Preference Elicitation in Fully Probabilistic Design of Decision Strategies

Miroslav Kárný, Tatiana V. Guy

Abstract— **Any systematic decision-making design selects a decision strategy that makes the resulting closed-loop behaviour close to the desired one. Fully Probabilistic Design (FPD) describes modelled and desired closed-loop behaviours via their distributions. The designed strategy is a minimiser of Kullback-Leibler divergence of these distributions. FPD: i) unifies modelling and aim-expressing languages; ii) directly describes multiple aims and constraints; iii) simplifies an (inevitable) approximate design as it has an explicit minimiser.**

The paper enriches the theory of FPD, in particular, it: i) improves its axiomatic basis; ii) quantitatively relates FPD to standard Bayesian decision making showing that the set of FPD tasks is a dense extension of Bayesian problem formulations; iii) opens a way to a systematic data-based preference elicitation, i.e., quantitative expression of decision-making aims.

I. INTRODUCTION

A systematic choice of the optimal decision strategies, mapping available knowledge on optional actions influencing a World's part, is the common topic of decision-making [1], [2] and control communities [3]. CDC is one of still rare events where these communities meet and are aware that they address essentially the same problem. The paper uses a terminology closer to dynamic statistical Decision Making (DM) but stresses closed loop as the central notion of control theory. The paper further develops Fully Probabilistic Design (FPD) of DM strategies, in particular, it

- makes the axiomatic basis of FPD more firm comparing to the preliminary version published in [4];
- enhances the application potential of FPD by explicitly relating it to the standard Bayesian DM;
- provides a variant of FPD that opens a way to quantitative, data-based, elicitation of DM preferences.

The decision strategy influences the joint distribution of variables in the optimised closed loop. The standard Bayesian design selects the strategy minimising expectation of the performance index expressing DM preferences, e.g., [2]. The discussed FPD, [5], [6], [7], chooses the strategy that makes the joint distribution of the closed-loop variables as close as possible the ideal distribution expressing DM preferences. FPD features motivating its development are:

- The ideal distribution well respects constraints and multiple aims [8], [9].
- FPD possesses an explicit minimiser in (minimisation, expectation)-pairs forming the DM design [3], which simplifies an (almost always) inevitable approximate design.
- Probabilities describe both models and decision aims. Consequently, knowledge elicitation methods [10] can be applied to the important and underdeveloped domain of the preference elicitation, i.e., quantitative expression of DM aims. Moreover, this expression reflects deterministic and non-deterministic properties of the closed DM loop.

Section II reviews DM theory exemplifying conditions leading to FPD. Section III: i) maps the standard Bayesian design on FPD; ii) opens a way to data-based learning of the ideal distribution; iii) indicates how this learning is to be prepared. Section IV provides concluding remarks.

Throughout, $\equiv$ is defining equality. $X^\star$ denotes a set of Xs with X being parameter, strategy, uncertainty, etc. $X_\star$ denotes a subset of $X^\star$. Sans serif fonts denote mappings.

Subscripts $_{i,\ j,\ k}$ distinguish different elements of an inspected set. The symbol $\prec$ denotes a strict ordering of set. The distinction between various orderings is obvious from the nature of compared elements.

II. DECISION MAKING UNDER UNCERTAINTY

Decision making works with sequences $X \equiv (X_t)_{t\in t^\star}$ labelled by discrete-time moments, $t \in t^\star \equiv \{1, \dots, h\}$, at which actions are taken. $h < \infty$ is a given decision horizon.

DM consists of the selection and application of DM *strategy*, i.e., a sequence of mappings $\mathsf{S} \equiv (\mathsf{S}_t)_{t\in t^\star}$. It is formed by *decision rules* S_t, $t \in t^\star$. A strategy maps a *knowledge* sequence $K \equiv (K_t)_{t=0}^{h+1}$ on an *action* sequence $A \equiv (A_t)_{t\in t^\star} \in A^\star \neq \emptyset$. Actions describe or influence *environment*, the World's part of interest.

The coupling of the strategy and the environment forms a closed loop. Its *behaviour* $B \in B^\star \neq \emptyset$ is identified with all variables observed, opted or considered in the closed loop. The set of *admissible strategies* $\mathsf{S}^\star$ is formed by a sequence of decision rules that are causal from information view point, i.e., $\mathsf{S}_t(K) = \mathsf{S}_t(K_t)$. The strategy S is selected from a set of *compared strategies* $\mathsf{S}_\star \neq \emptyset$, which is usually a proper subset of admissible strategies $\mathsf{S}_\star \subsetneq \mathsf{S}^\star$.

The processed knowledge sequence $K \in K^\star$ is assumed to be non-shrinking, i.e., the knowledge K_t is extended by *observations* $\Delta_t \in \Delta_t^\star \equiv K_{t+1}^\star \setminus K_t^\star$, $t \in t^\star$. By its definition, the behaviour contains, among others, opted and observed data $D_t \equiv (A_t, \Delta_t),\ t \in t^\star$. The knowledge K_t, available at time t for choosing the action A_t, coincides with $(D_1, \dots, D_{t-1})$ enriched by K_0 containing prior knowledge.

A. Numerical Representation of Orderings in DM

The addressed DM theory converts DM design, i.e., the selection of the used DM strategy, into an optimisation prob-

This work was supported by GAČR 102/08/0567.

Department of Adaptive Systems, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, P.O.Box 18, 182 08 Prague 8, Czech Republic, {school,guy}@utia.cas.cz

lem that respects the encountered uncertainty and incomplete knowledge. The design aims to select a strategy that induces the best possible behaviour with respect to (*wrt*) a user-specified *preference ordering* $\prec$

$$B_i \prec B_j \Leftrightarrow \text{behaviour } B_i \text{ is strictly preferred against } B_j.$$

This ordering, relating pairs $B_i, B_j \in B^\star$, can be made strict by identifying behaviours felt as equally preferred. Then, the relation $\prec$ is irreflexive, $\neg(B_i \prec B_i)$ and asymmetric, $B_i \prec B_j \Rightarrow \neg(B_j \prec B_i)$. Nontrivial DM arises for non-empty $\prec$. The ordering $\prec$ is assumed to be internally consistent (transitive) $(B_i \prec B_j \wedge B_j \prec B_k) \Rightarrow (B_i \prec B_k)$.

Generally, the user supplying $\prec$ is unable to compare all behaviours, $\prec$ is a *partial* ordering. There is, however, a multitude of complete orderings extending the given partial ordering $\prec$. Further on, such *complete ordering*, denoted also $\prec$, is considered. The preference elicitation discussed in Section III essentially counteracts its non-uniqueness.

