An Information and Control Framework for Optimizing User-Compliant Human–Computer Interfaces

This paper presents a framework for a human–computer interface, which provides a simplified method based on optimal transport theory to generate optimal feedback signals between the computer and human in high dimension.

By Justin Tantiongloc, Diego A. Mesa, *Student Member IEEE*, Rui Ma, *Member IEEE*, Sanggyun Kim, *Member IEEE*, Cristian H. Alzate, *Student Member IEEE*, Jaime J. Camacho, *Student Member IEEE*, Vidya Manian, *Member IEEE*, and Todd P. Coleman, *Senior Member IEEE*

ABSTRACT | We consider a general framework for a human-computer interface whereby the human's knowledge is represented as a point in Euclidean space, the intention of the human is signaled to the computer over a noisy channel, and the computer queries the human in a manner that is amenable to human operation. With these constraints at hand, we demonstrate a class of systems that are nonetheless information-theoretically optimal in that the computer very rapidly hones in on the intent of the human. Much recent work on feedback information theory has been dedicated to the exploration of methods by which optimal feedback may be derived for the purpose of expediting the communication of a message point between an inanimate encoder and decoder. Our framework not only takes advantage of previous work to demonstrate its communication optimality from this perspective as well as from an information-theoretic perspective but also contributes two distinct advantages. First, our framework provides a simplified method based on optimal transport theory to generate optimal feedback signals between the computer and human in high dimension, while still preserving communication optimality. Second, our framework specifically lends itself to the integration of a human user by attempting to moderate the difficulty of the task presented to the user, while still preserving optimality. We demonstrate applications of our framework within the context of multi-agent brain-computer interfaces.

KEYWORDS | Bayes procedures; communication channels; computer interfaces human factors; control systems; information theory; optimization methods; user interfaces; user interfaces human factors

I. INTRODUCTION

The idea of transferring knowledge from a human to a computer system has been a widely studied area of research across a variety of fields, including machine learning, education, and economics. Machine learning,

Manuscript received September 15, 2015; revised June 24, 2016; accepted September 6, 2016. Date of publication October 28, 2016; date of current version January 18, 2017. This work was supported by the National Science Foundation (NSF) Center for Science of Information under Grant CCF-0939370, ARO contract and ARO-62793-RT-REP, and the Multidisciplinary University Research Initiatives (MURI) Award under contract ARO-W911NF-15-1-0479.

J. Tantiongloc is with the Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: jctanti@gmail.com).

D. A. Mesa, **R. Ma**, **S. Kim**, and **T. P. Coleman** are with the Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: diego898@gmail.com; mrshuiuc@gmail.com; ifree77@gmail.com; tpcoleman@eng.ucsd.edu).

C. H. Alzate and **V. Manian** are with the Department of Electrical Engineering, University of Puerto Rico, Mayaguez 00682, Puerto Rico (e-mail: cristian.alzate@upr.edu; manian@ece.uprm.edu).

J. J. Camacho is with the Department of Electrical Engineering, University of Puerto Rico, Vieques 00765, Puerto Rico (e-mail: jaime.camacho@upr.edu).

Digital Object Identifier: 10.1109/JPROC.2016.2608723

in general, widely involves the practice of designing computer systems capable of learning abstract information from human users in the presence of arbitrary noise sources.

In the field of interactive reinforcement learning, systems have been designed in which a computer agent attempts to learn optimal control information from a human expert. In these scenarios, the human often provides input to the computer with the intention of steering it toward the optimal control policy [1]–[4].

In many cases, human–computer interfaces can be thought of as an instantiation of traditional coding-decoding frameworks with feedback [5]–[7]. In general, the human encodes the information to be transmitted over a noisy channel to the computer, who decodes the information and takes an appropriate control action with respect to the environment, which can then be used as feedback to the human. The human then uses the feedback to modulate its next input into the noisy channel to the end, reliably communicating the information to the computer.

Much previous work has gone into developing the necessary theory to also maximize the efficiency of such systems. One such notion of optimization that has garnered much attention in the information and statistics community is maximization of mutual information or minimization of Shannon entropy [8], [9].

A. Main Contributions

We describe the primary contributions of this work as follows.

We emphasize that the modulation task for the human agent should remain intellectually tractable for the sake of making the system "easy-to-use" from the perspective of the human agent. Although recent work has demonstrated information-theoretically optimal feedback communication systems, they do not all directly apply to human–computer interfaces. Specifically, if the human's task is too mentally taxing to exhibit a desired behavioral policy, we run the risk of welcoming user error into the system and thus further sources of noise. By keeping the task required of the human agent as simple and streamlined as possible, we further increase the efficacy at which the human agent can signal information to the computer agent in a desired manner, thus minimizing the possibility of erroneous behavior due to cognitive load.

Wedding to the aforementioned constraints, we construct and demonstrate systems that balance this "user-compatibility" constraint with optimality from an information-theoretic perspective. We also present a novel and generalized means by which maximally informative feedback signals to the human can be derived, even in settings in which the human knowledge is in higher dimensions, through a connection to the theory of optimal transport [10] and the orientation-preserving properties of monotonic diffeomorphisms.

Keeping these central themes in mind, we also demonstrate a class of "collaborative brain–computer interfaces," where multiple agents with common knowledge interactively signal through brain signals to a computer. We demonstrate that these crowd-sourced systems are nonetheless information-theoretically optimal.

B. Outline

The remainder of this work will be organized as follows. Section II describes the framework from the traditional information and control perspectives. Section III presents a slightly modified model that emphasizes the design of a practical and intellectually tractable system specifically for use with a human user. Section IV discusses a novel method by which optimal feedback for such communication systems can be derived, even in high dimension. Finally, Section V provides a proof-of-concept application instantiating this framework in the form of an optimal multi-user brain–computer interface.

II. A DECENTRALIZED INFORMATION AND CONTROL VIEWPOINT

In its most general sense, a human–computer interface is a system that facilitates signaling between human and computer agents in some meaningful way. The computer can relay sensory information to the user about what it interpreted so that subsequent signaling can be corrective in nature. This causal feedback loop allows the human and computer agents to cooperate via an iterative process, as shown in Fig. 1.

In our case, we focus on applications whereby the human agent is modeled as possessing some form of abstract information or knowledge, denoted as $W \in \mathsf{W} \subset \mathbb{R}^d$, that it would like to reliably convey to the computer agent over a noisy channel with feedback from the computer agent; we will sometimes refer to this abstract piece of information as the "message point."

Specifically, the human agent iteratively generates input signals to the noisy channel, X_t, as a function of the underlying knowledge of interest, W, and feedback of all

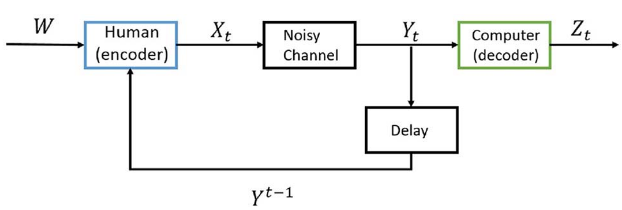


Fig. 1. ***General model for decentralized information communication approach to designing a human–computer interface.***

the information the computer has causally seen so far, Y^{t-1}. Furthermore, each channel output Y_t is governed by the noisy channel $(P_{Y|X=x})_{x\in\mathsf{X}}$, and gives rise to a control action, Z_t, performed by the computer agent, which can manifest itself in a variety of ways depending on the application (e.g., Z_t could actuate a signal in the environment or on an external device, it may represent a probability distribution update, etc).

