in a principled statistical manner, and is used across a wide
98 range of scientific research domains (e.g., engineering physics [70], nuclear weapons stockpile management [74],
99 astronautical engineering [24], medicine and healthcare [9]). Applied to the problem of SSD, UQ could estimate
100 the uncertainty of model parameters and how the uncertainty changes with different dataset sizes (i.e., how
101 increase in number of data samples reduces model parameter uncertainty). Note that in this paper we focus on
102 determining the number of samples (e.g., how many people to collect behavior instances from), rather than the
103 length of individual behavior instances.

104 Here, we present a UQ-based method for estimating how much behavior data to collect to obtain an accurate
105 estimate of behavior model parameters. We do this for a specific IRL algorithm, MaxCausalEnt [95], which has
106 been used to model human behaviors [4, 8, 94, 96] as a Markov Decision Process (MDP) [75]. Our goal is not to
107 compute and declare the optimal number of data points for any specific application problem; instead, our method
108 enables model designers to make an educated decision about the trade-off between resources required to collect a
109 number of data points and the uncertainty in model parameters that will result from that size of data.

110 We cast our problem of SSD [66] in a simulation-based Bayesian experimental design setting [19, 41, 68]. The
111 main insight behind our method is that the probability of model parameters given training data can be updated
112 from prior to posterior through Bayesian inference [10, 11, 51, 85, 88]. We hypothesize that the probability density
113 function (PDF) of the posterior will be narrower (on average) than that of the prior, and that the shrinkage of
114 posterior from the prior, which we quantify using the Information Gain (IG) based on the Kullback-Leibler (KL)
115 divergence [47], will continue to increase as the dataset size increases. Our method leverages this shrinkage to
116 determine whether adding more behavior instances will contribute significant improvements in the accuracy of
117 the model and its parameters.

118 Our method is applicable in two different cases relevant to data collection for modeling behavior:

- 119 • *pre-hoc* experiment design, conducted in the planning stage before any data is collected, to guide the
120 estimation of how many samples to collect; and
- 121 • *post-hoc* dataset analysis, performed after data is collected, to decide if the existing dataset has sufficient
122 samples and whether more data is needed.

123
124 We validate our approach and illustrate its relevance to HCI applications in the domain of healthcare, where
125 SSD is a crucial part of data collection and modeling. We use a realistic example of modeling behaviors of people
126 with Multiple Sclerosis (MS), a progressive autoimmune disease of the central nervous system [40]. In our *pre-hoc*
127 experiment design, we: 1) generate synthetic datasets with various samples of the true model parameters and
128 dataset sizes based on the parameter priors, 2) draw samples from the posterior distributions using Markov Chain
129 Monte Carlo (MCMC) [16, 37], and 3) quantify the IG from the prior to the posterior. We then construct a *pre-hoc*
130 curve that shows how the expected IG (EIG) among synthetic datasets changes as the synthetic dataset size
131 increases. For the *post-hoc* analysis, we perform our computations on an existing real-world MS dataset [53–55]
132 to obtain its realized IG curve and compare it with the *pre-hoc* EIG curve to show that dataset size determined
133 from the *pre-hoc* stage is indeed indicative and can be applied to the MS dataset collection.

134 Having validated our method, we present a procedure that the model designer (i.e., the end-user of our method)
135 can directly use to determine the sample size that results in a desired target EIG. Overall, our method serves as a
136 tool for supporting decision-making; the final decision of exactly how many samples to collect rests with the
137 model designer. In our *pre-hoc* design procedure, the model designer needs only specify: 1) the absolute maximum
138 number of samples allowed by their resource constraints, and 2) a percentage target of resulting maximum EIG.
139 We provided an illustration of SSD for MS behavioral modeling, where with a maximum of 10,000 samples and a
140 target of 80%; our method returned that only 807 samples are needed. Our *post-hoc* dataset analysis procedure
141

allows the model designer to check whether their collected dataset is sufficiently large to have achieved the target EIG. In this procedure, the model designer supplies: 1) the collected dataset, and 2) the target EIG from their *pre-hoc* design. Our method then computes the ratio of the realized IG to the target EIG. In our illustration, we showed that with 707 real-life samples, it already exceeded our EIG target from *pre-hoc* design, and thus no additional samples are needed.

The key contribution of this work is a Bayesian experimental design approach to creating a generalizable SSD method for IRL. Our work enables model designers to perform a deeper, principled investigation of effects of dataset size on model parameters in IRL. It allows for the determination of sample size, one of the hallmark problems of scientific research, as a precursor to building more accurate models of human behavior. Insisting on building accurate models will become particularly important with the rise of behavior-aware user interfaces that automatically reason and act in response to people's behaviors in almost every aspect of their lives.

2 HUMAN BEHAVIOR MODELS SAMPLE SIZE DETERMINATION CHALLENGES

Computational modeling in Human-Computer Interaction (HCI) [7, 72] and behavioral science [89] more broadly aims to use precise mathematical models to make better sense of behavioral data collected from people's personal, mobile, and wearable devices, and their instrumented environments. Such models range from simple regression models (e.g., how people point at targets [5, 86]) to more complicated ML-based models [15] that make predictions about people's behaviors. In recent years, researchers have proposed a number of models of the latter kind that capture purposeful human behaviors (in particular routine behaviors) [4, 26, 27, 61, 64, 80]. Such models are characterized by the need for empirically collected behavior instances data to train the models.

Inverse Reinforcement Learning (IRL) [69] is one such approach that offers a principled way to formalize the definition of purposeful (routine) behaviors [6] and model such behaviors [4]. Unlike other supervised ML methods and data mining approaches [26, 27, 61, 64, 80] that simply correlate people's current situation (or recent past situations) with the next action, IRL methods predict people's next action (in their current situation) that allows them to accomplish a future goal; thus capturing the purposefulness of human behavior [6].

Similar to existing Reinforcement Learning (RL) [48] approaches to modeling human behavior [33, 34, 38, 57, 58, 82], IRL approaches capture human behavior as an MDP [75] and assume that human behavior can be estimated as a computationally rational policy [35, 60]. Lacking the *existing, known* rewards that are present in RL, IRL algorithms optimize the parameters of the model based on history of behavior instances from the data (each guided by a policy that specifies which actions to perform in which situations), to estimate an *unknown* reward function (the preference that people have for specific situations and actions) that could have led to their behavior. Thus, policies used in both RL and IRL imitate human behaviour because such policies emerge as solutions to optimization problems that maximize long-term cumulative reward given cognitive and task bounds [21, 72].

However, lack of precise rules for determining how much data we need to accurately estimate the parameters of such behavior models remains one of the challenges when modeling human behavior *via* IRL. Although SSD methods [66] have seen application for estimating the number of samples in related field of supervised ML classification [14, 29, 32], such methods have not seen much application to IRL, or more specifically, in the domain of modeling human behavior *via* IRL.

Note that methods that aid training models from scarce datasets [30, 79, 87] do not provide information if the amount of data they train on is enough to train a model. Artificially increasing collected dataset size to fix class imbalance in datasets (e.g., using oversampling [3]) or to reduce overfitting (e.g., using data warping [45]) gives no indication about how much data to generate. For example, using Generative Adversarial Networks [92] to generate more records, simply samples from narrow probability distributions built on existing small sample sizes, thus increasing the coverage around the existing data, but not in underexplored data spaces. Active Learning approaches [1, 84] compute uncertainty over new data samples to find a subset of data that requires

189 labeling. However, data scarcity challenges are different from data labeling challenges, where massive amount of
 190 unlabelled data is readily available.

191 Uncertainty Quantification (UQ) methods [36] hold potential to aid in SSD by estimating the uncertainty of
 192 model parameters and how the uncertainty changes with different dataset sizes (i.e., how increase in number
 193 of data samples reduces model parameter uncertainty). While the literature often refers UQ to be the *forward*
 194 propagation of uncertainty from inputs (uncertainty sources) to outputs, SSD emerges as part of the *inverse* UQ
 195 problem: given observation data, what is the resulting uncertainty on model parameters in the presence of this
 196 evidence? The Bayesian inference formalism [10, 11, 51, 85, 88], which seeks to compute the posterior probability
 197 of unknown parameters conditioned on given observation data, offers a mathematically rigorous approach for
 198 solving this inverse problem. The Bayesian framework is particularly suitable for SSD [62] in situations of sparse
 199 and noisy data, where the residual uncertainty in the model remains non-negligible. Yet, it is not immediately
 200 clear how to apply such a method to the problem of SSD when modeling human behavior *via* IRL, which we
 201 explore in our work.

202 3 MODELING HUMAN BEHAVIOR VIA INVERSE REINFORCEMENT LEARNING

203 In this work, we focus on sample size determination for a specific model of human behaviors [4], which uses
 204 the MaxCausalEnt algorithm [95] to estimate model parameters from behavior instances data. Thus, we start
 205 by describing the classical (non-Bayesian) construction of that IRL model, followed by a presentation of the
 206 Bayesian framework [10, 11, 51, 85, 88] for inferring model parameters with quantified uncertainty given datasets
 207 of various sizes in Section 4.

208 3.1 Modeling Human Behavior with Markov Decision Process

209 Following [4], we formalize purposeful behaviors as a MDP [75], given by a tuple:

$$210 \mathcal{M}_{MDP} = \{\mathcal{S}, \mathcal{A}, R(s, a), P_0(s), P(a | s), P(s' | s, a)\}. \quad (1)$$

211 Here $s \in \mathcal{S}$ is the state from a finite state space representing different situations that a human participant (i.e.,
 212 agent) can be in, and $a \in \mathcal{A}$ is the action from a finite action space that the participant can take. The reward
 213 function $R(s, a)$ defines the reward that the participant incurs when performing an action a while in the state s .
 214 The reward function thus represents the participant preference for being in various situations and for taking
 215 different actions in those situations.

216 The initial state probability $P_0(s)$ captures the probability that a state initiates a behavior instance. The
 217 conditional probability of actions given states $P(a | s)$ represents the probability that the participant will choose
 218 to perform a particular action a in a particular state s . The action-dependent transitional probability $P(s' | s, a)$
 219 designates the probability of transition into the next state s' after the participant performs action a in state s . The
 220 transitional probability thus captures how the actions that participants choose in various situations influence
 221 their surroundings. However, since participants rarely have full control over their environments, the transition
 222 probability also encodes how the environment changes at each step independent of the participants' actions.

223 3.2 Estimating Model Parameters via MaxCausalEnt IRL Algorithm

224 In an IRL setting, $P_0(s)$, $P(s' | s, a)$, $P(a | s)$, and $R(s, a)$ are all unknown, and the goal is to estimate them from an
 225 available dataset of observed behavior instances, where each behavior instance sample is defined as a sequence
 226 of state and action pairs over time: $\left\{ \left(s_1^{(i)}, a_1^{(i)} \right), \left(s_2^{(i)}, a_2^{(i)} \right), \dots, \left(s_T^{(i)}, a_T^{(i)} \right) \right\}$, where $i = 1, \dots, n_d$ denotes the i th
 227 sequence sample of length T in the dataset for a total of n_d samples. Note that we can collect more than one
 228 behavior instance from a single person in our dataset. For now, we assume all sequences have the same length T .

Following the framework introduced in [4], $P_0(s)$ and $P(s' | s, a)$ can be estimated by constructing two separate Bayesian networks [12] from the dataset samples. Furthermore, we define a linear parametric reward function:

$$R(s, a) = \theta^T \mathcal{F}_{s,a}, \quad (2)$$

where $\mathcal{F}_{s,a}$ is a general feature vector and θ is a vector of unknown weight parameters representing the strength of preferences for individual features. MaxCausalEnt IRL algorithm [95] then estimates the parameter θ of the reward function $R(s, a)$. The algorithm ensures that the reward function $R(s, a)$ relates to the stochastic policy $P(a | s)$, so that the probability of each action is proportional to the sum of rewards over a sequence of future states and actions starting at the next transition state. It does so by computing the policy probability $P(a | s)$ that maximizes the ‘‘causal’’ entropy $H(a_{1:T} || s_{1:T})$ (please see [95] for the precise definition, we altered the superscript notation here to a subscript to avoid confusion with the transpose operation):

$$\operatorname{argmax}_{\substack{P(a | s), \\ a \in \mathcal{A}, s \in \mathcal{S}}} H(a_{1:T} || s_{1:T}) \quad (3)$$

such that

$$\begin{aligned} \mathbb{E}_{P(s,a)} [\mathcal{F}_{s,a}] &= \mathbb{E}_{\tilde{P}(s,a)} [\mathcal{F}_{s,a}] \\ P(a | s) &\geq 0, \quad \forall s \in \mathcal{S}, a \in \mathcal{A} \\ \sum_{a \in \mathcal{A}} P(a | s) &= 1, \quad \forall s \in \mathcal{S}. \end{aligned}$$

Most importantly, the first constraint imposes the expected feature counts $\mathbb{E}_{P(s,a)} [\mathcal{F}_{s,a}]$ computed from the model using the estimated policy $P(a | s)$ matches the empirical expected feature counts $\mathbb{E}_{\tilde{P}(s,a)} [\mathcal{F}_{s,a}]$ observed in the dataset, thus ensuring the model to capture the policy that guided the behavior instances in the data [69].