For the optimisation-based design, the preference orderings have to be represented numerically. Proposition 1 (see below) delimits a widely acceptable condition under which such a representation exists. It can be found, for instance, in [11] and needs the notion of topology basis.

Let a strict complete ordering $\prec$ be specified on $B^\star \neq \emptyset$. Pairs $B_i, B_j \in B^\star$ define open intervals $(B_i, B_j) \equiv \{B \in B^\star : B_i \prec B \prec B_j\}$. The open-interval topology $\mathcal{B}$ on $B^\star$ is formed by sets $\emptyset$ and $B^\star$ complemented by unions of an arbitrary number of open intervals and by intersections of a finite number of open intervals. A basis of the topology $\mathcal{B}$ is its subset fulfilling: i) any $B \in B^\star$ belongs to a member of the topology $\mathcal{B}$; ii) if $B \in B^\star$ belongs to an intersection of a pair of $\mathcal{B}$-members then there is a topology member contained in the intersection of this pair and containing B.

*Proposition 1 (**Numerical representation of ordering**):* If the set $B^\star \neq \emptyset$ is equipped with a strict complete ordering $\prec$ and there is a countable basis of the topology $\mathcal{B}$ generated by open intervals, then, a *loss* $\mathsf{L} : B^\star \to (-\infty, \infty)$ representing this ordering $\prec$ exists, i.e.,

$$B_i \prec B_j \Rightarrow \mathsf{L}(B_i) < \mathsf{L}(B_j). \tag{1}$$

Proof: See [11] ∎

Loosely, a loss exists if the behaviour set is not topologically richer than the real line. This is supposed further on.

The loss L is non-unique at least due to non-uniqueness of the extension of the user's partial preference ordering.

In the considered DM, the used strategy S determines the behaviour B only partially. To inspect consequences of a specific choice of S, this non-determinism has to be modelled. To determine B uniquely, a mapping W acting on the strategy and an inaccessible *uncertainty* $N \in N^\star \neq \emptyset$

$$\mathsf{W} : (\mathsf{S}, N)^\star \to B^\star \tag{2}$$

has to be introduced. Pairs (W_i, N_i), (W_j, N_j) providing the same behaviour are equivalent. Thus, without a loss of generality, $\{\mathsf{W}(\mathsf{S}, \cdot)\}_{\mathsf{S} \in \mathsf{S}^\star}$ can be assumed to be bijective mappings of $N^\star$ on $B^\star$. The adopted notion of uncertainty respects also non-uniqueness of the extension of the partial preference ordering to the complete one: it simply includes labels of possible extensions into the considered behaviour.

The composition of the mapping W (2) and the loss L (1) generates the set $\mathsf{Z}^\star$ (3) of functions mapping uncertainty $N^\star$ on the real line. They are "indexed" by strategies $\mathsf{S} \in \mathsf{S}^\star$.

$$\begin{aligned} \mathsf{Z}^\star \quad \equiv \quad & \{\mathsf{Z}_\mathsf{S} : N^\star \to (-\infty, \infty) \ \exists \mathsf{S} \in \mathsf{S}^\star \text{ such that} \\ & \mathsf{Z}_\mathsf{S}(N) = \mathsf{L}(\mathsf{W}(\mathsf{S}, N)), \ \forall N \in N^\star\}. \end{aligned} \tag{3}$$

DM has to select and apply a strategy. Thus, it introduces, possibly implicitly, a complete ordering $\prec$ on the set of compared strategies $\mathsf{S}_\star \neq \emptyset$. The complete ordering of strategies introduces the complete ordering of $\mathsf{Z} \in \mathsf{Z}^\star$ (3)

$$\mathsf{Z}_{\mathsf{S}_i} \prec \mathsf{Z}_{\mathsf{S}_j} \Leftrightarrow \mathsf{S}_i \prec \mathsf{S}_j. \tag{4}$$

Assuming that conditions of Proposition 1 are met if Z_S replaces B, the equivalence (4) implies the existence of the functional $\mathsf{T} : \mathsf{Z}^\star \to (-\infty, \infty)$ such that

$$\mathsf{S}_i \prec \mathsf{S}_j \Rightarrow \mathsf{T}(\mathsf{Z}_{\mathsf{S}_i}) < \mathsf{T}(\mathsf{Z}_{\mathsf{S}_j}). \tag{5}$$

Assuming additivity of T on functions $\mathsf{Z}_\mathsf{S} \in Z^\star$ with non-overlapping supports and a range of technical, practically non-restrictive, assumptions (see [12], p. 479, Theorem 5), the following representation of the functional T exists

$$\mathsf{T}(\mathsf{Z}_\mathsf{S}) = \int \mathsf{U}(\mathsf{Z}_\mathsf{S}(N), N)\, \mu(\mathrm{d}N), \tag{6}$$

where an appropriately measurable *utility* U is zero when the first argument is zero and μ is a probabilistic Borel measure.

The numerical representation of ordering treats the uncertainty, and due to (2), the behaviour as random, which is the core of the Bayesian DM. Further on, the same identifier points to a random variable, its value and *realisation*. A verbal description is used if needed. Also, the rnds having time-dependent arguments are generally time-dependent functions.

On $\mathsf{Z}^\star$ (3) a partial ordering exists that induces the strict partial dominance ordering $\prec_d$ of strategies $\mathsf{S} \in \mathsf{S}^\star$

$$\begin{aligned} \mathsf{S}_i \prec_d \mathsf{S}_j \quad \Leftrightarrow \quad & \mathsf{Z}_{\mathsf{S}_i}(N) \leq \mathsf{Z}_{\mathsf{S}_j}(N), \ \forall N \in N^\star \\ \wedge \quad & \mathsf{Z}_{\mathsf{S}_i}(N) < \mathsf{Z}_{\mathsf{S}_j}(N) \ \text{ on } \ N_\star \subset N^\star, \end{aligned} \tag{7}$$

where $N_\star$ has positive measure μ, cf. (6).

Any meaningful ordering of strategies $\prec$ (4) has to be an extension of the dominance ordering $\prec_d$ in (7). Otherwise, the strategy S_j in (7) could be selected as the optimal one, which is obviously the bad choice as the strategy S_i leads to a smaller loss irrespectively of the uncertainty N.

Obviously, the dominance is avoided iff utility U is increasing in its first argument. This is considered further on.