Therefore, we can say that the encoder is governed by a collection of time-varying encoding functions $e = \{e_t : \mathsf{W}\times\mathsf{Y}^{t-1}\rightarrow\mathsf{X}\}_{t=1}^{T}$, and the decoder governed by a collection of time-varying decoding functions, $d = \{d_t : \mathsf{Y}^t\rightarrow\mathsf{Z}\}_{t=1}^{T}$, whereby each channel input X_t and decoder control action Z_t are given by

$$X_t = e_t(W, Y^{t-1}) \tag{1a}$$
$$Z_t = d_t(Y^t). \tag{1b}$$

We will see in further sections how this perspective can be applied to a wide range of scenarios, but the most general framework can be modeled as shown by Fig. 1.

A. Connection to Team Decision Theory

The human and computer are cooperating to achieve a common goal, and this can be manifested from the lens of team decision theory as a joint maximization of a reward function, subject to constraints on the information structure

$$(e^*, d^*) = \arg\max_{e,d} \sum_{t=1}^{T} \mathbb{E}\left[r_t(W, X^t, Z^t)\right] \tag{2a}$$
$$\text{s.t.}\quad X_t = e_t(W, Y^{t-1}) \tag{2b}$$
$$Z_t = d_t(Y^t). \tag{2c}$$

These classes of decentralized control problems fall into the category of "non-classical information structure" problems that are in general notoriously "difficult" to solve [11]. However, for certain types of reward functions, information-theoretic convex relaxations can be tight and solve the original problem [12].

For certain subclasses of such team decision problems, it is known [5]–[7] that there exist optimal signaling strategies for which the encoder and decoder strategies do not vary with time and obey the following minimal structure:

$$x_t = \bar{e}(w, z_{t-1}) \tag{3a}$$
$$z_t = \bar{d}(z_{t-1}, y_t). \tag{3b}$$

Here, $\bar{e} : \mathsf{W}\times\mathsf{Z}\rightarrow\mathsf{X}$ still represents the strategy embodying the human agent, modeled as the "encoder" in the classical framework, and similarly, $\bar{d} : \mathsf{Z}\times\mathsf{Y}\rightarrow\mathsf{Z}$ still represents the strategy embodying the computer agent, modeled as the "decoder" in the classical framework. One point to emphasize here, that moves toward improved usability, is that $\bar{e}$ and $\bar{d}$ here are now time-invariant, with no expense in optimal performance. The channel input X_t is a function $\bar{e}$ of the message to be conveyed, W, and the previous decision made by the computer agent, Z_{t-1}, yet the encoding process itself does not vary across time. Likewise, the next computer agent decision Z_t is a function $\bar{d}$ of the previous decision made and the current channel output, yet the decoding strategy itself does not vary across time. The exact nature of functions $\bar{e}$ and $\bar{d}$ will vary by application. It is on the time-invariant minimal structure of (3) that we will base our ongoing discussion.

B. Connection to Information Theory and Posterior Matching

In this discussion, we will primarily focus on a certain class of reward functions pertaining to maximizing how much uncertainty about W is reduced on average. More specifically, we consider maximizing the mutual information between W and the observations Y^T, which is equivalent to minimizing the amount of uncertainty about W, given the observations Y^T (e.g., minimizing posterior entropy) [8], [9]. As such, here we consider $\mathsf{Z} = \mathsf{P}(\mathsf{W})$ to be a space of probability distributions over W, and Z_t can be treated as a distribution on W given Y^t. We define the sequential information gain reward function [5] as

$$r_t(W, X^t, Z_t) = \log\frac{dZ_t}{dZ_{t-1}}(W). \tag{4}$$

If we define $\pi_t \in \mathsf{P}(\mathsf{W})$ to be the posterior distribution on W given Y^t, then for any interval A

$$\pi_t(A) = \mathbb{P}\left(W\in A|Y^t\right). \tag{5}$$

It can be shown that for the reward function given by (4), for any encoder e, the optimal decoder strategy $d : \mathsf{Y}^T\rightarrow\mathsf{Z}$ is achieved when the decision signal of the computer agent Z_t is π_t. We note that because of Bayes' rule, this can be accomplished with a minimal decoder of the form (3b)

$$\pi_t = \bar{d}(\pi_{t-1}, Y_t) \tag{6}$$
$$\bar{d}^*(\pi_{t-1}, y)(dw) \triangleq \frac{\pi_{t-1}(dw)P\left(y|\bar{e}(w,\pi_{t-1})\right)}{\beta} \tag{7}$$

where β is a normalization constant. For the duration of this work, we will consider the sequential information gain reward function (4), and without loss of generality, we will assume that $Z_t = \pi_t$. As such, for any encoder strategy e, it follows that:

$$\sum_{t=1}^{T} \mathbb{E}\left[r_t(W, X^t, Z^t)\right] = \sum_{t=1}^{T} I(W; Y_t | Y^{t-1}) = I(W; Y^T).$$

The capacity of a noisy channel $(P_{Y|X=x})_{x\in X}$ is given by maximum mutual information between the input and output of the channel over all input probability distributions $P(X)$ [13]

$$C = \max_{P(X)} I(X; Y).$$

We can therefore designate the input probability distribution that achieves capacity as

$$P_X^* = \arg\max_{P(X)} I(X; Y). \tag{8}$$

For many commonly-used channels, P_X^* is well known. An example using one such common channel, the binary symmetric channel (BSC), will be instantiated in Section II-C.

The mutual information is upper bounded by capacity

$$\frac{1}{T} I(W; Y^T) \leq C \tag{9}$$

where equality holds if and only if [13], [14]:

- X_t is statistically independent of Y^{t-1}.
- $X_t \sim P_X^*$, given in (8).

This gives rise to the following problem we consider for the remainder of this discussion:

$$(\bar{e}^*, \bar{d}^*) = \arg\max_{\bar{e},\bar{d}} I(W; Y^T) \tag{10a}$$
$$\text{s.t.} \quad X_t = \bar{e}(W, Z_{t-1}) \tag{10b}$$
$$Z_t = \bar{d}(Z_{t-1}, Y_t). \tag{10c}$$

Since $\bar{d}^*$ is given by (7), we will focus our attention on the design of $\bar{e}^*$.

Within the context of a human computer interface, maximizing mutual information is a necessary condition to guarantee that π_t converges to a point mass at W as rapidly as possible; in many situations, given other technical conditions, it is also sufficient [14]. The original 1–D posterior matching scheme in [14], as well as our group's generalization of the scheme to arbitrary dimension [10], attains the upper bound of (9) with equality. We also demonstrated in [5] and [15] that this iterative scheme can be interpreted as a sequential team decision problem consistent with that presented in Section II-A, and thus the optimal time-invariant encoder/decoder structure in (3) is applicable to the posterior matching system as well.