MaxCausalEnt IRL algorithm [95] solves the optimization problem in Equation (3) using Stochastic Gradient Descent (SGD) [12] over the reward function parameter θ . At each gradient step, it iteratively computes action-based value function $Q_\theta^{\text{soft}}(s, a)$ that represents the expected value of performing a specific action a in a specific state s , and state-based value function $V_\theta^{\text{soft}}(s)$ that represents the expected value of being in a specific state s :

$$Q_\theta^{\text{soft}}(s_t, a_t) = \sum_{s_{t+1}} P(s_{t+1} | s_t, a_t) V_\theta^{\text{soft}}(s_{t+1}) + \theta^T \mathcal{F}_{s_t, a_t} \quad (4)$$

$$V_\theta^{\text{soft}}(s_t) = \operatorname{softmax}_{a_t} \{Q_\theta^{\text{soft}}(s_t, a_t), c\} \quad (5)$$

where $\operatorname{softmax}_x \{f(x), c\} := \frac{1}{c} \ln \sum_x e^{cf(x)}$ and c is a hyperparameter. These two value functions are then used to compute the policy *via* following equation:

$$P(a_t | s_t) = \frac{1}{Z(s_t)} e^{c(Q_\theta^{\text{soft}}(s_t, a_t) - V_\theta^{\text{soft}}(s_t))}, \quad (6)$$

where $Z(s_t) = \sum_{b \in \mathcal{A}} e^{c(Q_\theta^{\text{soft}}(s_t, b) - V_\theta^{\text{soft}}(s_t))}$ ensures the overall expression is a proper probability mass function. The

SGD algorithm uses the stochastic policy $P(a_t | s_t)$ in a forward pass to update the estimated expected feature counts $\mathbb{E}_{P(s,a)} [\mathcal{F}_{s,a}]$, and updates the parameter θ using the following equation until convergence:

$$\Delta \theta = \mathbb{E}_{P(s,a)} [\mathcal{F}_{s,a}] - \mathbb{E}_{\tilde{P}(s,a)} [\mathcal{F}_{s,a}]. \quad (7)$$

Overall in the MaxCausalEnt IRL algorithm [95], the dataset, and the number of sequence samples n_d in the dataset, affects the estimation of $P_0(s)$, $P(s' | s, a)$, and $\mathbb{E}_{\tilde{P}(s,a)} [\mathcal{F}_{s,a}]$. While MaxCausalEnt IRL provides an efficient method for estimating θ , it only produces a single point estimate and does not offer a measure

of uncertainty or confidence surrounding the estimated value that results from n_d . This difficulty is further compounded by the highly non-concave nature of the causal entropy objective, leading to local maxima and non-unique solutions in θ (equivalently, multimodal distributions [76]). As a result, in the above framework, it is very challenging to quantify the estimation quality, and the potential improvement of estimation quality when more data samples are acquired. In the next section, we introduce the Bayesian inference approach that computes the full probability distribution of θ , thus allowing a rigorous quantification of the uncertainty in estimating θ . We will then use metrics derived from the uncertainty of θ to guide the determination of sample size n_d .

4 METHOD FOR SAMPLE SIZE DETERMINATION

We introduce a Bayesian approach to quantify the effects of dataset size (i.e., number of collected data samples) on the uncertainty of parameter estimation for a specific IRL-based model of human behavior [4] that we described in the above section. Once we quantify these effects, we can compute the number of data samples needed to achieve some targeted estimation performance or quality, or when the cost of collecting additional data outweighs the added benefit. Our method can guide model designers to decide how much data to collect, or to assess if an existing dataset contains sufficient data samples.

4.1 Bayesian Inference for Quantifying Parameter Uncertainties

Bayesian inference [10, 11, 51, 85, 88] is a framework that provides a rigorous quantification of uncertainty *via* the formalism of probability theory. It is particularly suitable for incorporating sparse, noisy, and incomplete data from different sources, and a versatile entryway to inject domain knowledge, historical data, and opinions from subject matter experts. Performing Bayesian inference for the IRL problem is known as Bayesian IRL (BIRL) [76]. BIRL has been studied in many different use-cases, including modeling the behavior of distinct conversational agents in a virtual environment [78], automatically generating trajectories for active learning from critiques [25], and evaluating the upper bound of the policy loss of IRL [17]. Such existing work attempts to directly infer a distribution for $R(s, a)$, which has a dimensionality equal to the cardinality of $\mathcal{S} \times \mathcal{A}$, typically prohibitive for most Bayesian inference algorithms in practice.

Instead, we proceed in a manner similar to the previously described MaxCausalEnt approach to use a linear parameterized form of $R(s, a)$ in Equation (2), and then infer θ . We build upon an existing parametric BIRL work [42], which previously illustrated only on a 4-dimensional θ . Under the Bayesian framework, we treat θ as a (continuous) random vector with an associated PDF. Given a dataset with n_d behavior instances (sequence samples) $D = \left\{ \left(s_t^{(i)}, a_t^{(i)} \right) \right\}, t = 1, \dots, T, i = 1, \dots, n_d$, the uncertainty of our unknown model parameter θ is updated via Bayes' rule:

$$p(\theta | D) = \frac{P(D | \theta)p(\theta)}{P(D)}, \quad (8)$$

where $p(\theta | D)$ is the posterior PDF¹, $p(\theta)$ is the prior PDF, $P(D | \theta)$ is the likelihood function (i.e., the probability density of having observed the state and action trajectories in D if the true feature weights were θ), and $P(D)$ is the model evidence (a normalization constant with respect to θ). The prior thus represents the “before-data” uncertainty, and the posterior is the “after-data” uncertainty. Solving the Bayesian inference problem entails characterizing the posterior $p(\theta | D)$. We note that the evidence (also known as the marginal likelihood) $P(D) = \int_{\Theta} P(D | \theta)p(\theta) d\theta$ is an intractable quantity and expensive to numerically approximate, but can be avoided altogether by employing MCMC algorithms [16] that generate samples from the posterior distribution.

¹We use lower case $p(\cdot)$ for PDF of a continuous random variable or vector, and upper case $P(\cdot)$ for probability mass function of a discrete random variable or vector.

For a Bayesian inference problem, we generally have the ability to evaluate the prior and likelihood PDFs. Usually the model designer would select the prior, which represents their knowledge or belief about θ before having seen any data. For example, a prior PDF with finite support brings bound constraints to the parameter value, a non-informative maximum entropy prior indicates maximum initial ignorance and minimal assumptions [46], and an informative prior can incorporate domain knowledge, previous experience, and expert opinions [71].

The likelihood function can be directly derived using the Markovian structure of the MDP model with the assumption of conditional independence among different observed behavior instances [76]:

$$\begin{aligned}
P(D | \theta) &= \prod_{i=1}^{n_d} P \left(\left\{ (s_t^{(i)}, a_t^{(i)}) \right\}_{t=1}^T \mid \theta \right) \\
&= \prod_{i=1}^{n_d} P \left(s_1^{(i)} \right) P \left(a_1^{(i)} \mid s_1^{(i)} \right) P \left(s_2^{(i)} \mid s_1^{(i)}, a_1^{(i)} \right) \cdots P \left(s_T^{(i)} \mid s_{T-1}^{(i)}, a_{T-1}^{(i)} \right) P \left(a_T^{(i)} \mid s_T^{(i)} \right) \\
&= \prod_{i=1}^{n_d} P \left(s_1^{(i)} \right) \left(\prod_{t=2}^T P \left(s_t^{(i)} \mid s_{t-1}^{(i)}, a_{t-1}^{(i)} \right) \right) \left(\prod_{t=1}^T P \left(a_t^{(i)} \mid s_t^{(i)} \right) \right) \\
&= \prod_{i=1}^{n_d} P \left(s_1^{(i)} \right) \left(\prod_{t=2}^T P \left(s_t^{(i)} \mid s_{t-1}^{(i)}, a_{t-1}^{(i)} \right) \right) \left(\prod_{t=1}^T \frac{e^{c(Q_\theta^{\text{soft}}(s_t^{(i)}, a_t^{(i)}) - V_\theta^{\text{soft}}(s_t^{(i)}))}}{\sum_{b \in \mathcal{A}} e^{c(Q_\theta^{\text{soft}}(s_t^{(i)}, b) - V_\theta^{\text{soft}}(s_t^{(i)}))}} \right) \\
&= K_{[\sim \theta]} \prod_{i=1}^{n_d} \prod_{t=1}^T \frac{e^{c(Q_\theta^{\text{soft}}(s_t^{(i)}, a_t^{(i)}) - V_\theta^{\text{soft}}(s_t^{(i)}))}}{\sum_{b \in \mathcal{A}} e^{c(Q_\theta^{\text{soft}}(s_t^{(i)}, b) - V_\theta^{\text{soft}}(s_t^{(i)}))}},
\end{aligned} \tag{9}$$

where c is the softmax hyperparameter, and $K_{[\sim \theta]}$ collects all terms that do not depend on θ . Note that $K_{[\sim \theta]}$ does not need to be computed if MCMC is used since this term is a constant with respect to θ . However, the posterior still depends on the transition probability since it enters through the policy computation in the exponential terms, as seen in Equation (4).

We note that the likelihood here differs from classical Bayesian likelihoods often designed to capture measurement noise or model inadequacy, such as those discussed in [51]. Instead, the likelihood in Equation (10) (and in the current BIRL literature) stems from the stochasticity of the policy $P(a | s)$. Consequently, this formulation has a limitation that inherently assumes the dataset is free of measurement noise and the model is absent of error. Incorporating these factors is non-trivial, and requires substantial new formulations to the BIRL framework. Thus in this work, we follow existing BIRL formulation, and leave such improvements as future work.

Lastly, once we obtain the posterior $p(\theta | D)$ or its samples, we can propagate this uncertainty to any other θ -dependent quantities of interest in the model—these are known as *posterior-predictive distributions*. For example, model designers and domain experts may be interested in the uncertainty of specific policy probabilities $P(a^* | s^*)$, for some subset of a^* and s^* of interest, resulting from the residual posterior uncertainty in θ . This can be achieved numerically via Monte Carlo sampling [77] from the MCMC posterior samples of θ .

4.2 Quantifying Data Scarcity

Following the Bayesian framework introduced in the previous section, we quantify data scarcity based on the extent of uncertainty reduction on θ . Specifically, we employ the KL divergence [56] from the prior to the posterior

(i.e., IG on θ from the data D):

$$D_{\text{KL}}(p(\theta | D) || p(\theta)) = \int_{\Theta} p(\theta | D) \ln \left[\frac{p(\theta | D)}{p(\theta)} \right] d\theta. \quad (11)$$

The KL divergence is non-negative, and equals zero if and only if $p(\theta | D) = p(\theta)$. It provides a measure of dissimilarity between two probability distributions and is strongly rooted in information theory [23]. Furthermore, since SSD is desired before data collection, D would not be available yet at this *pre-hoc* stage. Thus, we need to take the expectation over all possible realizations of D , to arrive at the final *expected KL divergence* (or EIG):

$$\begin{aligned} \text{EIG}(n_d) &= \mathbb{E}_D [D_{\text{KL}}(p(\theta | D) || p(\theta))] \\ &= \sum_{D \in \{(S_t^{(i)}, \mathcal{A}_t^{(i)})_{t=1}^{n_d}\}} \int_{\Theta} p(\theta | D) \ln \left[\frac{p(\theta | D)}{p(\theta)} \right] d\theta P(D), \end{aligned} \quad (12)$$

where we explicitly show the dependence on n_d . This quantity is also known as the *expected utility* in Bayesian experimental design [19, 41, 68] and often used as the criterion to be maximized in statistical designs of experiments.

In general, the KL divergence has no closed-form and must be approximated numerically. We adopt a Monte Carlo estimator using posterior samples generated from the MCMC algorithm:

$$\text{EIG}(n_d) \approx \frac{1}{LM} \sum_{j=1}^L \sum_{k=1}^M \left[\ln p(\theta^{(j,k)} | D^{(j)}) - \ln p(\theta^{(j,k)}) \right], \quad (13)$$

where L and M are the Monte Carlo sample sizes in this estimator. Specifically, L is the number of realizations of the data D that we synthetically generate, and M is the number of posterior θ samples produced from MCMC given each of these data realizations. We choose M so that MCMC reasonably converges, and L based on available computational resources. We emphasize that these Monte Carlo sample sizes L and M are purely for numerically estimating the EIG, and should not be confused with the sample size n_d that is the number of behavior instances to be collected (which affects the dimension of D). $D^{(j)}$ is then the j th Monte Carlo sample drawn from $P(D)$, generated by sampling the prior $\theta' \sim p(\theta)$ then the likelihood $D^{(j)} \sim P(D | \theta')$ conditioned on the prior sample. More specifically, to generate each (i th) behavior instance within the j th data realization $D^{(j)}$ that corresponds to θ' , we sample the behavior instance's states from the transition probability and actions from policy computed using Equations (4), (5), and (6) given θ' . Lastly, $\theta^{(j,k)}$ is the k th MCMC sample drawn from $p(\theta | D^{(j)})$. However, we do not have the ability to evaluate the posterior PDF $p(\theta | D)$ since MCMC produces samples but not PDF values. Thus, we use kernel density estimation (KDE) [83] to approximate the PDF from the MCMC samples of θ .