Fundamental theorem of probability ([12], p. 155, Corollary 7) used for a backward substitution corresponding to (2), (3), expresses the strategy-dependent functional (6) as the functional on strategy-dependent functions of behaviour

$$\mathsf{T}_\mathsf{S}(\mathsf{L}) \equiv \mathsf{T}(\mathsf{Z}_\mathsf{S}) = \int_{B^\star} \mathsf{U}\left(\mathsf{L}(B), \mathsf{W}^{-1}(\mathsf{S}, B)\right) \mu_\mathsf{S}(\mathrm{d}B). \tag{8}$$

The resulting measure $\mu_\mathsf{S}(\mathrm{d}B)$ on σ-algebra of Borel sets in $B^\star$ describes the closed loop with the strategy S.

By construction, the most preferred strategy minimises the functional (8) within the set of compared strategies $\mathsf{S}_\star$.

The presence of the second argument of the utility U in (8) generalises the standard Bayesian DM [13]. It allows scaling of the loss value by the utility in dependence on the uncertainty. Moreover, it leads to FPD, see Section III-A.

Below, the utility U and the loss L are mostly discussed together. This allows a simplified notation by introducing and using the *performance index* I

$$\mathsf{I}\left(B, \mathsf{W}^{-1}(\mathsf{S}, B)\right) \equiv \mathsf{U}\left(\mathsf{L}(B), \mathsf{W}^{-1}(\mathsf{S}, B)\right). \tag{9}$$

The optimal (the most preferred) strategy ${}^{\mathsf{IS}_\star}\mathsf{S}$, for a performance index I (9) and compared strategies in $\mathsf{S}_\star$, is defined

$$\begin{aligned} {}^{\mathsf{IS}_\star}\mathsf{S} &\in \operatorname{Arg}\min_{\mathsf{S}\in\mathsf{S}_\star} \mathsf{T}_\mathsf{S}(\mathsf{L}) \\ &= \operatorname{Arg}\min_{\mathsf{S}\in\mathsf{S}_\star} \int_{B^\star} \mathsf{I}\left(B, \mathsf{W}^{-1}(\mathsf{S}, B)\right)\, \mu_\mathsf{S}(\mathrm{d}B). \end{aligned} \tag{10}$$

B. Structuring of DM

The measures $\{\mu_\mathsf{S}\}_{\mathsf{S}\in\mathsf{S}_\star}$ in (10) are assumed to be absolutely continuous wrt a probabilistic measure ν, i.e., for any measurable subset $B_\star \subset B^\star$ with $\nu(B_\star) = 0$ also $\mu_\mathsf{S}(B_\star) = 0$, $\forall \mathsf{S} \in \mathsf{S}_\star$. Then, there exist *Radon-Nikodým derivatives* (rnd) $\mathsf{f}_\mathsf{S}(B)$ [12] such that

$$\begin{aligned} &\mu_\mathsf{S}(\mathrm{d}B) = \mathsf{f}_\mathsf{S}(B)\nu(\mathrm{d}B) \text{ with } \nu\text{-probability 1} \\ &\mathsf{f}_\mathsf{S}(B) \geq 0,\ \int_{B^\star} \mathsf{f}_\mathsf{S}(B)\, \nu(\mathrm{d}B) = 1. \end{aligned} \tag{11}$$

The rnd $\mathsf{f}_\mathsf{S}(B)$ can be interpreted as the *closed-loop model*.

Definitions from the beginning of Section II imply that any admissible strategy $\mathsf{S} \in \mathsf{S}^\star$ is formed by the decision rules S_t, mapping knowledge K_t, gradually enriched by observations Δ_t, on actions A_t. Formally,

$$\begin{aligned} &\mathsf{S} \in \mathsf{S}^\star \equiv \left\{(\mathsf{S}_t : K_t^\star \to A_t^\star)_{t\in t^\star}\right\} \\ &K_t = (K_0, D_1, \ldots, D_{t-1}),\ D_t = (A_t, \Delta_t),\ K = K_{h+1}. \end{aligned} \tag{12}$$

Generally, the behaviour B embraces also *internals*, i.e., variables thought of but never observed. In this text, the considered internals $\Theta \in \Theta^\star$ have character of parameters, i.e., internals uninfluenced by the applied actions. Thus,

$$B = (K, \Theta) = \text{(knowledge, internals)}. \tag{13}$$

Assuming that the dominating measure ν is of product form on B constituents, the chain rule for the rnd $\mathsf{f}_\mathsf{S}(B)$ provides the following factorisation of this closed-loop model

$$\begin{aligned} \mathsf{f}_\mathsf{S}(B) &= \mathsf{f}_\mathsf{S}(\Theta) \prod_{t\in t^\star} \mathsf{f}_\mathsf{S}(\Delta_t|A_t, K_t, \Theta) \times \prod_{t\in t^\star} \mathsf{f}_\mathsf{S}(A_t|K_t) \\ &\equiv \mathsf{M}(B) \times \mathsf{S}(B). \end{aligned} \tag{14}$$

The last product expresses the adopted assumption that the *environment model* $\mathsf{M}(B) \equiv \mathsf{f}(\Theta)\prod_{t\in t^\star} \mathsf{f}(\Delta_t|A_t, K_t, \Theta)$ is common to all compared strategies $\mathsf{S} \in \mathsf{S}_\star$. This allows the dropping of the strategy-identifying subscript S at its factors.

The omission of the internals Θ in the strategy model $\prod_{t\in t^\star} \mathsf{f}_\mathsf{S}(A_t|K_t)$ formalises the notion "unknown": the values of the internals $\Theta \in \Theta^\star$ cannot be used by an admissible strategy. This assumption is known as natural conditions of control [14].

With the environment model M fixed, the strategies $\mathsf{S}_i, \mathsf{S}_j \in \mathsf{S}^\star$ with identical models of their DM rules $\mathsf{f}_{\mathsf{S}_i}(A_t|K_t) = \mathsf{f}_{\mathsf{S}_j}(A_t|K_t)$, $t \in t^\star$, provide the same closed-loop model $\mathsf{f}_\mathsf{S}(B)$. Consequently, the DM strategy S can be identified with its model $\mathsf{S}(B) \equiv \prod_{t\in t^\star} \mathsf{f}(A_t|K_t)$. Similar identification applies to decision rules, $\mathsf{S}_t \equiv \mathsf{f}(A_t|K_t)$.

In the environment model M, the rnd $\mathsf{f}(\Theta)$ describes prior knowledge on internals Θ. The unknown Θ may origin from any part of the modelled closed decision loop. Mostly, only a part ${}^{\mathsf{M}}\Theta$, called *parameters*, explicitly enters the *parametric environment model*

$$\mathsf{f}(\Delta_t|A_t, K_t, \Theta) = \mathsf{f}\left(\Delta_t|A_t, K_t, {}^{\mathsf{M}}\Theta\right). \tag{15}$$

The "correlated" rest ${}^{\mathsf{I}}\Theta$ of Θ then characterises the complete behaviour ordering quantified by the performance index I

$$\Theta = \left({}^{\mathsf{M}}\Theta,\ {}^{\mathsf{I}}\Theta\right) = \text{(model,performance index) parameters}. \tag{16}$$

III. FPD AND PREFERENCE ELICITATION

Section III-A gives conditions under which the design becomes fully probabilistic. Section III-B relates FPD to the standard Bayesian design. Section III-C solves a novel variant of FPD supporting data-dependent preference elicitation.