C. Example: One-Dimensional Binary Symmetric Channel

In this section, we will make the above discussion more concrete by providing a specific application of the posterior matching scheme in the team decision theory setting. Consider a binary symmetric channel model with crossover probability ϵ

$$P(y|x) \triangleq \mathbb{P}(Y = y | X = x) = \begin{cases} 1-\epsilon, & \text{if y=x} \\ \epsilon, & \text{if y} \neq \text{x} \end{cases} \tag{11}$$

and $\mathsf{W} = [0, 1]$, $\mathsf{X} = \mathsf{Y} = \{0, 1\}$. The encoder strategy, $\bar{e}$, updates X_t given W and π_{t-1}, and, as such, our selection of strategy should reflect that property. It is commonly known that for a BSC, the input probability distribution that achieves capacity is the uniform distribution: $P_X(0) = P_X(1) = 0.5$. Therefore, to fulfill the minimal information structure in (3a) and thereby achieve optimality, one can define the posterior matching scheme on a BSC with the following structure [14], [16]:

$$S_{t-1}(w) \triangleq \int_0^w \pi_{t-1}(dw) \tag{12a}$$
$$\lambda(w) \triangleq \begin{cases} 0, & \text{w} \leq 0.5 \\ 1, & \text{otherwise} \end{cases} \tag{12b}$$
$$W_t = S_{t-1}(W) \tag{12c}$$
$$X_t = \lambda(W_t). \tag{12d}$$

The idea here is that we can define $S_{t-1}(\cdot)$ to be the cumulative distribution function of the posterior π_{t-1} as given in (12a), and if we evaluate S_{t-1} at some point $W \in \mathsf{W}$, we generate a random variable $W_t \in \mathsf{W}$, as shown in (12c), that is uniformly distributed and statistically independent of Y^{t-1}. This is then turned into the signal X_t by simply comparing W_t to some threshold value, as shown in (12d). Here, we set this threshold to 0.5 in (12b), as that is the value that guarantees the property that $X_t \sim P_X^* = [0.5, 0.5]^T$ in this setting. We also note that the decision for X_t is indeed only a function of W and π_{t-1}, thus showing that the posterior matching scheme falls within the class of minimal schemes with structure in (3).

Keep the 1-D BSC in mind, because we will continue to revisit the model and use it as an example application as we develop and discuss the framework presented in this work.

III. PRACTICAL USE OF POSTERIOR MATCHING: THE PROBLEM OF ENGAGING A HUMAN AGENT

Although the communication scheme previously presented is indeed information-theoretically optimal and minimal in the team decision theory sense of (3), it is somewhat difficult to apply with a human agent in the loop. Referring to (12c) and (12d), the human agent plays the role of the encoder and, as such, is expected to evaluate some function λ dependent on $\tilde{W}_t$, which itself is dependent on maintaining π_{t-1} by virtue of (12a). Even with access to noiseless channel outputs Y^{t-1}, maintaining and updating the posterior, $Z_{t-1} \equiv \pi_{t-1}$, is unrealistic or outright unfeasible for a human. The challenge then becomes how to simplify the system so that information captured by π_{t-1} is consolidated in some concise way, thus simplifying the task for the human agent **while still preserving optimality**. This is a persistent problem that exists in the design of optimal human–computer interfaces. Despite the fact that the theory may lend itself toward an optimal communication scheme, whether an actual human being can manage to engage the system in an effective way is still an *interface design* challenge that must also be considered. In this section, we present a guiding framework that can potentially aid the system designer in creating such interfaces for feasible human use.

Consider the communication framework in Fig. 2. Relative to the structure of Fig. 1, we have added two components to the diagram.

- Define the function $\bar{\tau} : \mathsf{Z}_{t-1} \rightarrow \mathsf{Q}$ that represents a mapping from posterior distributions to the set of possible "queries to the human," Q. We can interpret this to mean that at every time point, the computer agent returns a single query, $Q_t \in \mathsf{Q}$, as a minimal amount of querying feedback to the human

$$Q_t = \bar{\tau}(Z_{t-1}). \qquad (13)$$

Here, if Z_{t-1} corresponds to the posterior, $\bar{\tau}$ operates directly on it to determine the query Q_t.
- Define the function $\bar{\phi} : \mathsf{W} \times \mathsf{Q} \rightarrow \mathsf{X}$ as a mapping from the latent information that the human aims to convey to the computer (W) as well as a feedback signal from the computer agent, (Q_t), to the alphabet of the noisy channel. Therefore, the next channel input X_t will be generated as

$$X_t = \bar{\phi}(W, Q_t). \qquad (14)$$

$\bar{\phi}$ and τ can be interpreted as behavioral policies being enforced on the human and computer agents, respectively, to facilitate an effective interaction between the two. Note that when comparing this scheme to the general 1-D posterior matching scheme, specifically (3) and (6), we have simply decomposed the encoder module as

$$X_t = \bar{e}(W, Z_{t-1}) = \bar{\phi}(W, Q_t) = \bar{\phi}(W, \bar{\tau}(Z_{t-1})).$$

In other words, the combination of both $\bar{\phi}$ and $\bar{\tau}$ jointly represent the encoder strategy, as the encoder operates on the posterior; however, to simplify the human agent's task, we delegate the operation on the posterior to the computer agent instead via $\bar{\tau}$. Specifically, note that the green boxes in Fig. 2 represent computer agent tasks, and the blue box represents the human agent task. We have also drawn dotted boxes around the encoder–decoder modules with respect to the original figure. Also, note a second change in this figure relative to the first is that the feedback loop to the encoder module now involves Z_{t-1} as opposed to Y_{t-1}. As the $\bar{\tau}$ task has been allocated to the computer agent, this is functionally equivalent to the previous model as $Z_{t-1} = \pi_{t-1}$, which is a sufficient statistic of Y_{t-1}. In other words, the encoder no longer pertains only to the human agent: our system takes on a somewhat different structure. As such, we will drop the vocabulary "encoder" and "decoder" for the remainder of the paper and refer only to the structure of Fig. 2 in further sections.

Here, we can also easily make the connection to the concept of maximization of mutual information from

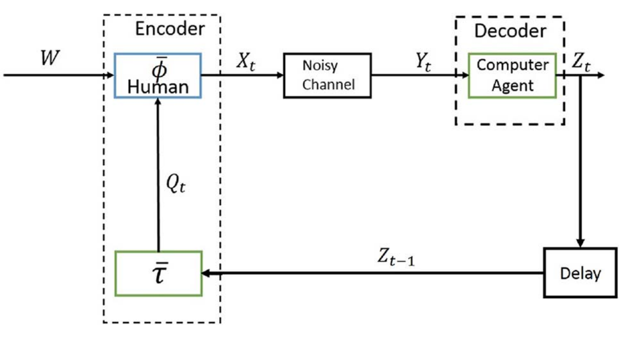


Fig. 2. ***General model for decentralized information communication approach to designing a human–computer interface, taking into the interface design constraint of usability from the human agent's perspective.***

Section II-A with respect to the new information structure involving $\bar{\tau}$ and $\bar{\phi}$

$$(\bar{\tau}^*, \bar{\phi}^*, \bar{d}^*) = \underset{\bar{\tau},\bar{\phi},\bar{d}}{\arg\max}\ I(W; Y^t) \tag{15a}$$
$$\text{s.t. } X_t = \bar{\phi}(W, Q_t) \tag{15b}$$
$$Q_t = \bar{\tau}(Z_{t-1}) \tag{15c}$$
$$Z_t = \bar{d}(Z_{t-1}, Y_t). \tag{15d}$$