4.3 Determining Sample Size

Intuitively, one expects EIG to increase as more data is collected (i.e. with large n_d), but the additional benefit from each new sample generally diminishes as the overall dataset grows. Finding the critical point where the rate of benefit is no longer worthwhile can be valuable in guiding the decision-making of SSD.

In the *pre-hoc* experiment design, the model designer (i.e., the experimenter) seeks to estimate the number of samples to collect. We propose a procedure to arrive at this number, where the experimenter needs to supply *two* parameters. The first parameter is the absolute maximum number of samples n_d^{max} allowed by the resource constraints (e.g., time, funding, number of people in the target population). Here we require the specification of the maximum allowable samples instead of a target EIG since the former is more intuitive and tangible, and relatively easier to determine for model designers (e.g., divide maximum resources by unit cost of each data sample). The second parameter is the target percentage $p \in (0, 100]$ of the maximum EIG (i.e., the EIG corresponding to n_d^{max}) that the model designer wishes to achieve (e.g., 80%, 90%). This parameter allows the flexibility of conserving

resources since often we do not wish to expend all available resources on a single data-collection campaign, and for the model designer to exercise their valuation of information-versus-cost trade-off. The dataset sample size $n_d(p)$ for achieving p percent of the maximal EIG can then be solved *via*:

$$\text{EIG}(n_d(p)) = \frac{p}{100} \text{EIG}(n_d^{\max}). \quad (14)$$

In the *post-hoc* dataset analysis, the model designer can calculate if the number of samples in their collected dataset is sufficiently large. For example, we can compute the percentage of the maximal EIG reached by the realized *post-hoc* IG, and compare it with the desired percentage p . If this *post-hoc* percentage is equal or greater than p , then we have collected sufficient data samples to achieve our target EIG. We note that it is possible for the *post-hoc* percentage to exceed 100%, since the realized *post-hoc* IG is for the particular collected dataset, while the *pre-hoc* EIG is the expectation over all possible datasets.

We emphasize that our method is a general decision-support framework, not decision-making system. The precise decision-making rules and criteria (e.g., choice of n_d^{\max} and p) will depend on many additional factors such as the specific goals of the model and data usage, monetary cost and scheduling of experiments, value of information and knowledge, consequences and risks, regulatory and policy requirements, and even the degree of risk-aversion of the experimenter. Incorporating these components systematically and comprehensively, as is pursued in the research of decision theory (e.g., [10, 73]), is challenging and beyond the scope of our work.

5 ILLUSTRATING APPLICATION OF OUR SAMPLE SIZE DETERMINATION METHOD

We demonstrate our *decision support* method for SSD on a behavior model of people with Multiple Sclerosis (MS) [97]. People with MS experience physical impairment and chronic pain, fatigue, depressed mood, and cognitive problems. Such symptoms relate to numerous negative outcomes, including unemployment, disability, social impairment, life dissatisfaction, interference with daily activities, deterioration of general mental and physical health, and lack of community integration. Thus, interventions that target the most severe or impactful symptoms, such as pain and fatigue, could improve people's healthcare outcomes and their quality of life by guiding the timing of medications and selection of behavior-based self-management strategies.

Such a behavior model is of high interest to the clinicians to test the hypothesis that participants' activities of daily living that we can sense, identify, and collect can predict their pain, fatigue, and overall well-being. Therefore, our goal is to illustrate how to provide guidance to model designers, *via* a mathematically principled framework, in deciding how many samples (i.e., how many behavior instances of people with MS) to collect for learning people's behaviors. We leave the application of IRL to modelling the MS data including validation of such models and any resulting interventions for future work, which can only be done after we have estimated the required number of samples to train our models. We illustrate the effects of dataset size under both *pre-hoc* and *post-hoc* scenarios.

5.1 Model of Behaviors of People with Multiple Sclerosis

To model behaviors of people with MS, we followed the modeling approach from [4] and used the MDP framework described in Section 4. For this investigation, we consulted a domain expert from our institution, who is a Research Non-Clinical Psychologist in Physical Medicine and Rehabilitation specializing in MS, for formulation of the initial state and state transition probabilities. Here, we describe how we model different aspects of behaviors of people with MS, and in the following sections we provide details about how we perform the *pre-hoc* experiment design and *post-hoc* dataset analysis.

5.1.1 States and Actions. We define our state space \mathcal{S} and action space \mathcal{A} by defining state and action features (Tables 1 and 2). These features in turn define feature vectors $\mathcal{F}_{s,a}$, which are designed to be one-hot-encoding of all possible state and action pairs. The state features (Table 1) describe participants' demographics (gender

and age), contextual information such as the time of day (wake, morning, afternoon, evening, and bed), self-reported symptoms and health indicators (i.e., momentary assessment of pain and fatigue at each time of day, and self-reported positive affect and well-being (PAW) [81]), and information about their last objective measures of activity intensity (measured as activity bouts) and pace (measured as the number of breaks participants take while performing an activity).

The action features (Table 2) represent participant objective measure of activity intensity and pace (e.g., as measured by an ActiGraph watch) and whether or not a participant filled out the momentary assessment of their symptoms and their PAW (i.e., end-of-day functional outcome) at each time of day. Note that people with MS do not have full control over their symptoms and their healthcare outcomes—they only have control over their decision to record them or not. We followed the same time of day intervals from the original dataset [53, 55] according to common momentary assessment times for this participant population.

Table 1. State features that define the different situations that a participant with MS can be in.

State Feature	Description
Gender	Gender of the people with MS {Male, Female}
Age	Age of the people with MS {Younger than 30, Between 30 to 60, 60 and older}
Current Daytime Interval	Time of the day {Wake, Morning, Afternoon, Evening, Bed}
Current Pain	Current interval pain score {Low, Medium, High, Not Recorded}
Current Fatigue	Current interval fatigue score {Low, Medium, High, Not Recorded}
Last Activity Bouts	Last interval activity bouts based on average activity bouts per 15s from an ActiGraph watch {Low, Medium, High, Not Recorded}
Last Activity Pace	Last interval pace (determines whether last activity was performed with/without breaks) from an ActiGraph watch {Low, Medium, High, Not Recorded}
End-of-Day Positive Affect and Well-being	Positive impact on bed interval signifies how much positive impact (sense of well-being, feeling hopeful and satisfying, cheerful, etc.) the participants with MS had on that particular day (recorded at bedtime only) {None (Not Applicable), Moderate, Mild, Normal, Not Recorded}

5.1.2 Behavior Instances. We define a behavior instance as a sequence of states and actions that captures situations that a particular participant found themselves in and the actions they performed in those situations, in each single day. Thus, we treat each participant’s day as *one sample* of a total of n_d samples (n_d thus equals the number of participants times the number of days they participated in the study). Because each behavior instance represents behaviors of a single participant, *Gender* and *Age* variables remain the same throughout a behavior instance, time of day advances at each transition, and the rest of the variables change either in response to the participant’s actions, or depending on the changes in symptoms and healthcare outcomes between different times of day.

5.1.3 Initial State Probabilities. Each behavior instance starts with an initial state s with probability $P_0(s)$, where *CurrentDaytimeInterval* of s is set to *Wake*. To estimate initial state probabilities $P_0(s)$, we build a Bayesian network [12] (Figure 1.a). We consider 6 features that describe each state: age, gender, pain level, fatigue level, last activity bout, and last pace. Note that, all initial states start when a person with MS wakes

Table 2. Action features representing different actions that a participant with MS can perform.

Action Feature	Description
Current Activity Bouts	Current interval activity bouts based on average activity bouts per 15s from an ActiGraph watch {Low, Medium, High, Not Recorded}
Current Activity Pace	Current interval pace (determines whether current activity was performed with/without breaks) {Low, Medium, High, Not Recorded}
Record Next Pain	Status of next state pain {Recorded, Not Recorded}
Record Next Fatigue	Status of next state fatigue {Recorded, Not Recorded}
Record Next Positive Affect and Well-being	Status of next state positive affect and well-being {Not Applicable, Recorded, Not Recorded}

up; therefore, we only consider states with $Current_Daytime_Interval = Wake$ in our probability calculations (participants did not report End-of-Day PAW at wake time), and set all other initial state probabilities to 0. Age and $Gender$ both influence $Pain$ and $Fatigue$, which are mutually independent, as are Age and $Gender$. On the other hand, $LastAc bout$ and $LastPace$ only influence level of $Fatigue$ and they ($LastAc bout$ and $LastPace$) are mutually independent except for the *Not Recorded* case. If $LastAc bout$ is *Not Recorded*, then $LastPace$ will also be *Not Recorded*, and vice-versa. Thus, we compute the initial state probability as follows:

$$\begin{aligned}
 P_0(s) &= P(Age, Gender, Pain, Fatigue, LastAc bout, LastPace) \\
 &= P(Pain|Age, Gender) \times P(Fatigue|Age, Gender, LastAc bout, LastPace) \\
 &\quad \times P(Age) \times P(Gender) \times P(LastAc bout) \times P(LastPace).
 \end{aligned} \tag{15}$$

where $s \in \mathcal{S}_0$ and \mathcal{S}_0 represents all possible initial states.

5.1.4 State Transition Probabilities. Each state transition is driven by the action that the participant performs and changes in participants' symptoms irrespective of their actions, as influenced by their demographics, time of day, and previously reported symptoms and healthcare outcomes. To capture the probability of these transitions, we built another Bayesian network (Figure 1.b) to estimate state transition probabilities $P(s' | s, a)$. Here, we used all of the features from current state s and action a to estimate the joint probability of pain, fatigue, and positive affect and well-being in state s' , which corresponds to the probability of next state s' . This is because all of the other state features are deterministic: $DaytimeInterval$ transitions are fixed, Age and $Gender$ stay the same at each transition, and values of $LastActivityBouts$ and $LastActivityPace$ in s' are the same as in action a . We drew edges between nodes in the Bayesian network in Figure 1.b according to instructions from the domain expert.

5.1.5 Stochastic Action Policy. In this model, the participants decide on their next action based on a reward function $R(s, a)$ (Equation (2)), which represents the preference that people with MS have for certain situations (e.g., specific pain and fatigue levels) and performing certain actions in those situations. Given model parameters θ , we can compute the conditional probability of participants' actions given their current situation $P(a | s)$. This stochastic policy captures the probability of each participant action relative to their preference for different features of states and actions.

5.2 Computational Setup for BIRL

To demonstrate and validate our method, we used an existing MS dataset [53–55] containing 749 behavior instances collected from a total of 107 participants with MS. To illustrate *pre-hoc* experimental design and *post-hoc*

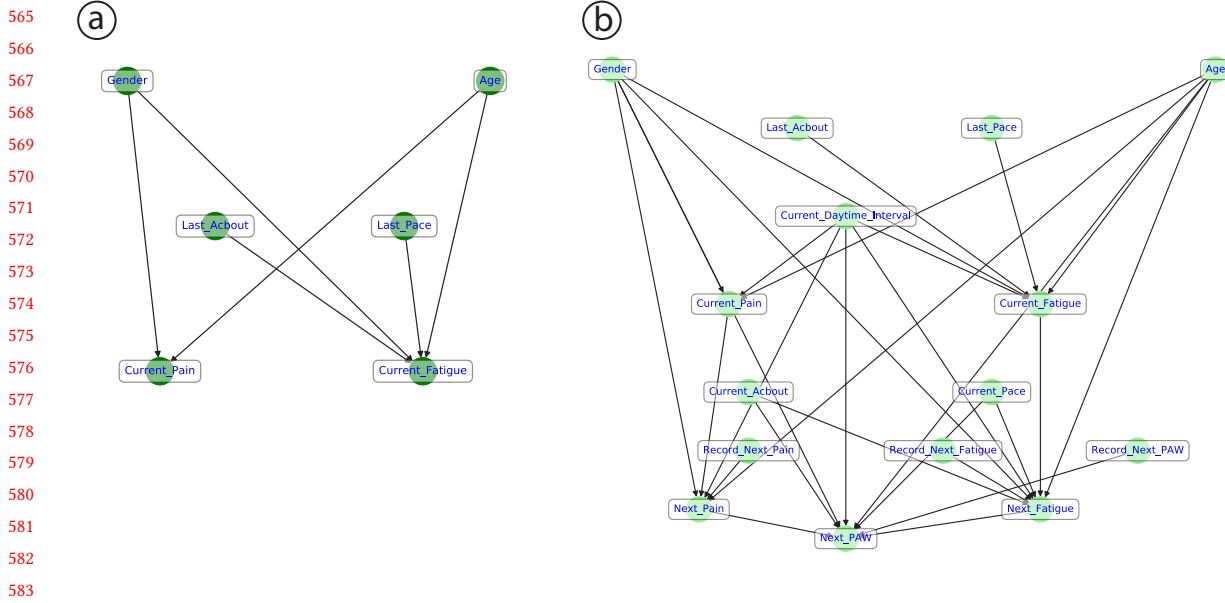


Fig. 1. Bayesian networks for computing: a) initial state probabilities $P_0(s)$, and b) state transition probabilities $P(s' | s, a)$.

dataset analysis, we split this dataset into two subsets: 1) *pre-collected* dataset with 6 participants resulting in 42 behavior instances (i.e., samples), and 2) *post-collected* dataset with 101 participants, or 707 samples. This split mimicked common practice of piloting a data collection study before running the main data collection. We used the pre-collected dataset to estimate the initial state and transition probabilities of the MDP model, but not for building an informative prior; thus avoiding leaking information between decision points.