A. Fully Probabilistic Design

If realisations $B_i,\ B_j \in B^\star$ lead to the same loss $\mathsf{L}(B_i) = \mathsf{L}(B_j)$ and have the same probability $\mathsf{f}_\mathsf{S}(B_i)\nu(\mathrm{d}B_i) = \mathsf{f}_\mathsf{S}(B_j)\nu(\mathrm{d}B_j)$ then they are equivalent for the DM task. This motivates the assumption, see (9),

$$\mathsf{I}(B, \mathsf{W}^{-1}(\mathsf{S}, B)) = \mathsf{I}(B, \mathsf{f}_\mathsf{S}(B)), \tag{17}$$

which says that the behaviour enters the second argument of the performance index only via the closed-loop model $\mathsf{f}_\mathsf{S}(B)$. Under (17) and (14), the optimal strategy ${}^{\mathsf{IS}_\star}\mathsf{S}$ (10) is defined

$${}^{\mathsf{IS}_\star}\mathsf{S} \in \operatorname{Arg}\min_{\mathsf{S}\in\mathsf{S}_\star} \int_{B^\star} \mathsf{I}\left(B, \mathsf{M}(B)\mathsf{S}(B)\right) \mathsf{M}(B)\mathsf{S}(B)\, \nu(\mathrm{d}B). \tag{18}$$

The closed-loop model f_S (14) with the optimal strategy ${}^{\mathsf{IS}_\star}\mathsf{S}$ (18) gives an *ideal closed-loop model*

$${}^{\mathsf{IS}_\star}\mathsf{f} \equiv \mathsf{M}\, {}^{\mathsf{IS}_\star}\mathsf{S}, \tag{19}$$

which is an image of the user-specified performance index I and of the set of compared strategies $\mathsf{S}_\star$. The FPD consider the DM design in which the user specifies an ideal (desired) closed-loop model ${}^{IS_\star}\mathsf{f}$ instead of an performance index.

There is whole set $\{\mathsf{I}\}$ of performance indices leading via (18), (19) to the given ideal closed-loop model ${}^{\mathsf{IS}_\star}\mathsf{f}$. They are equivalent for the DM design. The following exposition founds a performance index ${}^{R}\mathsf{I}$ representing them.

By its definition, the performance index ${}^{R}\mathsf{I}$, when used in (18) in the role of I, provides the optimal strategy with which the closed-loop model (19) coincides with the given ideal closed-loop model. Moreover, ${}^{R}\mathsf{I}$ is required to fulfill

$${}^{R}\mathsf{I}\left(B,\ {}^{\mathsf{IS}_\star}\mathsf{f}(B)\right) = {}^{S_\star}constant,\ \forall B \in B^\star, \tag{20}$$

i.e., the value of this performance index with the optimally tuned closed loop is required to be independent of the behaviour realisation. This requirement stresses that the choice of the optimal strategy is made a priori without knowing a specific behaviour realisation.

The assumption (20) inserted (18) with $\mathsf{I} = {}^{R}\mathsf{I}$ implies

$$ {}^{R}\mathsf{I}\left(B,\, {}^{\mathsf{IS}_\star}\mathsf{f}(B)\right) \geq {}^{R}\mathsf{I}\left(B,\, {}^{\mathsf{IS}^\star}\mathsf{f}(B)\right) \text{ on } B^\star.$$

This guarantees that the realisation of the performance index does not increase due to an extension of the set of the compared strategies. It means that the intuitively favourable extension of the set of compared strategies has a positive influence on realisations of the performance index.

Onwards, the explicit reference to $\mathsf{S}_\star$ can be dropped.

*Proposition 2 (**DM as FPD**):* Let the performance index ${}^{R}\mathsf{I}(B, \mathsf{M}(B)\mathsf{S}(B))$ represent all performance indices that lead, via (18), (19), to a given ideal closed-loop model ${}^{\mathsf{I}}\mathsf{f} = {}^{\mathsf{IS}_\star}\mathsf{f}$. If the performance index ${}^{R}\mathsf{I}$ has a finite first derivative wrt to the second argument and meets (20), then,

$$ {}^{R}\mathsf{I}\left(B, \mathsf{f}_\mathsf{S}(B)\right) = \ln\left(\frac{\mathsf{f}_\mathsf{S}(B)}{{}^{\mathsf{I}}\mathsf{f}(B)}\right). \tag{21}$$

With $\mathsf{I} = {}^{R}\mathsf{I}$, the optimised functional (18) becomes Kullback-Leibler Divergence (KLD) $\mathsf{D}\left(\mathsf{f}_\mathsf{S}||\,{}^{\mathsf{I}}\mathsf{f}\right)$ of the closed-loop model $\mathsf{f}_\mathsf{S} = \mathsf{MS}$ on the ideal closed-loop model ${}^{\mathsf{I}}\mathsf{f}$

$$\int_{B^\star} \mathsf{M}(B)\mathsf{S}(B)\left(\frac{\mathsf{M}(B)\mathsf{S}(B)}{{}^{\mathsf{I}}\mathsf{f}(B)}\right)\nu(\mathrm{d}B) \equiv \mathsf{D}\left(\mathsf{f}_\mathsf{S}||\,{}^{\mathsf{I}}\mathsf{f}\right).$$

KLD of a pair of rnds g, f on $B^\star$, fulfils, [15],

$$\mathsf{D}\left(\mathsf{g}||\mathsf{f}\right) \geq 0,\ \mathsf{D}\left(\mathsf{g}||\mathsf{f}\right) = 0 \text{ iff } \mathsf{g} = \mathsf{f} \text{ with } \nu\text{-probability 1}$$
$$\mathsf{D}\left(\mathsf{g}||\mathsf{f}\right) = \infty \text{ iff } \mathsf{f} \text{ is not absolutely continuous wrt } \mathsf{g}. \tag{22}$$

Proof: The functional (18) with $\mathsf{I} = {}^{R}\mathsf{I}$ has to reach minimum for the given ideal closed-loop model ${}^{\mathsf{I}}\mathsf{f}$. A weak zero variation of (18) with $\mathsf{I} = {}^{R}\mathsf{I}$ at ${}^{\mathsf{I}}\mathsf{f}$ provides the necessary conditions that have to hold with ν-probability 1 on $B^\star$

$$x\frac{\partial\, {}^{R}\mathsf{I}\left(B,x\right)}{\partial x} + {}^{R}\mathsf{I}\left(B,x\right) = constant \text{ for } x = {}^{\mathsf{I}}\mathsf{f}(B).$$

Under (20), they are fulfilled by (21), which meets all requirements on ${}^{R}\mathsf{I}$. ∎

The DM design that uses the performance index ${}^{R}\mathsf{I}$ characterised by Proposition 2 is called fully probabilistic design. *FPD selects the optimal DM strategy* ${}^{O}\mathsf{S} \equiv {}^{R\mathsf{I}}\mathsf{S}$ *as the minimiser of KLD of the closed-loop model* $\mathsf{f}_\mathsf{S} = \mathsf{MS}$ *on a given ideal closed-loop model* ${}^{\mathsf{I}}\mathsf{f}$.