A. Applying $\bar{\tau}$ and $\bar{\phi}$ to the 1-D Binary Symmetric Channel

Referring back to (12), we can now more succinctly define a new, human-convenient structure for the 1-D posterior matching system described in Section II-C, where $\mathsf{Q} = \mathsf{W} = [0, 1]$ and $\mathsf{X} = \{0, 1\}$

$$Q_t = \bar{\tau}(Z_{t-1}) = \text{median}(\pi_{t-1}) \tag{16a}$$
$$X_t = \bar{\phi}(W, Q_t) \tag{16b}$$
$$\bar{\phi}(w, q) \triangleq \begin{cases} 0, & w \leq q \\ 1, & \text{otherwise} \end{cases}. \tag{16c}$$

To summarize, at every time step, the computer agent need only calculate the posterior distribution π_{t-1} and relay the median of the distribution to the human agent as Q_t. From the perspective of the human, their task becomes almost trivially simple now: given Q_t, the human need only signal whether the message point W is greater than or less than Q_t. This simple comparison yields the uniform input distribution we desire as well, as W has equal probability of lying on either side of Q_t from the perspective of the computer agent and its calculated posterior distribution. An important interpretation of the role of the query point Q_t is that it is the computer agent's "most informative question" to the end of learning what W actually is.

B. Selection of $\bar{\tau}$ and $\bar{\phi}$

With the framework presented here, we are introducing a model by which a system designer can imagine delegating tasks in the design of a human–computer interface so that the system not only remains optimal from an information-theoretic perspective but also becomes practical and realistic. Information and control optimality means almost nothing when humans are involved if a human cannot realistically utilize the system in an efficient manner in the first place. However, it should be noted that the selection of Q, $\bar{\tau}$ and $\bar{\phi}$ is dependent on the application that the system is to be tailored for, and is therefore a process that is difficult to standardize. While the selection of $\bar{\tau}$ and $\bar{\phi}$ is indeed inherently coupled to the nature of the channel model (specifically, they are directly related to P_X^*), there will usually never be a unique solution and is therefore subject to the creativity of the system designer as well. The extremely simple guideline that we generally find useful in using this model is the following: **select $\bar{\phi}$ to be simple, and $\bar{\tau}$ to be complicated**.

In previous work, our group has designed systems that fall into the class of systems represented by this framework for various applications. Using the simple 1-D BSC model presented here, our group has designed and validated brain–computer interfaces for the specification of paths in 2-D space as well as words of arbitrary length using an arithmetic coding scheme. These Iare examples of how theory and creativity can both be applied to the system for practical human-friendly applications that are also information-theoretically optimal. [17]–[19].

Remark 1: Clearly, selection of the $(\bar{\tau}, \bar{\phi})$ pair is, in general, up to the system designer and does not yield a unique pair for any given channel model. As such, one can just as easily select a pair that is valid with respect to the model we have presented, but that is not user-friendly. For a very naive example, consider a BSC scenario with $\mathsf{W} = [0, 1]$ where the at any time t, the posterior is piecewise-constant [16]. Consider encoding an approximation of the piecewise-constant posterior as a real number on the [0, 1] line as follows. At time t, if there are K piecewise-constant intervals in the posterior, we can encode every kth interval's probability as a sequence of bits $B_k = 0.b_0^k b_1^k b_2^k \ldots b_n^k$, where n is some number of bits required to encode the smallest probability value in the posterior. Suppose N_0 is a bit representation of n, then we can effectively represent the entire posterior as a concatenation of these smaller sequences, prefaced by N_0: $Q_t = 0.N_0 B_0 B_1 B_2 B_3 \ldots B_K$.

Although this strategy is consistent with the information in (15), giving the human the entire posterior to work with imposes a "complicated" operation on the part of the human if $\bar{\phi}$ is the process of comparing W to the median of the posterior.

Furthermore, others in the community have also established this 1–D example in which $\mathsf{W} = \mathsf{Q} = [0, 1]$ using query point representations as in (16) and applied it to human–computer interface applications [8], [9].

In the following sections, we will demonstrate that this simple example underlies a viewpoint from the theory of optimal transport that allows for an extension to when W is a subset of arbitrary-dimensional Euclidean space, thus allowing the framework to accommodate even more complex information models, and giving the system designer more creative freedom as well. We will then demonstrate the true power of this framework by showcasing a general class of optimal queries through the optimal maps lens and instantiate this with a collaborative brain–computer interface.

IV. GENERALIZING POSTERIOR MATCHING FOR HIGHER DIMENSION PROBLEMS

The previous discussion applies to scenarios in which $\mathsf{W} \subset \mathbb{R}$. Recently, we have also shown that the framework can be generalized to arbitrary dimension [10], namely $\mathsf{W} \subset \mathbb{R}^d$. We will first present a high-level theoretical discussion of the arbitrary dimension scenario, and at the end of this section, we will present an experimental implementation of the framework.

A. Posterior Matching in Arbitrary Dimension

Consider a scenario with $\mathsf{X}, \mathsf{Y}, \mathsf{W} \subset \mathbb{R}^d$, with some prior distribution P_W that is uniform over W (which we assume to be convex and compact) and P_X a probability mass function over a finite subset of $\mathbb{R}^d$. We would like to build a generalized scheme that provides analogous definitions to those of (12).

Consequently, we still seek some function $S_{t-1} : \mathsf{W} \rightarrow \mathsf{W}$ and some function $\lambda : \mathsf{W} \rightarrow \mathsf{X}$ analogous to the corresponding functions in (12). As mentioned previously, in order to maximize mutual information as in (9), it must be that $P_{X_t|Y^{t-1}} \equiv P_X^*$. We can achieve this by defining $X_t = \lambda(W_t)$, and by enforcing the following independence and stationarity property:

$$P_{W_t|Y^{t-1}=y^{t-1}} = P_{W_t} = P_W. \tag{17}$$

We can accomplish this by performing a construction $W_t = S_{t-1}(W)$, where S_{t-1} transforms the posterior $\pi_{t-1} \equiv P_{W|Y^{t-1}=y^{t-1}}$ to P_W so that $W_t = S_{t-1}(W)$ obeys (17). As discussed in [10], we can utilize the theory of optimal transport to this end. Given two probability measures $P, Q \in \mathsf{P}(\mathsf{W})$, it is our role to design a map S that pushes P to Q, meaning that if a random variable V with distribution P is transformed into a random variable U via $U = S(V)$, then U has distribution Q. Under appropriate technical conditions (about P having a density with respect to the Lebesgue measure), the Monge-Kantorovich L^2 optimal transport theory [27] shows that there always exists such a map

$$S^* \triangleq MK_{L^2}(P, Q)$$

which is also invertible and monotonic. Moreover, for large classes of problems where P and Q have a log-concave structure, we have shown that identifying the unique invertible and monotonic map can be solved efficiently with convex optimization methods [20]–[22]. We have shown in [10] that by defining $P = \pi_{t-1}$ to be the posterior and $Q = P_W$ to be the prior, then we can build $S_{t-1} = MK_{L^2}(\pi_{t-1}, P_W)$ so that it generalizes (12a) for arbitrary dimension. Analogously, by defining $P = P_W$ to be the prior on W and $Q = P_X^*$, the capacity achieving distribution for X, then the following scheme maximizes mutual information [10, Th. 4.3]:

$$S_{t-1}(w) \triangleq MK_{L^2}(\pi_{t-1}, P_W) \tag{18a}$$

$$\lambda \triangleq MK_{L^2}\left(P_W, P_X^*\right) \tag{18b}$$

$$W_t = S_{t-1}(W) \tag{18c}$$

$$X_t = \lambda(W_t). \tag{18d}$$

Note the similarity between (18) and (12); indeed, the latter is a special case of the former for $\mathsf{W} = [0, 1]$ and the binary symmetric channel. Once again, it is also clear that this scheme fulfills the functional requirements of (3). Also note, however, that although we present a strict definition of λ as some arbitrary map that pushes P_W to P_X^*, we can often simplify this function by defining the encoder rule explicitly and consistently with the noisy channel model in question (similar to how we designated the higher/lower comparison presented in (12d) for the BSC case).