We kept these two subsets strictly separate to prevent any statistical contamination (i.e., “cheating”) between *pre-hoc* experimental design and *post-hoc* dataset analysis. We judiciously chose the 6 pre-collected participants to include one person in each of the three age groups and for each gender. The remaining 101 participants followed the original data collection study [53–55] inclusion and exclusion criteria.

We then performed *pre-hoc* experimental design for SSD of the main data collection using only the 42 pre-collected samples (and without using or seeing any of the 707 post-collected samples). We then mimicked the main data collection with the remaining 707 samples and conduct a *post-hoc* dataset analysis to update the model uncertainty and provide guidance on whether additional samples are needed.

5.2.1 Estimating Initial State and Transition Probabilities. We used the pre-collected dataset to estimate the initial state probability and transition probabilities. Since our 42 pre-collected samples in the *pre-hoc* experimental design came from 6 participants with different combinations of age and gender each, the samples were not representative of the distributions of age and gender in the study population. We thus consulted our domain expert (who collected the original dataset [53–55]) to provide an estimate of the marginal distributions of age and gender groups envisioned for the targeted study group, and used these values to appropriately scale the initial state and transition probabilities extracted from the Bayesian networks; the scaling factors supplied by the expert are shown in Table 3. Note that the post-collected dataset protocol used those same marginal distributions to control the ratio of age and gender in the dataset, thus showing excellent agreement.

Table 3. Participant distribution of age and gender suggested by an expert compared to the observed distribution from the post-collected dataset.

State Features	Descriptors	Expert's Scaling	Post-collected Dataset Distribution
Age	60 and Older	0.10	0.09
	Between 30 to 60	0.80	0.80
	Younger than 30	0.10	0.11
Gender	Female	0.70	0.70
	Male	0.30	0.30

5.2.2 Bayesian Prior Selection. Having introduced the model and data, we now define the prior distribution for our model parameters. Without any initial knowledge about correlation between different θ , we prescribe an independent prior for θ :

$$p(\theta) = \prod_i^{n_\theta} p(\theta_i), \quad (16)$$

where n_θ is the total number of feature coefficients. Furthermore, we endow each $p(\theta_i)$ to have a weakly-informed structure, using a truncated normal distribution with mean 0.5 and truncated outside $[0, 1]$. The truncated normal distribution thus imposes hard constraints for θ_i to remain within 0 and 1 and also injects a preference for region near the center of this interval—these regularize the non-uniqueness effects of θ (i.e., multiple θ values may produce the same policy). Truncating the θ space also helps the MCMC sampling to be more stable and efficient. As shown in Figure 2, we will compare priors with 5 different standard deviations, where the smallest standard deviation represents the most informative prior, and the largest standard deviation provides near-maximum entropy and approximates the uniform distribution. We discuss the impact of different priors on the results of EIG, and then select one for the remainder of the paper.

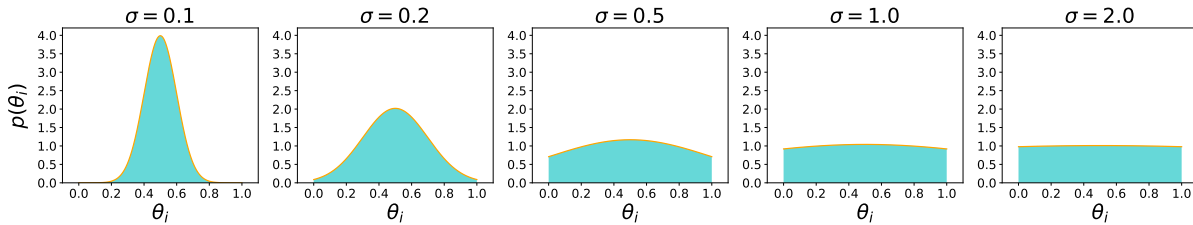


Fig. 2. The prior PDF for θ_i is a truncated normal distribution with mean 0.5, standard deviation σ , and truncation between 0 and 1. We tested the effects of prior choice under different values of σ above.

5.2.3 Estimating Model Parameters and Corresponding Stochastic Action Policy. Participants' reward function, and the model parameters θ we use to compute it, are not known ahead of time. Existing IRL approaches use the MaxCausalEnt IRL algorithm [95] (Equation (6)) to estimate feature weights to compute a reward function and the resulting policy. In our illustration, we apply our BIRL method from Section 4.1 to compute the Bayesian posterior distribution of θ . Then, for any given θ from this distribution, we can also obtain the corresponding policy *via* Equations (4), (5) and (6). Note that a converged policy is not myopic; it is dependent on the cumulative future reward (i.e., Q and V functions), not only on the immediate reward.

5.3 Pre-hoc Experimental Design and Post-hoc Dataset Analysis

We present results of the *pre-hoc* experiment design and *post-hoc* dataset analysis for the MS dataset. We first compute the EIG as a function of data sample size n_d . We then illustrate how an optimal sample size can be determined from an example decision-making rule. Here, we estimate how many samples to collect for different criteria of data collection. Finally, we illustrate the *post-hoc* calculation of the realized IG from collected dataset.

We employed the affine invariant MCMC ensemble sampler [39] to sample from the posterior distribution $p(\theta | D)$, specifically the Emcee Python package version 3.0 [31], for its ability to parallelize and explore multimodal distributions. We performed all EIG calculations using the Monte Carlo estimator in Equation (13), with $L = 60$ and $M = 92,000$ (92 ensemble MCMC chains in parallel, each with 1,000 samples). We discarded the first 20% of the chains as burn-in. In order to determine a suitable chain length of MCMC, we compared MCMC results with: i) 1,000 MCMC samples versus ii) continuing the chain longer until 10,000 MCMC samples. We do this on the same testing case and plot their corresponding IG curves in Figure 3. These two curves match well with each other, which suggests that 1,000 MCMC samples is sufficiently converged for estimating the IG for this problem, and running a longer MCMC chain is not likely to make a significant difference.

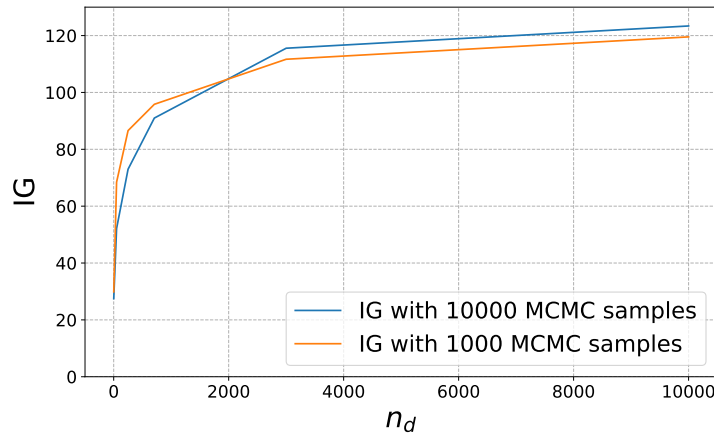


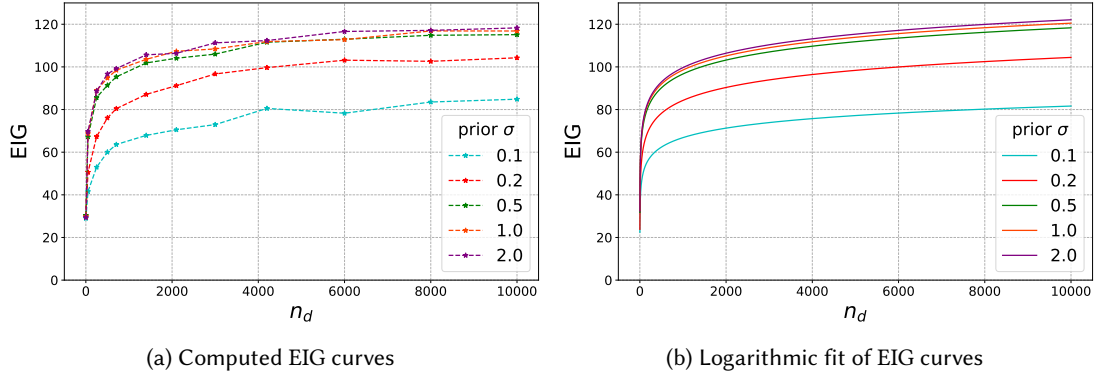
Fig. 3. IG curves under different number of MCMC samples (i.e. chain lengths).

5.3.1 Pre-hoc Experiment Design. In *pre-hoc* experiment design, we seek to construct the EIG curve and use it for SSD. We also present these results with the different prior choices introduced in Section 5.2.2, to illustrate their impact on the EIG and SSD.

Figure 4a plots EIG versus dataset sample size n_d under the priors with different σ , offering a quantitative overview of their trade-offs. For all these curves, we observe a sharp initial increase of the EIG followed by a gradual flattening of the curve—a “kneebend”-like transition—which is consistent with the intuition of diminishing return as the total dataset size grows. The EIG values are generally higher for larger prior σ , because a higher-variance prior presents more of an opportunity to reduce uncertainty as measured by the KL divergence. The different curves also appear to share a similar shape and trend, hinting that the impact of these prior choices may not significantly affect the optimal sample size, which we will show in more detail shortly. Furthermore, the EIG curves appear to be roughly logarithmic. Indeed, Figure 4b plots the logarithmic fitting of EIG curves in the form:

$$\widehat{\text{EIG}}(n_d) = a \log n_d + b, \quad (17)$$

706 which match well with the original EIG curves in Figure 4a. The fitted functions thus can be used to guide future
 707 interpolation and extrapolation analyses.



725 Fig. 4. EIG curves from *pre-hoc* experiment design under different prior σ . The dashed lines in (a) are the computed EIG
 726 values, while the solid lines in (b) are their logarithmic fits.

727 Having obtained the EIG curves, we can now use them to perform SSD. Following the procedure described in
 728 Section 5.2.2, the model designer needs to specify two parameters: 1) n_d^{\max} —absolute maximum number of samples
 729 as dictated by resource constraints, and 2) p —the percentage of the maximum EIG that the model designer wishes
 730 to achieve. In our illustration with MS dataset, we choose n_d^{\max} based on maximum number of 1,500 participants
 731 (approximately 30% participants in MS tertiary clinics associated with our university’s Medical School) that
 732 we could recruit for a 7-day data collection study (the length of the usual seven-day protocols [53–55]) at the
 733 standard compensation of \$200 per participant per week, which totals to a prohibitively costly budget of \$300,000.
 This equates to approximately $n_d^{\max} = 10,000$ samples; as a demonstration, we select $p = 80\%$ as a reasonable
 target percentage of the maximal EIG.

734 Substituting the logarithm fit of EIG from Equation (17) into Equation (14), we can solve for the sample size as
 735 a function of p :

$$736 \quad n_d = \exp \left[\frac{p}{100} \log n_d^{\max} + \frac{(p - 100)b}{100a} \right]. \quad (18)$$

737 Figure 5 shows n_d versus p in semi-log plot under different priors, which appear almost identical except for
 738 a small discrepancy from $\sigma = 0.2$. This suggests the determination of n_d under our procedure is robust and
 739 unaffected by the different prior choices considered here.