B. Relation of Standard Bayesian DM to FPD

If the performance index $\mathsf{I}(B, \mathsf{f}_\mathsf{S}(B))$ (17) is independent of the second argument, the functional (18) becomes linear in the optimised strategy S. This section inspects relation of this standard (textbook) Bayesian design to FPD.

FPD specifies the design aims via the ideal closed loop model. At the same time, it is known [2] that the standard Bayesian design, optimising expectation of a performance index $\mathsf{I}(B)$, provides the optimal strategy formed by deterministic DM rules. They are formally described by

$$\mathsf{f}(A_t|K_t) = \delta\left(A - {}^{O}A_t(K_t)\right), t \in t^\star, \tag{23}$$

where δ concentrates the full probability mass on the optimal actions ${}^{O}A_t(K_t)$. Thus, the corresponding ideal closed-loop model (19) is singular and fully concentrated on actions $({}^{O}A_t(K_t))_{t\in t^\star}$, which are unknown when the DM problem is formulated.

The above discrepancy can be resolved by taking into account that such singular ideal closed-loop model can be arbitrarily closely approximated by employing DM rules that are positive on $(A_t, K_t)^\star$.

Indeed, for a continuous-valued action the approximating rnds can be taken as the normal rnd $\mathcal{N}_{A_t}({}^{O}A_t(K_t), \mathbf{R})$ with an expected value ${}^{O}A_t(K_t)$ and covariance $\mathbf{R}$ approaching to zero. For a discrete-valued action, a mixture $\varepsilon\mathcal{U}_{A_t}(A_t^\star) + (1-\varepsilon)\delta(A - {}^{O}A_t(K_t))$, with $\mathcal{U}_{A_t}(A_t^\star)$ being uniform rnd of A_t on $A_t^\star$ and a positive ε approaching zero, serves to this purpose. These cases and their combinations can be covered by requiring

$$\begin{aligned}\int_{B^\star} \mathsf{M}(B)\mathsf{S}(B)\ln(\mathsf{S}(B))\nu(\mathrm{d}B) &= \text{finite value}\\ < \sup_{\mathsf{S}\in S_\star}\int_{B^\star} \mathsf{M}(B)\mathsf{S}(B)\ln(\mathsf{S}(B))\nu(\mathrm{d}B).\end{aligned} \tag{24}$$

Let us consider a set of DM tasks in which the value on the right-hand side of (24) is an optional part of the DM design. Obviously, the larger this value is the less restrictive is this constraint.

The optimal strategy ${}^{\mathsf{I}}\mathsf{S}$ (18) on a subset $\mathsf{S}_\star \subset \mathsf{S}^\star$ of strategies meeting (24) results from minimisation of the functional, given by a multiplier $\lambda > 0$,

$$\begin{aligned}{}^{\mathsf{I}}\mathsf{S} &\in \mathrm{Arg}\min_{\mathsf{S}\in\mathsf{S}_\star}\int_{B^\star}\mathsf{M}(B)\mathsf{S}(B)\left[\mathsf{I}(B) + \lambda\ln(\mathsf{S}(B))\right]\nu(\mathrm{d}B)\\ &= \mathrm{Arg}\min_{\mathsf{S}\in\mathsf{S}_\star}\int_{B^\star}\mathsf{M}(B)\mathsf{S}(B)\ln\left(\frac{\mathsf{M}(B)\mathsf{S}(B)}{\mathsf{M}(B)\exp\left(-\frac{\mathsf{I}(B)}{\lambda}\right)}\right)\nu(\mathrm{d}B)\\ &= \mathrm{Arg}\min_{\mathsf{S}\in\mathsf{S}_\star}\mathsf{D}\left(\mathsf{f}_\mathsf{S}||\,{}^{\mathsf{I}}\mathsf{f}\right) \text{ with the ideal closed-loop model}\\ {}^{\mathsf{I}}\mathsf{f}(B) &\equiv \frac{\mathsf{M}(B)\exp\left(-\frac{\mathsf{I}(B)}{\lambda}\right)}{\int_{B^\star}\mathsf{M}(B)\exp\left(-\frac{\mathsf{I}(B)}{\lambda}\right)\nu(\mathrm{d}B)}.\end{aligned} \tag{25}$$

The explicit formula (25) relates any performance index $\mathsf{I}(B)$, determining the standard Bayesian design, with the corresponding ideal closed-loop model ${}^{\mathsf{I}}\mathsf{f}$. The value of the multiplier λ is to respect constraint (24). When this constraint is asymptotically relaxed, i.e., $\lambda \to 0$, the optimal strategy with deterministic DM rules is recovered. The construction is closely related to simulated annealing techniques like Boltzmann machine [16].

This implies that DM tasks formulated in terms of FPD are dense in the set of the standard Bayesian DM tasks.

C. Incomplete Knowledge of the Ideal Closed-Loop Model

The discussion of this section needs a modified version of the general FPD presented in [7]. Its presentation also shows that the minimisation in FPD can be made explicitly.