B. A Visual Example of Posterior Matching in Two-Dimensions

One way to interpret (18c) is that the map S_{t-1} "warps" the message space W to transform the random variables. To illustrate this, consider a scenario in which $\mathsf{W} = [0, 1]^2$ is the unit square. We begin with a uniform prior $\pi_0 = P_W$ at time step 0; we can then make an observation giving rise to the posterior π_1. Next, we construct a map $S_1 = MK_{L^2}(\pi_1, P_W)$, where P_W is uniform. If we evaluate $S_t(W)$, we essentially "warp" points from the posterior to a new warped space; from the perspective of this new warped space, the posterior once again appears uniform, with the space itself changing form. Fig. 3 illustrates this process more clearly using an image to represent the space being warped.

Thus, in the general structure of posterior matching in arbitrary dimension, we can imagine that W_t in (18c) is a point in this warped space, and so the encoder determines the next channel input X_t according to the function λ operating on the warped point.

C. Human-Compatible Posterior Matching Schemes in Higher Dimensions

We now consider the generalized posterior matching scheme given by (18) and, without sacrificing optimality, transform it to a more human-friendly model, analogous to the 1-D case in Section III and Fig. 2.

To make the system more consistent with a higher-dimension version of the framework presented in (16), we will define the noisy channel input of the model as

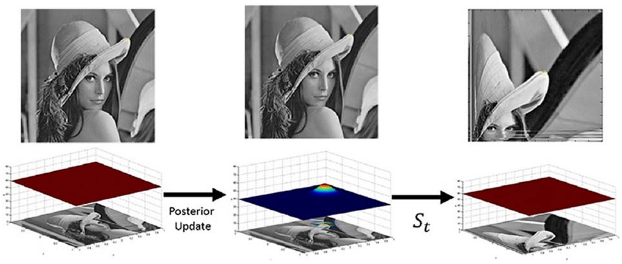


Fig. 3. ***Example of application of S_t map to warp the space of posterior, given a mapping $S_t = MK_{L^2}(\pi_t, P_W)$, where P_W is uniform. The first step of the image is the regular posterior update after making an observation Y_t. However, by using the map S_t, we can make the posterior appear "uniform" once again by warping the actual space itself accordingly to give rise to a new "warped" space; the image illustrates this warping of the space. (Notice that no part of the image is actually "lost" during the warping, it is simply resized and stretched to accommodate the simultaneous stretching of the distribution mass.)***

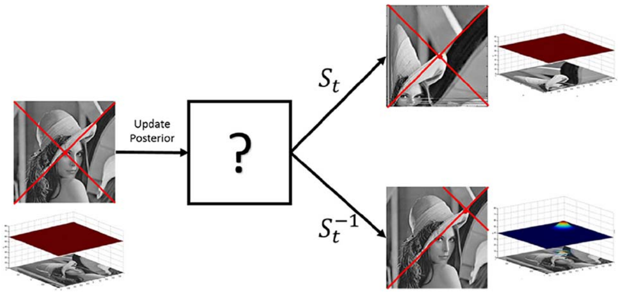


Fig. 5. ***The two equivalent scenarios of query point comparison are shown, with $Q_1 = [0, 0]^T$. After the posterior update, the top case shows the scenario in which the space warps to uniform such that $X_t = \phi_t(W_t, Q_1)$, where $W_t = S_t(W)$ and S_t is given by (18a). The bottom case shows the scenario in which we instead move the dividing lines to reflect where the positioning of the new Q_t should be such that $X_t = \phi_t(W, Q_t)$ where $Q_t = S_t^{-1}(Q_1)$. Notice that the bottom case is more desirable from a human-computer interface perspective.***

$X_t \in \{0, 1, 2, 3\}$. We can provide an analogous design to $\bar{\phi}$ in (16c) but for the 2-D case as follows.

Consider a geographic map that represents a 2-D space. We can divide this map, with $\mathsf{W} = [-1, 1]^2$, into four quadrants, as shown in Fig. 4. The query space Q here represents the set of possible pairs of curves in W that define the boundaries to the four regions at time t $(\Gamma_{k,t})_{k=0:3}$ that are the inverse image of λ evaluated at $X = k$

$$\{W \in \Gamma_{k,t}\} = \{X_t = k\} \quad \mathbb{P} - a.s. \tag{19}$$

Note that since $X_t \in \{0, 1, 2, 3\}$, it follows that $\{\Gamma_{0,t}, \Gamma_{1,t}, \ldots, \Gamma_{3,t}\}$ form a partition of W. For $t = 1$, when we have not conditioned upon anything, then we have that $\Gamma_{1,k} \equiv \Gamma_k$, where Γ_k is given by each of the regions in Fig. 4.

As such, $\bar{\phi}(W, Q_1)$ allows us to extend the 1-D BSC to two dimensions and the human now only needs to signal in which quadrant their point of interest, W, lies. In this way, we can utilize a four-input noisy channel model. Without loss of generality, we assume that P_X^* is equally likely to be in {0, 1, 2, 3}, and Q_t now represents some specific division of the map into four quadrants that changes at every iteration. As such, Q_t will now represent a set of dividing boundaries presented to the human. Similar to the BSC, we can take advantage of our knowledge that P_X^* must be uniform to achieve optimality by designing the computer to select Q_t as the set of dividing boundaries that separates the image into four regions of equal posterior probability.

In multiple dimensions, it is normally a nontrivial task to identify the optimal query boundaries, Q_t in any scenario. In the context of our 2-D map example, the challenge is to identify the boundaries for which the four regions have equal posterior probability.

We now define $\Gamma_{k,t}$ in terms of Γ_k and S_{t-1} that pushes π_{t-1} to $\pi_0$as

$$\Gamma_{k,t} = S_{t-1}^{-1}(\Gamma_k). \tag{20}$$

With this, we can state our main theorem:

Theorem IV.1: Assume that $\mathsf{W} \subset \mathbb{R}^d$, that $|\mathsf{X}| = 2^d$, and P_X^* is the distribution on X that maximizes $I(X; Y)$. Then there exists an (ϕ^*, τ^*, d^*) that is an optimal solution to (2), (10), (15), where $\phi(W, Q_t)$ is given by combining (19) with (20) as follows:

$$\{X_t = k\} = \left\{W \in S_{t-1}(\Gamma_k)\right\} \tag{21}$$
$$Q_t = S_{t-1}^{-1}(Q_1). \tag{22}$$

Fig. 4. ***An example of how the dividing lines designated by Q_t evolve over time as the computer receives outputs from the noisy channel and updates its posterior. At each time point, the dividing lines are changed in a way that maximizes the amount of information learned from the human. More concretely, Q_t is selected as the boundaries that divide the space such that the resulting quadrants have equal mass within them. Furthermore, the figure illustrates one way that the quadrants can potentially correspond to X_t, with the northern quadrant corresponding to $X_t = 0$, and so forth.***

For instance, if $d = 2$ and P_X^* is uniform over X, then we have that $\Gamma_0, \ldots, \Gamma_3$ are defined in terms of $Q_1 = \{p_1, p_2\}$, where p_1 is the line segments connecting $(-1, -1)$ to (1, 1), and p_2 is the line segment connecting $(-1, 1)$ to $(1, -1)$. Then by (22), we mean that $Q_2 = \{S_1^{-1}(p_1), S_1^{-1}(p_2)\}$.