740 Having considered the effects of different priors, we now focus only on the case with $\sigma = 0.5$ for the remainder
 741 of the paper. Figures 6 (left) shows both its expected (mean) IG (i.e., the EIG) as well as the standard deviation of
 742 IG plotted as 95% confidence intervals (i.e., ± 1.96 standard deviations). The logarithmic fit of EIG is $\widehat{\text{EIG}}(n_d) =$
 743 $9.35 \ln(n_d) + 31.51$, and in Figure 6 (right) it appears to have excellent agreement with the non-fitted points.
 744 Following our SSD procedure, we compute the required sample size for 80% maximal EIG to be $n_d(80) \approx 807$
 745 samples. This corresponds to a seven-day data collection with approximately 116 participants totaling \$23,200
 746 based on a rate of \$200 per participant per week, providing a saving of \$276,800 compared to the \$300,000 required
 747 for $n_d^{\max} = 10,000$ samples.

748
749
750 5.3.2 *Post-hoc Dataset Analysis.* In the *post-hoc* dataset analysis, our goal is twofold: 1) to validate our *pre-hoc*
 751 experiment design, and 2) to illustrate how to assess if an existing dataset contains sufficient samples. We compute
 752

753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799

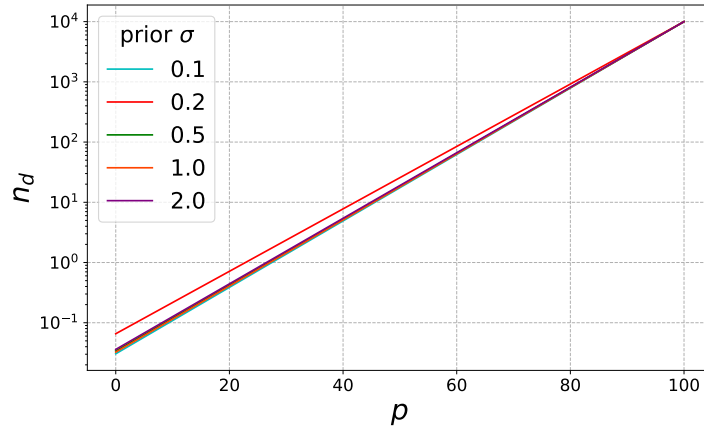


Fig. 5. Dataset sample size n_d required to achieve a target percentage p , under different prior σ .

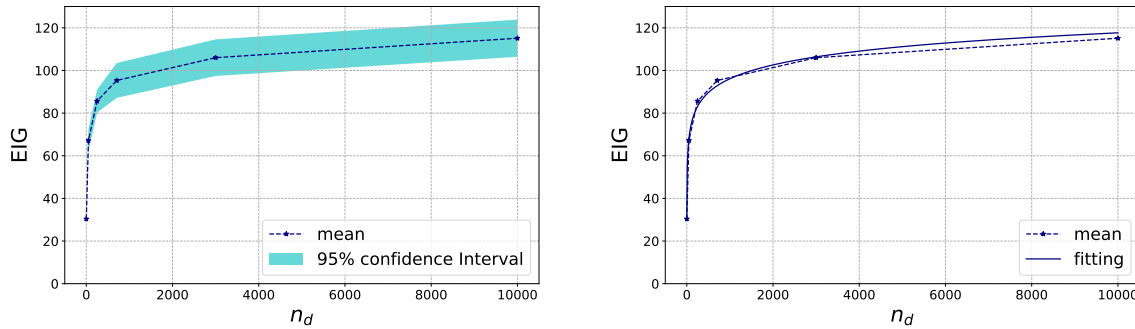


Fig. 6. *Pre-hoc* EIG curve with 95% confidence interval (± 1.96 standard deviations) (left) and logarithmic fit of the EIG curve (right), for the prior with $\sigma = 0.5$.

the realized IG based on the 707 post-collected data instances collected in real life and plot it in Figure 7 (in red/purple). The confidence interval envelope for the purple realized IG is from bootstrapping the 707 samples (e.g., for $n_d = 50$, there are many different ways to choose 50 samples from the total of 707) as well as the randomness from MCMC sampling. While the *pre-hoc* EIG curve (in blue/green) and its envelope captures the uncertainty due to different possible θ values, the *post-hoc* curve represents a specific realization of θ (i.e., data generated from the true θ in real life). Indeed, both curves share the same trend, and the *post-hoc* curve has lower uncertainty compared to its *pre-hoc* counterpart.

A SSD based on the the *pre-hoc* EIG curve is very representative of the realized IG from the collected dataset. Moreover, in this illustration, the *post-hoc* IG curve is higher than *pre-hoc* EIG curve, indicating that the 707 samples we have collected is sufficient for achieving our targeted EIG. The *post-hoc* percentage, which is the ratio between the realized *post-hoc* IG and the maximal *pre-hoc* EIG, is around 84%, exceeding the target of 80%.

Note that our method does not require the model designer to actually investigate these curves if they do not wish to do so. Instead, our method allows for a simple function that takes in our *pre-hoc* data analysis parameters and returns the ratio between the realized *post-hoc* IG and the maximal *pre-hoc* EIG. When the ration exceeds the

target (as is the case in our illustration) the model designer can accept the dataset; otherwise, model designer needs to collect more data samples.

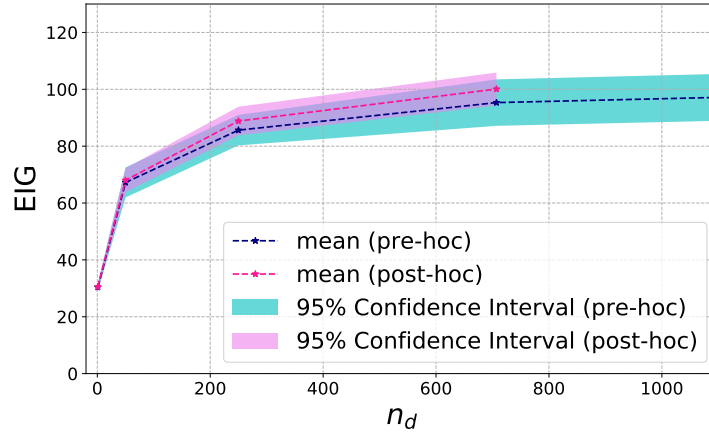


Fig. 7. Pre-hoc EIG curve, post-hoc realized IG curve, and their confidence intervals.

5.3.3 Interpreting Uncertainty Distribution of Model Parameters. Having quantified the overall uncertainty of model parameters for a given dataset size, the model designers may wish to visualize and explore the uncertainty associated with individual model parameters. The *post-hoc* dataset analysis provides the posterior distribution of model parameters θ , which allows model designers to interpret uncertainty information about the model parameters. This posterior distribution is also the formal solution to the Bayesian inference problem.

Figure 8 shows the marginal posterior histogram of each component of θ . While most marginal posteriors appear unimodal, the joint distribution is in fact highly multimodal due to the non-uniqueness of θ (i.e., different θ values can result in similar, or even the exact same stochastic policy), and the appearance of unimodality is due to the effect of marginalization (i.e. projection) onto the lower dimensional spaces.

Figure 9 further presents the top 20 pairwise-marginal scatter plots having the highest linear correlation magnitude. In these figures, red dots indicate the 10 MCMC samples having the highest posterior PDF value. Since these red points are scattered and do not appear to form a cluster, they suggest the joint distribution of θ posterior is highly multimodal. We verified that these high-PDF θ samples indeed result in similar likelihood values and policies, despite having different θ .

Some θ components cannot be interpreted as preference, such as those corresponding to the participants' gender and age, which are state features that do not change within a behavior instance. Other θ components correspond to participant preferences. For example, in this dataset, participants answered most of the self-reported momentary assessments, so *Not Recorded* features are always associated with lower θ values compared to *Recorded* features. The model is also highly confident that participants prefer *CurrentEODPAW : Normal* much more than *CurrentEODPAW : Mild* and *CurrentEODPAW : Moderate*, as the probability mass of *Normal* concentrate around 1 while that of *Mild* and *Moderate* are around 0. These observations and interpretations provide support that our IRL model captures meaningful patterns of behaviors from the dataset.

Having computed the posterior uncertainty of θ , we can then propagate it to other quantities of interest in the model and obtain their posterior-predictive distributions. For example, to compute the updated policy $P(a | s)$ based on the posterior distribution of θ , we can take the posterior samples $\theta^{(k)}$ from MCMC and for each sample compute the policy through Equation (6). The resulting distribution of policies then reflects the uncertainty in

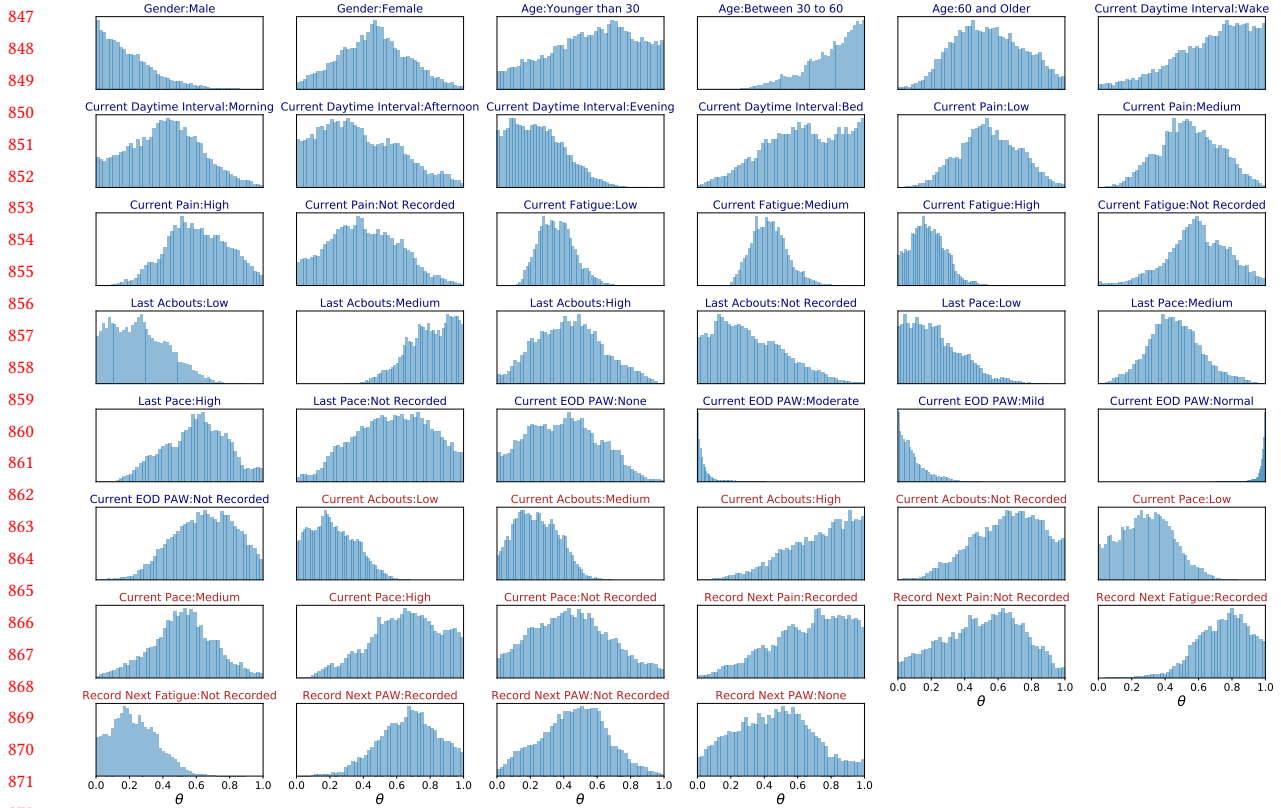


Fig. 8. Marginal posterior distributions (histograms of MCMC samples) for θ components from *post-hoc* dataset analysis. Blue titles indicate state features and orange titles indicate action features.

the policy due to the residual uncertainty in θ . Using this technique, we present in Figure 10 the *pre-hoc* EIG on the policy distribution (i.e., the KL divergence from the prior-predictive to the posterior-predictive distributions of the policy) averaged over all possible state and action pairs, and also the *post-hoc* realized IG. Both curves have a similar trend as the θ curves in Figures 6 and 7, although the uncertainty on the *pre-hoc* EIG curve of policy is higher. Worth noting is that the *post-hoc* realized IG curve is now below the *pre-hoc* EIG curve, in contrast to the curves for θ . This is an example that the value and information gained from data may be different depending on the quantity of interest.