For the decision horizon h, the closed-loop model (14) and behaviours (13) $\mathsf{f}_h(B) \equiv \mathsf{f}_{\mathsf{S}}(B) = \mathsf{f}_{\mathsf{S}}(K, \Theta)$ reads

$$\mathsf{f}_h(B) = \underbrace{\mathsf{f}(\Theta) \prod_{t \in t^\star} \mathsf{f}(\Delta_t | A_t, K_t, \Theta)}_{\text{environment model}} \underbrace{\prod_{t \in t^\star} \mathsf{f}(A_t | K_t)}_{\text{strategy}}. \tag{26}$$

Similarly, the ideal closed-loop model ${}^{\mathsf{I}}\mathsf{f}_h(B) \equiv {}^{\mathsf{I}}\mathsf{f}(B)$ is

$${}^{\mathsf{I}}\mathsf{f}_h(B) = \overbrace{{}^{\mathsf{I}}\mathsf{f}(\Theta) \prod_{t \in t^\star} {}^{\mathsf{I}}\mathsf{f}(\Delta_t | A_t, K_t, \Theta)}^{\text{ideal environment model}} \overbrace{\prod_{t \in t^\star} {}^{\mathsf{I}}\mathsf{f}(A_t | K_t, \Theta)}^{\text{ideal strategy}}. \tag{27}$$

In (27), the DM rules ${}^{\mathsf{I}}\mathsf{f}(A_t|K_t, \Theta)$ of the ideal strategy generally depend on the unknown internals $\Theta = ({}^{\mathsf{M}}\Theta, {}^{\mathsf{I}}\Theta) \in \Theta^\star$, see (16). This dependence is needed, for instance, when A_t is an estimate of ${}^{\mathsf{M}}\Theta$.

The FPD solution is summarised below under the adopted assumption that the measure ν is of a product form.

*Proposition 3 (**Solution of FPD**):* The DM strategy meeting the natural conditions of control $\left\{(\mathsf{f}(A_t|K_t, \Theta) = \mathsf{f}(A_t|K_t))_{t \in t^\star}\right\}$ and minimising KLD $\mathsf{D}\left(\mathsf{f}_h || {}^{\mathsf{I}}\mathsf{f}_h\right)$ (26), (27) is described by the following decision rules, $t \in t^\star$,

$$\begin{aligned}
{}^{O}\mathsf{f}(A_t|K_t) &= {}^{\mathsf{I}}\mathsf{f}(A_t|K_t) \frac{\exp[-\omega(A_t, K_t)]}{\gamma(K_t)} && (28)\\
\gamma(K_t) &= \int_{A_t^\star} {}^{\mathsf{I}}\mathsf{f}(A_t|K_t) \exp[-\omega(A_t, K_t)] \nu(\mathrm{d}A_t) \\
{}^{\mathsf{I}}\mathsf{f}(A_t|K_t) &\equiv && (29)\\
&\underbrace{\frac{\exp\left[\int_{\Theta^\star} \ln({}^{\mathsf{I}}\mathsf{f}(A_t|K_t, \Theta)) \mathsf{f}(\Theta|K_t) \nu(\mathrm{d}\Theta)\right]}{\int_{A_t^\star} \exp\left[\int_{\Theta^\star} \ln({}^{\mathsf{I}}\mathsf{f}(A_t|K_t, \Theta)) \mathsf{f}(\Theta|K_t) \nu(\mathrm{d}\Theta)\right] \nu(\mathrm{d}A_t)}}_{\phi(K_t)}.
\end{aligned}$$

Starting with $\gamma(K_{h+1}) \equiv 1$, the functions $\omega(A_t, K_t)$ are generated in the backward manner for $t = h, h-1, \ldots, 1$

$$\begin{aligned}
&\omega(A_t, K_t) \equiv \ln(\phi(K_t)) + \int_{\Theta^\star} \Omega(A_t, K_t, \Theta) \mathsf{f}(\Theta|K_t) \nu(\mathrm{d}\Theta) \\
&\Omega(A_t, K_t, \Theta) \equiv && (30)\\
&\int_{\Delta_t^\star} \mathsf{f}(\Delta_t|A_t, K_t, \Theta) \ln\left(\frac{\mathsf{f}(\Delta_t|A_t, K_t, \Theta)}{\gamma(K_{t+1}) {}^{\mathsf{I}}\mathsf{f}(\Delta_t|A_t, K_t, \Theta)}\right) \nu(\mathrm{d}\Delta_t).
\end{aligned}$$

Parameter estimate (the posterior rnd) $\mathsf{f}(\Theta|K_t)$ evolves recursively, from $\mathsf{f}(\Theta|K_0) = \mathsf{f}\left({}^{\mathsf{M}}\Theta, {}^{\mathsf{I}}\Theta\right)$, see (15), (16),

$$\mathsf{f}(\Theta|K_{t+1}) = \frac{\mathsf{f}\left(\Delta_t|A_t, K_t, {}^{\mathsf{M}}\Theta\right) \mathsf{f}(\Theta|K_t)}{\int_{\Theta^\star} \mathsf{f}(\Delta_t|A_t, K_t, {}^{\mathsf{M}}\Theta) \mathsf{f}(\Theta|K_t) \nu(\mathrm{d}\Theta)}. \tag{31}$$

Proof: The evolution (31) coincides with Bayes rule with the Θ-independent DM rules cancelled [14]. The derivation exploits the basic properties of KLD (22), Fubini theorem on multiple integration [12], the rnd properties (11), (14) and the fact that KLD is an expectation of the sum

$$\sum_{t \in t^\star} \ln\left(\frac{\mathsf{f}(\Delta_t|A_t, K_t, \Theta) \mathsf{f}(A_t|K_t)}{{}^{\mathsf{I}}\mathsf{f}(\Delta_t|A_t, K_t, \Theta) {}^{\mathsf{I}}\mathsf{f}(A_t|K_t, \Theta)}\right) + \ln\left(\frac{\mathsf{f}(\Theta)}{{}^{\mathsf{I}}\mathsf{f}}\right).$$

The last fact and the definition $\gamma(K_{h+1}) = 1$ imply

$$\begin{aligned}
&\min_{\{\mathsf{f}(A_t|K_t)\}_{t \in t^\star}} \mathsf{D}\left(\mathsf{f}_h || {}^{\mathsf{I}}\mathsf{f}_h\right) = \min_{\{\mathsf{f}(A_t|K_t)\}_{t=1}^{h-1}} \left\{ \mathsf{D}\left(\mathsf{f}_{h-1} || {}^{\mathsf{I}}\mathsf{f}_{h-1}\right) \right. \\
&+ \min_{\{\mathsf{f}(A_h|K_h)\}} \int_{(K_h, \Theta)^\star} \mathsf{f}_{h-1}(B) \nu(\mathrm{d}(K_h, \Theta)) \\
&\times \left[\int_{(\Delta_h, A_h)^\star} \nu(\mathrm{d}(\Delta_h, A_h)) \mathsf{f}(\Delta_h|A_h, K_h, \Theta) \mathsf{f}(A_h|K_h) \right. \\
&\left. \left. \times \ln\left(\frac{\mathsf{f}(\Delta_h|A_h, K_h, \Theta) \mathsf{f}(A_h|K_h)}{\gamma(K_{h+1}) {}^{\mathsf{I}}\mathsf{f}(\Delta_h|A_h, K_h, \Theta) {}^{\mathsf{I}}\mathsf{f}(A_h|K_h, \Theta)}\right) \right] \right\} && (32)
\end{aligned}$$