Proof: Define Q_1 as the collection of boundaries as mentioned above, for which

$$\frac{\mathrm{vol}(\Gamma_k)}{\mathrm{vol}(W)} = P_X^*(k) \tag{23}$$

and not that clearly $X_1 \sim P_X^*$. Given Y_1, define $z_1 = \pi_1 = d_1^*(\pi_0, y)$ given by the Bayes update. Define $S_1 = MK_{L_2}(\pi_1, P_W)$ and define $Q_2 = \tau_1(Y_1) \equiv S_1^{-1}(Q_1)$. Then note that for any $k \in \{0, 1 \ldots, 2^d - 1\}$

$$\begin{aligned} \mathbb{P}\{X_2 = k|Y_1\} &= \mathbb{P}(\phi(W, Q_2) = k|Y_1) \\ &= \mathbb{P}\big(W \in S_1^{-1}(\Gamma_k)|Y_1\big) && (24) \\ &= \mathbb{P}(S_1(W) \in \Gamma_k|Y_1) && (25) \\ &= \mathbb{P}(W_2 \in \Gamma_k|Y_1) && (26) \\ &= \mathbb{P}(W_2 \in \Gamma_k) && (27) \\ &= \mathbb{P}(W \in \Gamma_k) && (28) \\ &= P_X^*(k) && (29) \end{aligned}$$

where (24) follows from (21); (25) follows from the Monge-Kantorovich theory for which S_t at any t is invertible and monotonic; (26) follows from the formal definition of $W_2 = S_1(W)$; (27) follows because Γ_k is a non-random set and W_t is independent of Y_1 by virtue of (17); (28) follows from (18c), (18a), and (17); and (29) follows from (23).

Thus, $I(W; Y^2) = 2C$. More generally, define recursively $\pi_t = d_1^*(\pi_{t-1}, Y_t)$, $S_t = MK_{L_2}(\pi_t, P_W)$, and $Q_t = S_{t-1}^{-1}(Q_1)$. Then, it follows that from induction that $I(W; Y^T) = TC$ and the human encoder obeys the BCI-compatible scheme in Definition (18). □

Note that, conceptually, comparing $S_1(W)$ to Q_1 and comparing W to $S_t^{-1}(Q_1)$ are functionally equivalent operations, the only difference being the representation shown to the human. In the former case, the human would be comparing a "warped" version of the message point (in a warped message space) to some fixed center point of the image, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, as this corresponds to the initial set of boundaries denoted by Q_1, whereas in the latter, the human would be comparing the original message point (in the original message space) to a modified dividing boundary in the original unwarped space.

Fig. 5 illustrates this in more detail. In designing the system to be more human-friendly, notice we have naturally chosen the latter scenario for our implementation, as it makes the task far more intuitive to the human.

V. A TWO-DIMENSIONAL HUMAN-FRIENDLY APPLICATION

In this section, we will describe a 2-D application of the posterior matching framework applied to a brain–computer interface (BCI).

A. The Goal of the Task

The goal of this BCI is to allow a human user to specify a point on a geographic map to a computer system using only brain signals measured using electroencephalography (EEG). The notable concepts of merit that we wish to emphasize with this example are two-fold.

- By using the optimal transport method, we can still efficiently optimize the selection of the query point Q_t, even in high dimension, so that the system is still able to achieve optimality.
- The system utilizes our human-friendly framework so that the task imposed on the human is as simple as possible.

We will leverage motor imagery [23]–[25], a particularly well-studied signaling method for brain-computer interfaces, as our input signal of choice for facilitating communication between the subject and the computer. Motor imagery refers to an imagined movement of certain parts of the body such that the movement being imagined can be determined by observing the subject's EEG signals. One particular reason for our selection of motor imagery as our input signal is that it has also been demonstrated to be reasonably easy to obtain for most subjects. In the context of our BCI, we will utilize motor imagery to acquire a binary (0 or 1) signal from the subject by asking the subject to imagine moving either their left hand or their right hand. It is well known that this type of motor imagery can be characterized by changes in EEG power within the Mu (8–12 Hz) band between the left and right sides of the brain. Fig. 6 illustrates and describes these power-changes in slightly more detail.

The quad-symmetric channel will be governed by the following probabilistic law:

$$P(y|x) = \begin{cases} 1 - \epsilon, & \text{y = x} \\ \frac{\epsilon}{3}, & \text{otherwise} \end{cases} \tag{30}$$

for some probability of error ϵ. Conceptually, this noisy channel intuitively models the possibility of error in classifying EEG signals.

Given that we have restricted ourselves to the binary nature of left versus right motor imagery, we can still produce our four-symbol input alphabet for X_t by dividing each iteration into two subproblems, as illustrated by Fig. 7. We then simply ask the human whether W lies to the left or right of the two individual dividing lines. From the two binary responses, we can then determine which quadrant W lies in with respect to the original image.

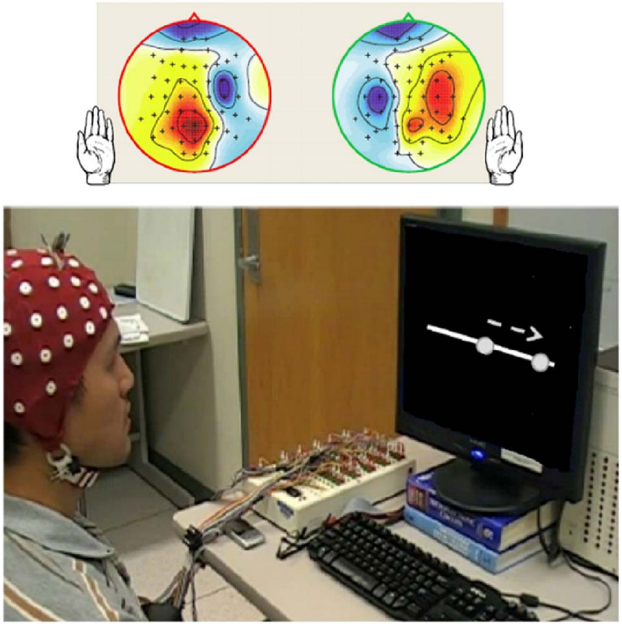
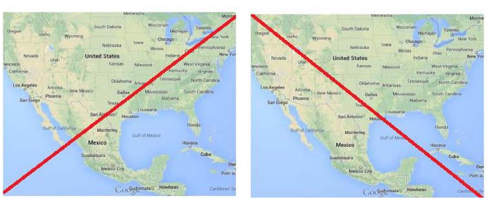