6 DISCUSSION

The results in the previous section serve as a diagnostic tool for SSD. In particular, the *pre-hoc* results allow one to decide how many new samples to collect, and the *post-hoc* results provide an assessment of whether the current dataset size is sufficient. Note that this is different from determining the length of each individual data collection (e.g., determining how many days of data to collect from each participant). In our demonstration, *post-hoc* also serves as a validation of the overall procedure, where the trend of the realized IG is captured within the *pre-hoc* EIG. These tools are particularly valuable when the cost of data is high, where a carefully made decision of how

894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940

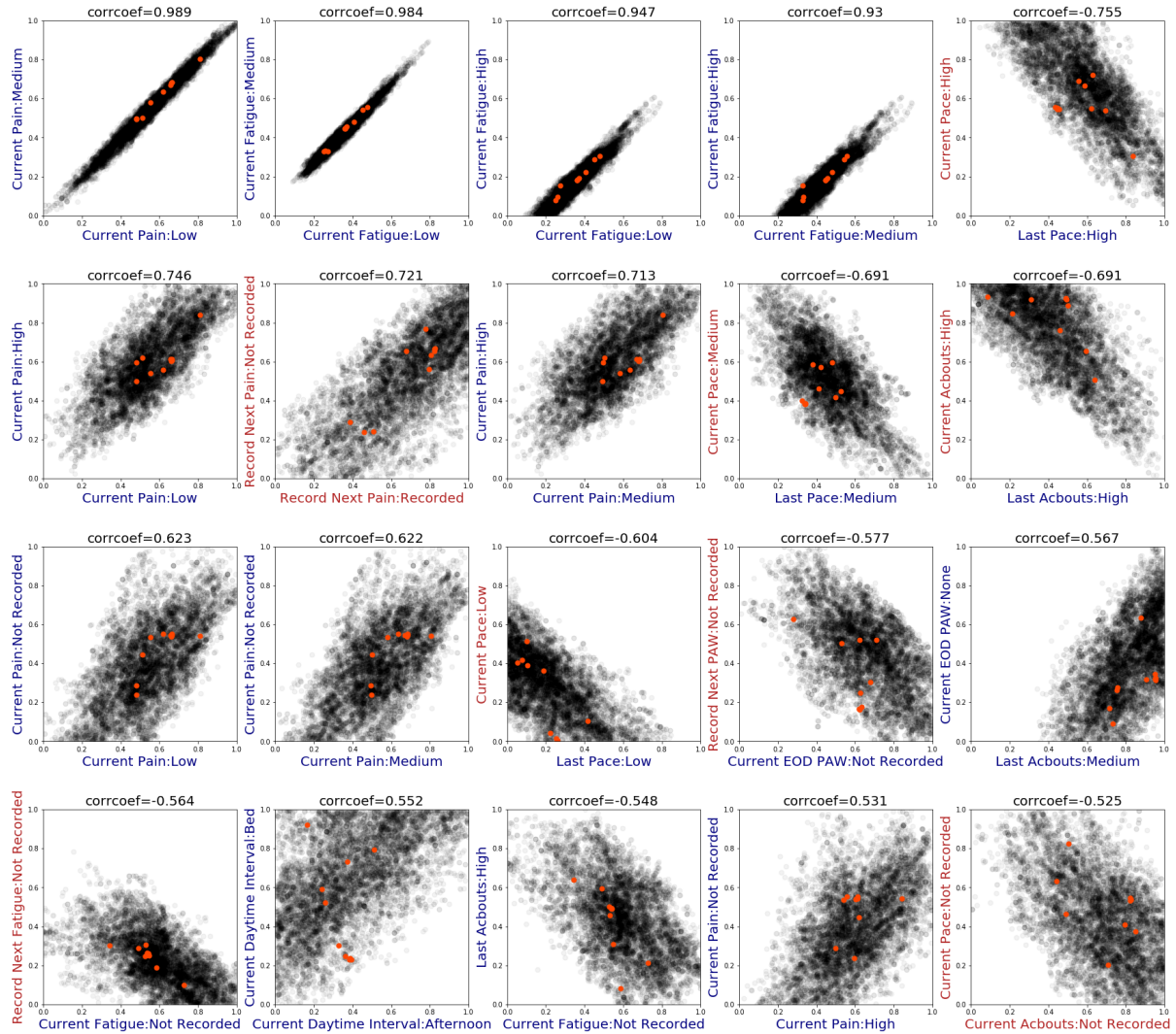


Fig. 9. Pairwise-marginal posterior distributions (MCMC samples) of θ component pairs from *post-hoc* dataset analysis. Every dot is a MCMC sample and its darkness reflects its relative posterior PDF value. Red dots indicate the 10 samples with the highest posterior PDF value. The subplots are sorted from high to low linear correlation magnitude.

much data to collect can lead to large resource savings and the alleviation of potentially unnecessary burden for the participants to undergo the data collection process.

We reiterate that our framework and methodology provide *support* for decision-making, but the decision of sample size selection ultimately rests with the model designer. This decision involves many additional factors such as the specific goals of the model and data usage, monetary cost and scheduling of experiments, value of information and knowledge, consequences and risks, regulatory and policy requirements, and even the degree of risk-aversion of the decision maker. Incorporating these components systematically and comprehensively, as is

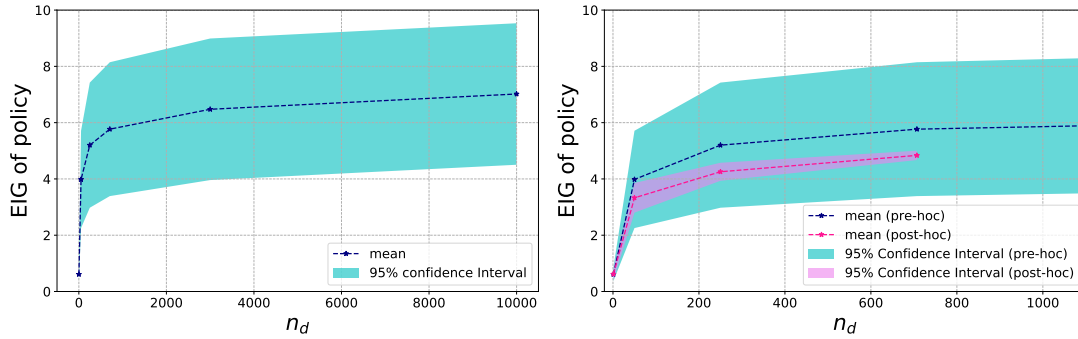


Fig. 10. *Pre-hoc* EIG curve, *post-hoc* realized IG curve, and their confidence intervals of the policy. The right figure is zoomed in from the left (note the horizontal axis values).

pursued in the research of decision theory (e.g., [10, 73]), is extremely challenging and beyond the scope of our work. The methods we provide are therefore a starting point for a potentially complex decision-making process, by offering a quantitative assessment of information content provided by the data. We thus provide a specific example where our framework can return the optimal number of data samples if a decision-maker provides it with maximal number of samples and a percentage of information to target.

Our framework and methodology in Section 4 is completely general, without any requirements on the problem context, dataset, decision rule, and feature choices. The concepts and building blocks from which we established our method are also mathematically principled and rigorously supported by theory; MDP, MaxCausalEnt, SGD, Bayesian inference, KL divergence and EIG, MCMC, and KDE, are all rooted in theoretical and applied computational research, and accompanied by their convergence proofs and conditions. Therefore, our framework is certainly applicable for a wide range of other use-cases that fit under the premise described in Section 4.

However, the computational requirements and difficulty in practice will depend on the size of the problem, cardinality of states and actions, number of features, and the conditioning of the specific problem and dataset. One major motivation for choosing the demonstration of the MS dataset is to showcase our method's ability to handle realistic, non-trivially sized problems.

One key novelty of our approach is the use of Bayesian inference for guiding SSD in an IRL context. Bayesian inference provides model parameter estimates along with uncertainty information, thus quantifying the quality of parameter estimation from a given dataset. This assessment quantity, the EIG, is tightly coupled with information-theoretic metrics, which allows us to explore the tradeoffs between information gain and additional data samples. Our work also advances a Bayesian alternative to the existing MaxCausalEnt IRL algorithm [95] while retaining the maximum entropy framework. Such a Bayesian approach bridges UQ with human behavior modeling, and opens the door for additional avenues of exploration such as Bayesian model selection, Bayesian experimental design, and robust design optimization.

Our proposed method also comes with limitations. For example, it may not immediately be clear how our method decouples epistemic uncertainty (i.e., uncertainty that is reducible given more data) from aleatoric uncertainty (i.e., inherent randomness that is irreducible from data). Although Bayesian inference typically centers around the update of epistemic uncertainty through the computation of the posterior distribution of model parameters, it certainly also involves (if not requires) the participation of aleatoric uncertainty. For example in our work, aleatoric uncertainty manifests *via* the initial probabilities, transition probabilities, and (in part) stochastic policy—all of which are crucial components of our overall model and essential in defining the Bayesian likelihood (if absent, the likelihood would be degenerate and the Bayesian problem would be ill-defined). Our UQ

988 results of posterior-predictive distributions on the policy encompass these aleatoric effects together with the
 989 epistemic uncertainty in the model parameters.

990 However, there are additional sources of aleatoric uncertainty that our method does not capture, notably data
 991 measurement noise and model discrepancy [51]. Although one can view these effects to be grouped into the overall
 992 likelihood probability, we acknowledge that it would be more accurate to model them explicitly. To consider
 993 the epistemic and aleatoric nature of different uncertainty sources, the model designer can introduce them into
 994 our current framework as long as the designer defines and represents their relationships to the other model
 995 variables in a manner faithful to their true behavior. However, from a practical perspective of the decision-maker,
 996 it is of greater importance to understand how the *overall* uncertainty reduces with more data, than specifically
 997 attributing to epistemic or aleatoric types—which is something that our method already enables.

998 Another challenge is computational cost. While a Bayesian framework offers rigorously quantified uncertainty,
 999 each inference solve with MCMC is generally much more expensive than finding point-estimates such as with
 1000 the SGD-based MaxCausalEnt IRL. Further compounding the numerical cost in the *pre-hoc* experiment design
 1001 stage is the *repeated* Bayesian inference solves (the outer loop Monte Carlo) under different samples of θ and D .
 1002 This becomes an extremely compute-intensive task that required the utilization of parallel and high-performance
 1003 computing. The cost would become even higher for more complex IRL models and higher dimensional state,
 1004 action, and feature spaces. These requirements may be alleviated with future advances on computational methods,
 1005 such as through dimension-reduction, efficient sampling (e.g., advanced MCMC), and approximate inference
 1006 (e.g., variational inference [13], Stein variational methods [63]). Nonetheless, even a high computational cost is
 1007 generally well worth the effort when comparing to the cost of real-life experiments needed for data acquisition.
 1008

1009 7 CONCLUSION AND FUTURE WORK

1010 In this paper, we presented a Bayesian SSD method for an existing IRL-based behavior modeling approach [4, 8].
 1011 We illustrated our method on a real problem of modeling behaviors of people with MS [97]. We showcased the
 1012 applicability of our method in different stages of data collection study design: 1) *pre-hoc* experimental design, to
 1013 plan and decide sample size before data collection, and 2) *post-hoc* dataset analysis, to analyse the sample size of
 1014 an existing dataset after data collection.
 1015

1016 Our numerical experiments validated our approach and illustrated how model designers can use the analysis
 1017 from our methods to make an informed decision about how many data samples to collect or assess the uncertainty
 1018 of parameter estimates from their existing datasets. Our method enables capabilities, which together with a
 1019 strategy for efficient data acquisition, can contribute to developing more accurate and reliable human behavior
 1020 models.

1021 Thus, there are three threads for future work. First, conducting SSD experiments under different domains
 1022 could uncover insights on how different domain-specific considerations drive the design of their respective data
 1023 collection studies. Such work could lead to establishing benchmark datasets for modeling human behavior using
 1024 IRL across domains. Second, having a SSD method allows us to study the considerations from model designers
 1025 and domain experts when trading off information of the model for the cost of data collection. Finally, our method
 1026 used a Bayesian formulation of IRL, which enables future exploration of conducting human behavior predictions
 1027 under uncertainty.
 1028

1029 ACKNOWLEDGMENTS

1030 We thank Dr. Anna Kratz for her invaluable input about Multiple Sclerosis that informed the design of our
 1031 models and for providing us with access to the datasets used in this work. We also thank our reviewers for their
 1032 constructive feedback that helped us improve our paper.
 1033
 1034

This research is based upon work supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award Number DE-SC0021398. This paper was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. This research was supported in part through computational resources and services provided by Advanced Research Computing (ARC), a division of Information and Technology Services (ITS) at the University of Michigan, Ann Arbor.