The second term in (32) is optimised over the last DM rule $\mathsf{f}(A_h|K_h)$ of the admissible strategy and the expression in its square brackets can be rearranged as follows

$$\begin{aligned}
&\maltese \equiv \int_{(\Delta_h, A_h)^\star} \mathsf{f}(\Delta_h|A_h, K_h, \Theta) \mathsf{f}(A_h|K_h) \\
&\times \ln\left(\frac{\mathsf{f}(\Delta_h|A_h, K_h, \Theta) \mathsf{f}(A_h|K_h)}{\gamma(K_{h+1}) {}^{\mathsf{I}}\mathsf{f}(\Delta_h|A_h, K_h, \Theta) {}^{\mathsf{I}}\mathsf{f}(A_h|K_h, \Theta)}\right) \nu(\mathrm{d}(\Delta_h, A_h)) \\
&= \int_{A_h^\star} \mathsf{f}(A_h|K_h) \nu(\mathrm{d}A_h) \left[\ln\left(\frac{\mathsf{f}(A_h|K_h)}{{}^{\mathsf{I}}\mathsf{f}(A_h|K_h, \Theta)}\right) + \right. \\
&\underbrace{\left. \int_{\Delta_h^\star} \mathsf{f}(\Delta_h|A_h, K_h, \Theta) \ln\left(\frac{\mathsf{f}(\Delta_h|A_h, K_h, \Theta)}{\gamma(K_{h+1}) {}^{\mathsf{I}}\mathsf{f}(\Delta_h|A_h, K_h, \Theta)}\right) \nu(\mathrm{d}\Delta_h) \right]}_{\Omega(A_h, K_h, \Theta)}.
\end{aligned}$$

Using definitions (29) of ${}^{\mathsf{I}}\mathsf{f}(A_h|K_h))$ and (30) of $\Omega(A_h, K_h, \Theta)$, the inspected second term in (32) becomes

$$\begin{aligned}
&\min_{\{\mathsf{f}(A_h|K_h)\}} \int_{K_h^\star} \mathsf{f}_{h-1}(B) \maltese \nu(\mathrm{d}K_h) && (33)\\
&= \int_{K_h^\star} \nu(\mathrm{d}K_h) \mathsf{f}(K_h) \left\{ \int_{A_h^\star} \mathsf{f}(A_h|K_h) \left[\ln\left(\frac{\mathsf{f}(A_h|K_h)}{{}^{\mathsf{I}}\mathsf{f}(A_h|K_h)}\right) \right. \right. \\
&\left. \left. + \underbrace{\ln(\phi(K_h)) + \int_{\Theta^\star} \mathsf{f}(\Theta|K_h) \Omega(A_h, K_h, \Theta) \nu(\mathrm{d}\Theta)}_{\omega(A_h, K_h)} \right] \nu(\mathrm{d}A_h) \right\}.
\end{aligned}$$

The function $\omega(A_h, K_h)$ defined in (33) uses the estimate $\mathsf{f}(\Theta|K_h)$ in integral term. Due to the natural conditions of control, the estimate $\mathsf{f}(\Theta|K_h)$ does not depend on the optimised DM rule $\mathsf{f}(A_h|K_h)$. The DM rule enters only the functional in the compound brackets in (33) as follows

$$\begin{aligned}
&\int_{A_h^\star} \mathsf{f}(A_h|K_h) \ln\left(\frac{\mathsf{f}(A_h|K_h)}{\frac{{}^{\mathsf{I}}\mathsf{f}(A_h|K_h) \exp[-\omega(A_h, K_h)]}{\gamma(K_h)}}\right) \nu(\mathrm{d}A_h) \\
&\underbrace{- \ln\left(\int_{A_h^\star} {}^{\mathsf{I}}\mathsf{f}(A_h|K_h) \exp[-\omega(A_h, K_h)] \nu(\mathrm{d}A_h)\right)}_{-\ln(\gamma(K_h))}. && (34)
\end{aligned}$$

Addition and subtraction of $\ln(\gamma(K_h))$ made the first term in (34) equal to the conditional version of KLD. The basic properties of KLD (22) imply that minimum $-\ln(\gamma(K_h))$ is reached for the rnd (28). Insertion of this minimiser into (32) shows that $-\ln(\gamma(K_h))$ enters $\mathsf{D}(\mathsf{f}_{h-1} || {}^{\mathsf{I}}\mathsf{f}_{h-1})$ exactly

in the same way as $-\ln(\gamma(K_{h+1}))$ enters $\mathsf{D}(\mathsf{f}_h||\,{}^{\mathsf{I}}\mathsf{f}_h)$, i.e., the optimisation can be repeated for $h-1, h-2, \ldots, 1$. ∎

Proposition 3, modifying the general FPD [7], suits to a systematic data-based preference elicitation. Formally, it suffices to relate all unknown internals $\Theta = \left({}^{\mathsf{M}}\Theta,\, {}^{\mathsf{I}}\Theta\right)$ (16) to ${}^{\mathsf{M}}\Theta$ parameterising the environment model M, see (15), and learn Θ. A closer look on the Bayes rule (31) reveals that the experience K_t accumulated via the Bayes rule influences only the marginal rnd $\mathsf{f}({}^{\mathsf{M}}\Theta|K_t)$. Thus, the joint rnd $\mathsf{f}(\Theta|K_t) = \mathsf{f}\left({}^{\mathsf{M}}\Theta,\, {}^{\mathsf{I}}\Theta|K_t\right) = \mathsf{f}\left({}^{\mathsf{I}}\Theta|\,{}^{\mathsf{M}}\Theta, K_t\right)\mathsf{f}\left({}^{\mathsf{M}}\Theta|K_t\right)$ informs about ${}^{\mathsf{I}}\Theta$ only when the rnd $\mathsf{f}({}^{\mathsf{I}}\Theta|\,{}^{\mathsf{M}}\Theta, K_t)$ relating ${}^{\mathsf{I}}\Theta$ to ${}^{\mathsf{M}}\Theta$ is supplied externally.

D. Regulation Problem with Normal Models

Example presented in this section indicates that the relation between ${}^{\mathsf{I}}\Theta$ and ${}^{\mathsf{M}}\Theta$, whose need for preference elicitation is revealed above, can be constructed.