Fig. 6. ***An example EEG setup in which the subject engages in motor imagery to try to drive a ball left and right on a computer screen. Motor imagery can be characterized by a shift in EEG power within the Mu (8–12 Hz) band between one side of the motor cortex and the other; upon engaging in motor imagery on the left side of the body, Mu power increases on the left side of the motor cortex, and vice versa when engaging in right-side motor imagery.***

B. Preliminary Experiment: A Cooperative BCI

By utilizing the framework described above, we first generated simulations of the system in action (Fig. 8). Here, we task the computer agent with converging on a specific point on a map of University of California at San Diego. On average, the simulation required less than 10 iterations to converge on the proper message point. Fig. 8 depicts both a subset of query images at time points 1, 2, 6, and 9, as well as the corresponding posterior distributions, which clearly demonstrate convergence to the true point of interest.



y	Quad(y)
[0,0]	3
[0,1]	0
[1,0]	2
[1,1]	1

Fig. 7. ***An example of how we can separate the map problem into two subproblems. We can effectively produce a four-input signal by asking the human to designate whether their W lies to the left or right of each dividing line. The combination of their responses can then be used to determine which quadrant of the original image W lies in.***

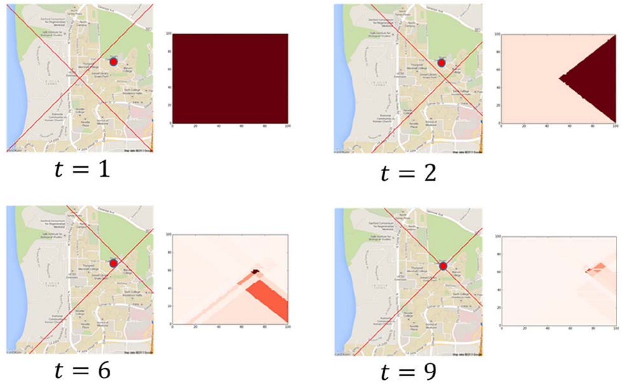


Fig. 8. ***A simulation of the 2-D BCI. After hardcoding W to be a specific point on the map (in this particular simulation, $W = (75, 60)$, with respect to the posterior axes, and is represented by the red dot), we observe rapid convergence to the true message point, as Q_t begins to center itself around W.***

Following our simulations, we then carried out actual experiments using human subjects generating motor imagery EEG signals to control the BCI. However, to be more provocative, we decided to utilize two humans engaging in a multi-user cooperative task whereby each subject solves one of the two subproblems presented above (Fig. 9). In essence, we showed how you can utilize multiple humans with a shared goal/state of mind to solve this problem even more quickly. In our

Fig. 9. ***A preliminary experiment designed to demonstrate the ability of the framework to accommodate real-world noisy EEG signals to converge on a point on a 2-D map in real time. The two subjects initially agree on a message point that they would cooperatively attempt to signal to the computer system and proceed to answer the binary left-right questions presented to them with respect to the agreed-upon message point. In this experiment, we observe the same functional convergence behavior as we saw in our simulations.***

experiment, we observed the same convergence behavior as in our simulations (for details on the motor imagery classifier used in this experiment, see [26]).

VI. DISCUSSION AND CONCLUSION

We have presented a general framework that may be used in the design of novel human–computer interfaces in the presence of arbitrary sources of noise as well as in scenarios in which the input signals from the human to the computer may be limited. With our design, we showcase the importance of achieving optimality between the human and the computer from an information-theoretic standpoint, and have presented examples of systems that achieve just that. We also emphasize the importance of separating tasks between the computer and human so that the tasks delegated to the human remain cognitively simple enough so as to not impose unnecessary difficulty on the subject, which may ultimately introduce additional noise into the system.

By utilizing posterior matching [14] in combination with the optimal transport method of finding optimal maps [10], we have also provided a framework for easy and optimal communication of potentially high-dimensional message points between a human and computer. We then instantiated this method with a high-level, proof-of-concept example utilizing the EEG signals from two separate subjects with a shared goal to effectively communicate a message point in 2-D space to the computer. This system demonstrated optimal convergence behavior in a practical setting, even while using a noisy source of information, such as EEG.

One point we would like to emphasize even further with regards to the work here is that while we have provided a few simple example instantiations of the framework, the true strength of the framework lies in the fact that W can lie in arbitrarily high-dimensional spaces, and even in those scenarios, it can still be reliably communicated. As such, W can be used as a representation of a broad range of information types beyond the literal point in $\mathbb{R}^d$ (e.g., a parameter vector of a basis function). We can imagine scenarios in which perhaps the goal of the system is to identify some common piece of information held by multiple humans or to characterize some common mode of behavior among many humans using a parameterized function. This framework may lend itself to being an extremely viable solution to solving such problems.

To achieve optimality, the designer of a system that uses this framework need guarantee only that the channel capacity maximizing conditions presented in Section II-B are met.

By utilizing this framework in designing future human–computer interfaces, we can guarantee that we are utilizing a communication channel between a human and computer in the most optimal way possible from an information-theoretic perspective. Furthermore, as a result of keeping the task of the human as simple as possible, we can design systems that apply to a very wide audience of potential users, which is a concept that should always be kept in mind when designing such systems for use in practical human scenarios. ■

REFERENCES

[1] W. B. Knox and P. Stone, "Tamer: Training an agent manually via evaluative reinforcement," in *IEEE 7th Int. Conf. Develop. and Learning (IDCL)*, 2008, pp. 292–297.

[2] N. A. Vien, W. Ertel, and T. C. Chung, "Learning via human feedback in continuous state and action spaces," *Applied Intelligence*, vol. 39, no. 2, pp. 267–278, 2013.

[3] C. Meriçli, M. Veloso, and H. L. Akin, "Complementary humanoid behaviour shaping using corrective demonstration," in *IEEE-RAS 10th Int. Conf. Humanoid Robots (Humanoids)*, 2010, pp. 334–339.

[4] C. Celemin and J. Ruiz-del-Solar, "Interactive Learning of Continuous Actions from Corrective Advice Communicated by Humans", in *Robot Soccer World Cup*, Berlin, Germany: Springer-Verlag, 2015, pp. 16–27.

[5] S. Gorantla and T. Coleman, "Information-theoretic viewpoints on optimal causal coding-decoding problems," 2011, [Online]. Available: https://arxiv.org/abs/1102.0250

[6] J. C. Walrand and P. Varaiya, "Optimal causal coding-decoding problems," *IEEE Trans. Inf˙. Theory*, vol. 29, no. 6, pp. 814–820, 1983.

[7] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 5, pp. 679–694, 1989.

[8] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Collaborative 20 questions for target localization," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2233–2252, 2014.

[9] B. Jedynak *et al.*, "Twenty questions with noise: Bayes optimal policies for entropy loss," *J. Appl. Probability*, vol. 49, no. 1, pp. 114–136, 2012.

[10] R. Ma and T. P. Coleman, "Generalizing the posterior matching scheme to higher dimensions via optimal transportation," in *49th Annu. Allerton Conf. Commun., Control, Comput.*, 2011, pp. 96–102.

[11] Y.-C. Ho, "Team decision theory and information structures," *Proc. IEEE*, vol. 68, no. 6, pp. 644–654, 1980.