REFERENCES

- [1] Rebecca Adami and Edison Thomaz. 2019. Leveraging Active Learning and Conditional Mutual Information to Minimize Data Annotation in Human Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 70 (Sept. 2019), 23 pages. <https://doi.org/10.1145/3351228>
- [2] C. J. Adcock. 1997. Sample size determination: a review. *Journal of the Royal Statistical Society: Series D (The Statistician)* 46, 2 (1997), 261–283. <https://doi.org/10.1111/1467-9884.00082>
- [3] Roberto Alejo, Vicente Garcia, and J. Pacheco. 2015. An Efficient Over-sampling Approach Based on Mean Square Error Back-propagation for Dealing with the Multi-class Imbalance Problem. *Neural Processing Letters* 42 (12 2015), 603–617. <https://doi.org/10.1007/s11063-014-9376-3>
- [4] Nikola Banovic, Tofi Buzali, Fanny Chevalier, Jennifer Mankoff, and Anind K. Dey. 2016. Modeling and Understanding Human Routine Behavior. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 248–260. <https://doi.org/10.1145/2858036.2858557>
- [5] Nikola Banovic, Tovi Grossman, and George Fitzmaurice. 2013. *The Effect of Time-Based Cost of Error in Target-Directed Pointing Tasks*. Association for Computing Machinery, New York, NY, USA, 1373–1382. <https://doi.org/10.1145/2470654.2466181>
- [6] Nikola Banovic, Jennifer Mankoff, and Anind K. Dey. 2018. Computational Model of Human Routine Behaviors. In *Computational Interaction*, Antti Oulasvirta, Per Ola Kristensson, Xiaojun Bi, and Andrew Howes (Eds.). Oxford University Press, Oxford, Chapter 14, 377–398.
- [7] Nikola Banovic, Antti Oulasvirta, and Per Ola Kristensson. 2019. Computational Modeling in Human-Computer Interaction. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (*CHI EA '19*). Association for Computing Machinery, New York, NY, USA, 1–7. <https://doi.org/10.1145/3290607.3299032>
- [8] Nikola Banovic, Anqi Wang, Yanfeng Jin, Christie Chang, Julian Ramos, Anind Dey, and Jennifer Mankoff. 2017. Leveraging Human Routine Models to Detect and Generate Human Behaviors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI '17*). ACM, New York, NY, USA, 6683–6694. <https://doi.org/10.1145/3025453.3025571>
- [9] Edmon Begoli, Tanmoy Bhattacharya, and Dimitri Kusnezov. 2019. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence* 1, 1 (jan 2019), 20–23. <https://doi.org/10.1038/s42256-018-0004-1>
- [10] James O. Berger. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer New York, New York, NY. <https://doi.org/10.1007/978-1-4757-4286-2>
- [11] Jose M. Bernardo and Adrian F. M. Smith. 2000. *Bayesian Theory*. John Wiley & Sons, New York, NY.
- [12] Christopher Bishop. 2006. *Pattern Recognition and Machine Learning*. Springer-Verlag New York.
- [13] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. 2017. Variational Inference: A Review for Statisticians. *J. Amer. Statist. Assoc.* 112, 518 (2017), 859–877. <https://doi.org/10.1080/01621459.2017.1285773>
- [14] Natthaphan Boonyanunta and Panlop Zeepongsekul. 2004. Predicting the Relationship Between the Size of Training Sample and the Predictive Power of Classifiers. In *Knowledge-Based Intelligent Information and Engineering Systems*, Mircea Gh. Negoita, Robert J. Howlett, and Lakhmi C. Jain (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 529–535.
- [15] Leo Breiman. 2001. Statistical modeling: The two cultures. *Statist. Sci.* (2001).

- 1082 [16] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng (Eds.). 2011. *Handbook of Markov Chain Monte Carlo*. Chapman and
1083 Hall/CRC. <https://doi.org/10.1201/b10905>
- 1084 [17] Daniel S Brown and Scott Niekum. 2017. Efficient probabilistic performance bounds for inverse reinforcement learning. *arXiv preprint*
1085 *arXiv:1707.00724* (2017).
- 1086 [18] Daniel S Brown and Scott Niekum. 2018. Efficient probabilistic performance bounds for inverse reinforcement learning. In *Thirty-Second*
1087 *AAAI Conference on Artificial Intelligence*.
- 1088 [19] Kathryn Chaloner and Isabella Verdinelli. 1995. Bayesian Experimental Design: A Review. *Statist. Sci.* 10, 3 (1995), 273–304. <https://doi.org/10.1214/ss/1177009939>
- 1089 [20] Youngjae Chang, Akhil Mathur, Anton Isopoussu, Junehwa Song, and Fahim Kawsar. 2020. A Systematic Study of Unsupervised Domain
1090 Adaptation for Robust Human-Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 39 (March 2020),
30 pages. <https://doi.org/10.1145/3380985>
- 1091 [21] Xiuli Chen, Sandra Dorothee Starke, Chris Baber, and Andrew Howes. 2017. A Cognitive Model of How People Make Decisions Through
1092 Interaction with Visual Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado,
1093 USA) (*CHI '17*). Association for Computing Machinery, New York, NY, USA, 1205–1216. <https://doi.org/10.1145/3025453.3025596>
- 1094 [22] Jacob Cohen. 1977. CHAPTER 1 - The Concepts of Power Analysis. In *Statistical Power Analysis for the Behavioral Sciences*, Jacob Cohen
1095 (Ed.). Academic Press, 1 – 17. <https://doi.org/10.1016/B978-0-12-179060-8.50006-2>
- 1096 [23] Thomas A. Cover and Joy A. Thomas. 2006. *Elements of Information Theory* (2nd ed.). John Wiley & Sons, Hoboken, NJ.
- 1097 [24] Luis G. Crespo, Sean P. Kenny, and Daniel P. Giesy. 2014. The NASA Langley Multidisciplinary Uncertainty Quantification Challenge.
1098 In *16th AIAA Non-Deterministic Approaches Conference*. American Institute of Aeronautics and Astronautics, Reston, Virginia. <https://doi.org/10.2514/6.2014-1347>
- 1099 [25] Yuchen Cui and Scott Niekum. 2017. Active learning from critiques via bayesian inverse reinforcement learning. In *Robotics: Science and*
1100 *Systems Workshop on Mathematical Models, Algorithms, and Human-Robot Interaction*.
- 1101 [26] Nathan Eagle and Alex Sandy Pentland. 2009. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology* 63,
1102 7 (01 May 2009), 1057–1066. <https://doi.org/10.1007/s00265-009-0739-0>
- 1103 [27] Katayoun Farrahi and Daniel Gatica-Perez. 2012. Extracting Mobile Behavioral Patterns with the Distant N-Gram Topic Model. In *2012*
1104 *16th International Symposium on Wearable Computers*. 1–8.
- 1105 [28] Rebecca Fiebrink, Perry R. Cook, and Dan Trueman. 2011. Human Model Evaluation in Interactive Supervised Learning. In *Proceedings*
1106 *of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). Association for Computing
1107 Machinery, New York, NY, USA, 147–156. <https://doi.org/10.1145/1978942.1978965>
- 1108 [29] Rosa L. Figueroa, Qing Zeng-Treitler, Sasikiran Kandula, and Long H. Ngo. 2012. Predicting sample size required for classification
1109 performance. *BMC Medical Informatics and Decision Making* 12 (Feb 2012), 8.
- 1110 [30] Chelsea Finn, Sergey Levine, and Pieter Abbeel. 2016. Guided Cost Learning: Deep Inverse Optimal Control via Policy Optimization.
1111 In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48* (New York, NY, USA)
1112 (*ICML'16*). JMLR.org, 49–58.
- 1113 [31] Daniel Foreman-Mackey, David W Hogg, Dustin Lang, and Jonathan Goodman. 2013. emcee: the MCMC hammer. *Publications of the*
1114 *Astronomical Society of the Pacific* 125, 925 (2013), 306.
- 1115 [32] K. Fukunaga and R. R. Hayes. 1989. Effects of sample size in classifier design. *IEEE Transactions on Pattern Analysis and Machine*
1116 *Intelligence* 11, 8 (1989), 873–885.
- 1117 [33] Christoph Gebhardt, Brian Hecox, Bas van Opheusden, Daniel Wigdor, James Hillis, Otmar Hilliges, and Hrvoje Benko. 2019. Learning
1118 Cooperative Personalized Policies from Gaze Data. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software*
1119 *and Technology* (New Orleans, LA, USA) (*UIST '19*). Association for Computing Machinery, New York, NY, USA, 197–208. <https://doi.org/10.1145/3332165.3347933>
- 1120 [34] Christoph Gebhardt, Antti Oulasvirta, and Otmar Hilliges. 2020. Hierarchical Reinforcement Learning as a Model of Human Task
1121 Interleaving. *arXiv:cs.AI/2001.02122*
- 1122 [35] Samuel J. Gershman, Eric J. Horvitz, and Joshua B. Tenenbaum. 2015. Computational rationality: A converging paradigm
1123 for intelligence in brains, minds, and machines. *Science* 349, 6245 (2015), 273–278. <https://doi.org/10.1126/science.aac6076>
1124 *arXiv:https://science.sciencemag.org/content/349/6245/273.full.pdf*
- 1125 [36] Roger Ghanem, David Higdon, and Houman Owahdi (Eds.). 2017. *Handbook of uncertainty quantification*. Springer International
1126 Publishing, Cham. <https://doi.org/10.1007/978-3-319-12385-1> *arXiv:1507.00398*
- 1127 [37] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York, NY.
- 1128 [38] Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konuyshkova, kumaripaba Athukorala, Samuel Kaski, and Giulio Jacucci. 2013. Directing
Exploratory Search: Reinforcement Learning from User Interactions with Keywords. In *Proceedings of the 2013 International Conference*
on *Intelligent User Interfaces* (Santa Monica, California, USA) (*IUI '13*). Association for Computing Machinery, New York, NY, USA,
117–128. <https://doi.org/10.1145/2449396.2449413>

- 1129 [39] Jonathan Goodman and Jonathan Weare. 2010. Ensemble samplers with affine invariance. *Communications in applied mathematics and*
1130 *computational science* 5, 1 (2010), 65–80.
- 1131 [40] Stephen L Hauser and Jorge R Oksenberg. 2006. The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration.
1132 *Neuron* 52, 1 (2006), 61–76.
- 1133 [41] Xun Huan and Youssef M. Marzouk. 2013. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. Comput.*
1134 *Phys.* 232, 1 (2013), 288–317. <https://doi.org/10.1016/j.jcp.2012.08.013>
- 1135 [42] Mahdi Imani and Ulisses M Braga-Neto. 2018. Control of gene regulatory networks using Bayesian inverse reinforcement learning.
1136 *IEEE/ACM transactions on computational biology and bioinformatics* 16, 4 (2018), 1250–1261.
- 1137 [43] Sozo Inoue, Paula Lago, Tahera Hossain, Tittaya Mairittha, and Nattaya Mairittha. 2019. Integrating Activity Recognition and Nursing
1138 Care Records: The System, Deployment, and a Verification Study. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article
1139 Article 86 (Sept. 2019), 24 pages. <https://doi.org/10.1145/3351244>
- 1140 [44] Sozo Inoue and Xincheng Pan. 2016. Supervised and Unsupervised Transfer Learning for Activity Recognition from Simple In-Home
1141 Sensors. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*
1142 (Hiroshima, Japan) (*MOBIQUITOUS 2016*). Association for Computing Machinery, New York, NY, USA, 20–27. <https://doi.org/10.1145/2994374.2994400>
- 1143 [45] Nathalie Japkowicz and Shaju Stephen. 2002. The Class Imbalance Problem: A Systematic Study. *Intell. Data Anal.* 6, 5 (Oct. 2002),
1144 429–449.
- 1145 [46] Edwin T. Jaynes and G. L. Bretthorst. 2003. *Probability Theory: the Logic of Science*. Cambridge University Press.
- 1146 [47] James M. Joyce. 2011. *Kullback-Leibler Divergence*. Springer Berlin Heidelberg, Berlin, Heidelberg, 720–722. [https://doi.org/10.1007/978-](https://doi.org/10.1007/978-3-642-04898-2_327)
1147 [3-642-04898-2_327](https://doi.org/10.1007/978-3-642-04898-2_327)
- 1148 [48] Leslie Pack Kaelbling, Michael L. Littman, and Andrew W. Moore. 1996. Reinforcement Learning: A Survey. *J. Artif. Int. Res.* 4, 1 (May
1149 1996), 237–285.
- 1150 [49] Antti Kangasrääsio and Samuel Kaski. 2018. Inverse reinforcement learning from summary data. *Machine Learning* 107, 8 (01 Sep 2018),
1151 1517–1535. <https://doi.org/10.1007/s10994-018-5730-4>
- 1152 [50] Antti Kangasrääsio, Jussi P. P. Jokinen, Antti Oulasvirta, Andrew Howes, and Samuel Kaski. 2019. Parameter Inference for Computational
1153 Cognitive Models with Approximate Bayesian Computation. *Cognitive Science* 43, 6 (2019), e12738. <https://doi.org/10.1111/cogs.12738>
1154 arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.12738>
- 1155 [51] Marc. C. Kennedy and Anthony O’Hagan. 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series*
1156 *B (Statistical Methodology)* 63, 3 (2001), 425–464. <https://doi.org/10.1111/1467-9868.00294>
- 1157 [52] Iuliia Kotseruba and John K. Tsotsos. 2020. 40 years of cognitive architectures: core cognitive abilities and practical applications.
1158 *Artificial Intelligence Review* 53, 1 (01 Jan 2020), 17–94. <https://doi.org/10.1007/s10462-018-9646-y>
- 1159 [53] Anna L Kratz, Tiffany J Braley, Emily Foxen-Craft, Eric Scott, John F Murphy III, and Susan L Murphy. 2017. How do pain, fatigue,
1160 depressive, and cognitive symptoms relate to well-being and social and physical functioning in the daily lives of individuals with
1161 multiple sclerosis? *Archives of physical medicine and rehabilitation* 98, 11 (2017), 2160–2166.
- 1162 [54] Anna L Kratz, Susan L Murphy, and Tiffany J Braley. 2017. Ecological momentary assessment of pain, fatigue, depressive, and cognitive
1163 symptoms reveals significant daily variability in multiple sclerosis. *Archives of physical medicine and rehabilitation* 98, 11 (2017),
1164 2142–2150.
- 1165 [55] Anna L Kratz, Susan L Murphy, and Tiffany J Braley. 2017. Pain, fatigue, and cognitive symptoms are temporally associated within but
1166 not across days in multiple sclerosis. *Archives of physical medicine and rehabilitation* 98, 11 (2017), 2151–2159.
- 1167 [56] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.
- 1168 [57] Katri Leino, Antti Oulasvirta, and Mikko Kurimo. 2019. RL-KLM: Automating Keystroke-Level Modeling with Reinforcement Learning.
1169 In *Proceedings of the 24th International Conference on Intelligent User Interfaces (Marina del Ray, California) (IUI ’19)*. Association for
1170 Computing Machinery, New York, NY, USA, 476–480. <https://doi.org/10.1145/3301275.3302285>
- 1171 [58] Katri Leino, Kashyap Todi, Antti Oulasvirta, and Mikko Kurimo. 2019. Computer-Supported Form Design Using Keystroke-Level Modeling
1172 with Reinforcement Learning. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion (Marina del*
1173 *Ray, California) (IUI ’19)*. Association for Computing Machinery, New York, NY, USA, 85–86. <https://doi.org/10.1145/3308557.3308704>
- 1174 [59] Russell V Lenth. 2001. Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician* 55, 3 (2001),
1175 187–193. <https://doi.org/10.1198/000313001317098149> arXiv:<https://doi.org/10.1198/000313001317098149>
- [60] Richard L. Lewis, Andrew Howes, and Satinder Singh. 2014. Computational Rationality: Linking Mechanism and Behavior
Through Bounded Utility Maximization. *Topics in Cognitive Science* 6, 2 (2014), 279–311. <https://doi.org/10.1111/tops.12086>
arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12086>
- [61] Nan Li, Subbarao Kambhampati, and Sungwook Yoon. 2009. Learning Probabilistic Hierarchical Task Networks to Capture User
Preferences. In *International Joint Conference on Artificial Intelligence*. <https://www.aaai.org/ocs/index.php/IJCAI/IJCAI-09/paper/view/417/874>