The paper [5] shown that FPD applied to normal rnds reduces to the standard linear-quadratic control design. In this case, the parametric environment model is assumed to be normal rnd

$$\mathsf{f}\left(\Delta_t|A_t, K_t,\, {}^{\mathsf{M}}\Theta\right) = \mathcal{N}_{\Delta_t}(\mathbf{C}[A_t', \Delta_{t-1}']', \mathbf{R}), \qquad (35)$$

where $'$ transposes column vectors of actions A and observations Δ. The environment model is parameterised by ${}^{\mathsf{M}}\Theta \equiv (\mathbf{C}, \mathbf{R})$, see (15). The unknown rectangular matrix $\mathbf{C}$ weights the action A_t and the past observation Δ_{t-1} and determines the expected value of Δ_t. The positive-definite covariance matrix $\mathbf{R}$ ($\mathbf{R} > 0$) is also unknown.

The considered regulation problem is the DM task that aims to keep the observed Δ_t, $t \in t^\star$, as close as possible to zero while the stationary covariance $\mathbf{R}_A$ of actions A_t should be bounded by a given matrix $\bar{\mathbf{R}}_A > 0$ that provides a soft upper bound on the action range.

For the given ${}^{\mathsf{M}}\Theta$, the optimal strategy, minimising an expected stationary value of the quadratic performance index given by a weighting matrix ${}^{\mathsf{I}}\Theta > 0$,

$$\mathsf{I}(B) = \lim_{h\to\infty} \frac{1}{h} \sum_{t\in t^\star} D_t'\, {}^{\mathsf{I}}\Theta D_t,\;\; D_t' = [A_t', \Delta_t'], \qquad (36)$$

is a linear feedback, see [17] and (16),

$$A_t = -\mathbf{L}'\left({}^{\mathsf{M}}\Theta,\, {}^{\mathsf{I}}\Theta\right)\Delta_{t-1} = -\mathbf{L}'(\Theta)\Delta_{t-1}. \qquad (37)$$

The corresponding ideal closed-loop model is normal rnd ${}^{\mathsf{I}}\mathsf{f}(D_t|K_{t-1}, \Theta) = \mathcal{N}_{D_t}\left(0, \lambda^{-1}\mathbf{P}\left({}^{\mathsf{M}}\Theta,\, {}^{\mathsf{I}}\Theta\right)\right) = \mathcal{N}_{D_t}\left(0, \lambda^{-1}\mathbf{P}(\Theta)\right)$.

The covariance matrix

$$\mathbf{P}(\Theta) = \begin{bmatrix} \mathbf{P}_A(\Theta) & \bullet \\ \bullet & \mathbf{P}_\Delta(\Theta) \end{bmatrix}$$

is found by solving the related stationary Riccati equation [5]. The scalar multiplier $\lambda > 0$ is optional, see (25).

The control preferences are followed the most tightly if the $\mathbf{P}_\Delta(\Theta)$ has the smallest trace while $\mathbf{P}_A(\Theta)$ is kept smaller than $\bar{\mathbf{R}}_A$. This defines the mapping relating ${}^{\mathsf{M}}\Theta$ to ${}^{\mathsf{I}}\Theta$

$${}^{\mathsf{M}}\Theta \to {}^{\mathsf{I}}\Theta \in \mathrm{Arg} \min_{\mathbf{Q}>0,\; \bar{\mathbf{R}}_A - \mathbf{P}_A({}^{\mathsf{M}}\Theta, \mathbf{Q}) \geq 0} \mathrm{tr}\left(\mathbf{P}_\Delta\left({}^{\mathsf{M}}\Theta, \mathbf{Q}\right)\right).$$

This mapping is conjectured to be well defined as the lowest potentially reachable stationary covariance of Δ_t coincides with the covariance of the parametric environment model $\mathbf{R}$. Its formal validation as well as numerical construction of this mapping is out of the scope of the paper. Numerical Monte Carlo evaluation was already found feasible in a related context [18].

IV. CONCLUSIONS

The paper deepens axiomatic basis of fully probabilistic design (FDP). It explicitly relates the performance index used in the standard Bayesian design to the Radon-Nikodým derivatives describing the ideal closed-loop model, which determines the performance index in FPD. The paper proves that FPD tasks are dense with respect to the set of standard Bayesian DM tasks. The support built for FPD covers the standard Bayesian DM, too. Importantly, the presented results open a new way to data-based preference elicitation.

REFERENCES

[1] A. Wald, *Statistical Decision Functions*, John Wiley, New York, London, 1950.

[2] M.H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.

[3] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, Nashua, US, 2001, 2nd edition.

[4] M. Kárný, "Bayesian paradigm and fully probabilistic design", in *Preprints of the 17th IFAC World Congress*. 2008, IFAC.

[5] M. Kárný, "Towards fully probabilistic control design", *Automatica*, vol. 32, no. 12, pp. 1719–1722, 1996.

[6] J. Šindelář, I. Vajda, and M. Kárný, "Stochastic control optimal in the Kullback sense", *Kybernetika*, vol. 44, no. 1, pp. 53–60, 2008.

[7] M. Kárný and T. V. Guy, "Fully probabilistic control design", *Systems & Control Letters*, vol. 55, no. 4, pp. 259–265, 2006.

[8] M. Kárný and J. Kracík, "A normative probabilistic design of a fair governmental decision strategy", *Journal of Multi-Criteria Decision Analysis*, vol. 12, no. 2-3, pp. 1–15, 2004.

[9] J. Böhm, T. V. Guy, and M. Kárný, "Multiobjective probabilistic mixture control", in *Preprints of the 16th IFAC World Congress*, P. Horáček, M. Šimandl, and P. Zítek, Eds., Prague, 2005, IFAC.

[10] P.H. Garthwaite, J.B. Kadane, and A. O´Hagan, "Statistical methods for eliciting probability distributions", *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 680–700, Jun 2005.

[11] P.C. Fishburn, *Utility Theory for Decision Making*, J. Wiley, New York, London, Sydney, Toronto, 1970.

[12] M.M. Rao, *Measure Theory and Integration*, John Wiley, New York, 1987.

[13] L.J. Savage, *Foundations of Statistics*, Wiley, New York, 1954.

[14] V. Peterka, "Bayesian system identification", in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.

[15] S. Kullback and R. Leibler, "On information and sufficiency", *Annals of Mathematical Statistics*, vol. 22, pp. 79–87, 1951.

[16] J. Si, A.G. Barto, W.B. Powell, and D. Wunsch, Eds., *Handbook of Learning and Approximate Dynamic Programming*, Danvers, May 2004. Wiley-IEEE Press.

[17] B.D.O. Anderson and J.B. Moore, *Optimal Control : Linear Quadratic Methods*, Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

[18] M. Kárný and A. Halousková, "Pretuning of self-tuners", in *Advances in Model-Based Predictive Control*, D. Clarke, Ed., pp. 333–343. Oxford University Press, Oxford, 1994.