[12] A. A. Kulkarni and T. P. Coleman, "An optimizer's approach to stochastic control problems with nonclassical information structures," *IEEE Trans. Automat. Control*, vol. 60, no. 4, pp. 937–949, Apr. 2015.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*, 2nd ed., Hoboken, NJ, USA: Wiley, Jul. 2006.

[14] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, 2011.

[15] R. Ma, N. Aghasadeghi, J. Jarzebowski, T. Bretl, and T. P. Coleman, "A stochastic control approach to optimally designing hierarchical flash sets in p300 communication prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 1, pp. 102–112, 2012.

[16] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, no. 3, pp. 136–143, 1963.

[17] C. Omar *et al.*, "A feedback information-theoretic approach to the design of brain–computer interfaces," *Int. J. Human-Comput. Interaction*, vol. 27, no. 1, pp. 5–23, 2010.

[18] T. P. Coleman, "A stochastic control viewpoint on posterior matching-style feedback communication schemes," in *IEEE Int. Symp. Inf. Theory*, 2009, pp. 1520–1524.

[19] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, no. 6, pp. 520–540, 1987.

[20] S. Kim, R. Ma, D. Mesa, and T. P. Coleman, "Efficient bayesian inference methods via convex optimization and optimal transport," in *Proc. IEEE Int. Symp. Inf. Theory*, 2013, pp. 2259–2263.

[21] S. Kim, C. J. Quinn, N. Kiyavash, and T. P. Coleman, "Dynamic and succinct statistical analysis of neuroscience data," *Proc. IEEE*, vol. 102, no. 5, pp. 683–698, May 2014.

[22] D. Mesa, S. Kim, and T. P. Coleman, "Efficient Bayesian inference methods via convex optimization and optimal transport," in *Proc IEEE ISIT*, 2013, pp. 2259–2263.

[23] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. Lopes da Silva, "Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks," *Neuroimage*, vol. 31, no. 1, pp. 153–159, 2006.

[24] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–325, 1998.

[25] L. Qin, L. Ding, and B. He, "Motor imagery classification by means of source analysis for brain–computer interface applications," *J. Neural Eng.*, vol. 1, no. 3, p. 135, 2004.

[26] M. McCormick, R. Ma, and T. P. Coleman, "An analytic spatial filter and a hidden Markov model for enhanced information transfer rate in EEG-based brain computer interfaces," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2010, pp. 602–605.

[27] C. Villani, *Optimal Transport: Old and New*, Germany: Springer-Berlin, 2008.

ABOUT THE AUTHORS

Justin Tantiongloc received the B.S. degree in computer engineering from the University of California at Davis, Davis, CA, USA, in 2011, and the M.S. degree in computer engineering from the University of California at San Diego, La Jolla, CA, USA, in 2014. He is currently pursuing the Ph.D. degree in computer engineering, also from the University of California, San Diego, under the guidance of Dr. Todd P. Coleman.

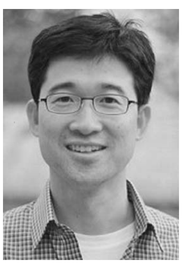
His research interests are rooted in the integration of software engineering and machine learning and mathematics with human–computer interfaces, and a variety of bioengineering-related applications.

Diego A. Mesa (Student Member, IEEE) received the B.S. degree in computer engineering from the University of Florida, Gainesville, Florida, USA, in 2010. He is currently pursuing the Ph.D. degree in bioengineering at the University of California at San Diego, La Jolla, CA, USA, in the Neural Interaction Laboratory.

His research interests include information theory, data science, and machine learning and statistical signal processing with applications to health care data.

Rui Ma (Member, IEEE) received the B.S. degree in neurobiology and biophysics from the University of Science and Technology of China, Hefei, China, in 2004, and the M.S. degree in applied mathematics and the Ph.D. degree in neuroscience, both in 2008, from the University of Illinois Urbana-Champaign, Champaign, IL, USA.

He was a Postdoctoral Scholar in Prof. T. Coleman's lab since 2008 until he joined Dexcom Inc., San Diego, CA, USA, in 2014. He is now working on wearable continuous glucose monitors.



Sanggyun Kim (Member, IEEE) received the Ph.D. degree in electrical engineering and computer science (EECS) from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2008.

He joined the Statistical Learning for Signal Processing Laboratory in EECS at KAIST in 2001. From January 2009 to June 2010, he was a Postdoctoral Researcher in the Department of Brain and Cognitive Sciences at the Massachusetts Institute of Technology, Cambridge, MA, USA. Dr. Kim is currently an Assistant Project Scientist in the Department of Bioengineering at the University of California at San Diego, La Jolla, CA, USA. His research interests include data science and machine learning and statistical signal processing with applications to neuroscience and health care data.

Cristian H. Alzate (Student Member, IEEE) was born in Cali, Colombia, in 1987. He received the B.E. degree in electronic engineering from the University of Valle, Cali, Colombia, in 2010. He also studied computer system engineering at the same university. He received the Ms.C. degree in electrical engineering from the University of Puerto Rico at Mayagüez (UPRM), Mayagüez, Puerto Rico, in 2014.

From 2011 to 2012, he was an Instructor in the Electrical and Computer Engineering department at UPRM. From 2012 to 2014, he was a Researcher in the Brain Computer Interface Lab at UPRM. He is currently the Leader of the development and research team at TPiPR, San Juan, Puerto Rico. His research interests include machine learning, artificial intelligence, statistical signal processing, and applied mathematics.

Jaime J. Camacho (Student Member, IEEE) received the B.S. degree in electrical engineering from the University of Puerto Rico at Mayagüez (UPRM), Mayagüez, Puerto Rico, in 2013. He is currently pursuing the M.S. degree under the supervision of Dr. Vidya Manian.

Since 2014, he has been working as a Researcher in the Brain Computer Interface Laboratory at UPRM. His research interests include imaging and statistical signal processing for neuroscience and biomedical applications.

Vidya Manian (Member, IEEE) received the Ph.D. degree in computing, information science and engineering from the University of Puerto Rico, Mayagüez (UPRM), Mayagüez, Puerto Rico, in 2004.

She was a Visiting Scholar at the Lane Department of Computer Science and Electrical Engineering at West Virginia University, Morgantown, West Virginia, USA, during the writing of this paper. In 2005, she was a Postdoctoral Associate in Electrical and Computer Engineering (ECE) at UPRM. In 2006, she joined ECE, UPRM, as an Assistant Professor. She is currently a Professor in the department of ECE, UPRM and the Director of the Brain Computer Interface Laboratory. Her research interests include signal and multispectral image processing, neuroscience and machine learning.

In 2013, the First Lady of Puerto Rico recognized Dr. Manian for her contribution to education.

Todd P. Coleman (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2005.

He was a Postdoctoral Scholar in neuroscience at MIT and Massachusetts General Hospital during the 2005–2006 academic year. He was an Assistant Professor in Electrical and Computer Engineering as well as Neuroscience at the University of Illinois at Urbana–Champaign, Champaign, IL, USA, from 2006 to 2011. He is currently an Associate Professor in Bioengineering, Principal Investigator of the Neural Interaction Laboratory, and co-Director of the Center for Perinatal Health at the University of California at San Diego, La Jolla, CA, USA. His research interests lie at the intersection of applied mathematics, bio-electronics, neuroscience, and medicine.

Dr. Coleman was named a Gilbreth Lecturer for the National Academy of Engineering in 2015.