- 1176 [62] Dennis V. Lindley. 1997. The Choice of Sample Size. *Journal of the Royal Statistical Society. Series D (The Statistician)* 46, 2 (1997), 129–138.
1177 <http://www.jstor.org/stable/2988516>
- 1178 [63] Qiang Liu and Dilin Wang. 2016. Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm. In *Advances in*
1179 *Neural Information Processing Systems 29 (NIPS 2016)*. Barcelona, Spain, 2378–2386.
- 1180 [64] Magnus S. Magnusson. 2000. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods,*
1181 *Instruments, & Computers* 32, 1 (01 Mar 2000), 93–110. <https://doi.org/10.3758/BF03200792>
- 1182 [65] Gideon S. Mann and Andrew McCallum. 2010. Generalized Expectation Criteria for Semi-Supervised Learning with Weakly Labeled
1183 Data. *Journal of Machine Learning Research* 11, 32 (2010), 955–984. <http://jmlr.org/papers/v11/mann10a.html>
- 1184 [66] Scott E. Maxwell, Ken Kelley, and Joseph R. Rausch. 2008. Sample Size Planning for Statistical Power and Accuracy in Param-
1185 eter Estimation. *Annual Review of Psychology* 59, 1 (2008), 537–563. <https://doi.org/10.1146/annurev.psych.59.103006.093735> PMID: 17937603.
1186 arXiv:<https://doi.org/10.1146/annurev.psych.59.103006.093735>
- 1187 [67] Roderick Melnik. 2015. *Universality of Mathematical Models in Understanding Nature, Society, and Man-*
1188 *Made World*. John Wiley & Sons, Ltd, Chapter 1, 1–16. <https://doi.org/10.1002/9781118853887.ch1>
1189 arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118853887.ch1>
- 1190 [68] Peter Müller. 2005. Simulation Based Optimal Design. *Handbook of Statistics* 25 (2005), 509–518. <https://doi.org/10.1016/S0169->
1191 [7161\(05\)25017-4](https://doi.org/10.1016/S0169-7161(05)25017-4)
- 1192 [69] Andrew Y. Ng and Stuart J. Russell. 2000. Algorithms for Inverse Reinforcement Learning. In *Proceedings of the Seventeenth International*
1193 *Conference on Machine Learning (ICML '00)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 663–670.
- 1194 [70] William L. Oberkamp, Timothy G. Trucano, and Charles Hirsch. 2004. Verification, validation, and predictive capability in computational
1195 engineering and physics. *Applied Mechanics Reviews* 57, 5 (sep 2004), 345–384. <https://doi.org/10.1115/1.1767847>
- 1196 [71] Anthony O'Hagan, Caitlin E. Buck, Alireza Daneshkhan, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley,
1197 and Tim Rakow. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons, Ltd, Chichester, United Kingdom.
1198 <https://doi.org/10.1002/0470033312>
- 1199 [72] Antti Oulasvirta, Jussi P. P. Jokinen, and Andrew Howes. 2022. Computational rationality as a theory of interaction. In *CHI '22*.
1200 Association for Computing Machinery (ACM), United States. Not yet published as of 29/03/2022.; CHI '22 : CHI Conference on Human
1201 Factors in Computing Systems ; Conference date: 29-04-2022 Through 05-05-2022.
- 1202 [73] Giovanni Parmigiani and Lurdes Y. T. Inoue. 2009. *Decision Theory: Principles and Approaches*. John Wiley & Sons, Inc., West Sussex,
1203 United Kingdom. 372 pages. <http://books.google.com/books?id=mnjGCYqWj7EC&pgis=1>
- 1204 [74] Martin Pilch, Timothy G. Trucano, and Jon C. Helton. 2006. *Ideas Underlying Quantification of Margins and Uncertainties (QMU): A*
1205 *White Paper*. Technical Report. Sandia National Laboratories.
- 1206 [75] Martin L. Puterman. 2014. Markov decision processes: discrete stochastic dynamic programming. (2014), 684 Pages.
- 1207 [76] Deepak Ramachandran and Eyal Amir. 2007. Bayesian Inverse Reinforcement Learning. In *IJCAI*, Vol. 7. 2586–2591.
- 1208 [77] Christian P. Robert and George Casella. 2004. *Monte Carlo Statistical Methods*. Springer New York, New York, NY. [https://doi.org/10.](https://doi.org/10.1007/978-1-4757-4145-2)
1209 [1007/978-1-4757-4145-2](https://doi.org/10.1007/978-1-4757-4145-2)
- 1210 [78] Lina M Rojas-Barahona and Christophe Cerisara. 2014. Bayesian inverse reinforcement learning for modeling conversational agents in
1211 a virtual environment. In *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 503–514.
- 1212 [79] Stephane Ross, Geoffrey J. Gordon, and J. Andrew Bagnell. 2011. No-Regret Reductions for Imitation Learning and Structured Prediction.
1213 In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- 1214 [80] Adam Sadilek and John Krumm. 2012. Far Out: Predicting Long-Term Human Mobility. In *AAAI Conference on Artificial Intelligence*.
1215 <https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/4845/5275>
- 1216 [81] John M Salsman, David Victorson, Seung W Choi, Amy H Peterman, Allen W Heinemann, Cindy Nowinski, and David Cella. 2013.
1217 Development and validation of the positive affect and well-being scale for the neurology quality of life (Neuro-QOL) measurement
1218 system. *Quality of Life Research* 22, 9 (2013), 2569–2580.
- 1219 [82] Sayan Sarcar, Jussi P.P. Jokinen, Antti Oulasvirta, Zhenxin Wang, Chaklam Silpasuwanchai, and Xiangshi Ren. 2018. Ability-Based
1220 Optimization of Touchscreen Interactions. *IEEE Pervasive Computing* 17, 1 (2018), 15–26. <https://doi.org/10.1109/MPRV.2018.011591058>
- 1221 [83] David W Scott. 2015. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.
- 1222 [84] Burr Settles. 2009. *Active learning literature survey*. Technical Report. University of Wisconsin-Madison Department of Computer
1223 Sciences.
- 1224 [85] D. S. Sivia and J. Skilling. 2006. *Data Analysis: A Bayesian Tutorial* (2nd ed.). Oxford University Press, New York, NY. 246 pages.
- 1225 [86] R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts'
1226 Law Research in HCI. *Int. J. Hum.-Comput. Stud.* 61, 6 (Dec. 2004), 751–789. <https://doi.org/10.1016/j.ijhcs.2004.09.001>
- 1227 [87] Arun Venkatraman, Martial Hebert, and J. Bagnell. 2015. Improving Multi-Step Prediction of Learned Time Series Models. In *AAAI*
1228 *Conference on Artificial Intelligence*. <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9592/9976>
- 1229 [88] Udo Von Toussaint. 2011. Bayesian inference in physics. *Reviews of Modern Physics* 83 (2011), 943–999. [https://doi.org/10.1103/](https://doi.org/10.1103/RevModPhys.83.943)
1230 [RevModPhys.83.943](https://doi.org/10.1103/RevModPhys.83.943)

- 1223 [89] Robert C Wilson and Anne GE Collins. 2019. Ten simple rules for the computational modeling of behavioral data. *eLife* 8 (nov 2019),
1224 e49547. <https://doi.org/10.7554/eLife.49547>
- 1225 [90] Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniella K. Villalba, Janine M. Dutcher, Michael J. Tumminia, Tim Althoff, Sheldon Cohen,
1226 Kasey G. Creswell, J. David Creswell, Jennifer Mankoff, and Anind K. Dey. 2019. Leveraging Routine Behavior and Contextually-Filtered
1227 Features for Depression Detection among College Students. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 116 (Sept.
1228 2019), 33 pages. <https://doi.org/10.1145/3351274>
- 1229 [91] Shuochao Yao, Yiran Zhao, Huajie Shao, Aston Zhang, Chao Zhang, Shen Li, and Tarek Abdelzaher. 2018. RDeepSense: Reliable Deep
1230 Mobile Computing Models with Uncertainty Estimations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 173 (Jan.
1231 2018), 26 pages. <https://doi.org/10.1145/3161181>
- 1232 [92] Shuochao Yao, Yiran Zhao, Huajie Shao, Chao Zhang, Aston Zhang, Shaohan Hu, Dongxin Liu, Shengzhong Liu, Lu Su, and Tarek
1233 Abdelzaher. 2018. SenseGAN: Enabling Deep Learning for Internet of Things with a Semi-Supervised Framework. *Proc. ACM Interact.*
1234 *Mob. Wearable Ubiquitous Technol.* 2, 3, Article 144 (Sept. 2018), 21 pages. <https://doi.org/10.1145/3264954>
- 1235 [93] Yunxiu Zeng, Kai Xu, Quanjun Yin, L. Qin, Yabing Zha, and William Yeoh. 2018. Inverse Reinforcement Learning Based Human Behavior
1236 Modeling for Goal Recognition in Dynamic Local Network Interdiction. In *AAAI Workshops*.
- 1237 [94] Brian Ziebart, Anind Dey, and J. Andrew Bagnell. 2012. Probabilistic Pointing Target Prediction via Inverse Optimal Control. In
1238 *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces* (Lisbon, Portugal) (*IUI '12*). Association for Computing
1239 Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/2166966.2166968>
- 1240 [95] Brian D. Ziebart, J. Andrew Bagnell, and Anind K. Dey. 2010. Modeling Interaction via the Principle of Maximum Causal Entropy. In
1241 *Proceedings of the 27th International Conference on International Conference on Machine Learning* (Haifa, Israel) (*ICML'10*). Omnipress,
1242 USA, 1255–1262. <http://dl.acm.org/citation.cfm?id=3104322.3104481>
- 1243 [96] Brian D. Ziebart, Andrew L. Maas, Anind K. Dey, and J. Andrew Bagnell. 2008. *Navigate like a Cabbie: Probabilistic Reasoning from Observed*
1244 *Context-Aware Behavior*. Association for Computing Machinery, New York, NY, USA, 322–331. <https://doi.org/10.1145/1409635.1409678>
- 1245 [97] Tjalf Ziemssen, Raimar Kern, and Katja Thomas. 2016. Multiple sclerosis: clinical profiling and data collection as prerequisite for
1246 personalized medicine approach. *BMC neurology* 16, 1 (2016), 124.