Fully Probabilistic Design Unifies and Supports Dynamic Decision Making Under Uncertainty

Miroslav Kárný

[a] The Czech Academy of Sciences, Institute of Information Theory and Automation, POB 18, 182 08 Prague 8, Czech Republic, school@utia.cas.cz

Abstract

The fully probabilistic design (FPD) of decision strategies models the closed decision loop as well as decision aims and constraints by joint probabilities of involved variables. FPD takes the minimiser of cross entropy (CE) of the closed-loop model to its ideal counterpart, expressing the decision aims and constraints, as the optimal strategy. FPD: a) got an axiomatic basis; b) extended the decision making (DM) optimising a subjective expected utility (SEU); c) was nontrivially applied; d) advocated CE as a proper similarity measure for an approximation of a given probability distribution; d) generalised the minimum CE principle for a choice of the distribution, which respects its incomplete specification; e) has opened a way to the cooperation based on sharing of probability distributions. When trying to survey the listed results, scattered in a range of publications, we have found that the results under b), d) and e) can be refined and non-trivially generalised. This determines the paper aims: to provide a complete concise description of FPD with its use and open problems outlined.

Keywords: dynamic decision making, uncertainty, cross entropy, performance indices

1. Introduction

Dynamic decision making under uncertainty, understood as a targeted choice among available options, covers a substantial portion of human activities and related theoretical domains. This paper describes FPD as its prescriptive theory. The discussion of DM and of a huge amount of available results is minimised.

1.1. Terminology and Notation

The paper adopts terminology close to DM community but mixes it with jargon used in control area where FDP arisen.

A solution of a DM task relies on a complete specification of its inputs called *DM elements*. A real decision maker, referred as an *agent* with the "gender" it, specifies them incompletely. Thus, the targeted algorithmic DM has their mapping on formal DM elements as its indispensable part. The construction of this mapping is a DM task. It is referred as the *supporting DM*. It serves to the "original" *supported DM*. The adjective *universal* stresses that the described DM theory serves to all decision makers dealing with the same behaviour set and facing the same uncertainty. They are free to choose their preferences, constraints and compared strategies.

Small letters denote elements of the supported DM. The supporting DM uses their capital counterparts. A set of z's is $\boldsymbol{z}$, $|\boldsymbol{z}|$ is its cardinality. $\mathbb{R}$ denotes the extended real line. San serif fonts mark $\mathsf{mappings}$, while $\mathcal{C}$aligraphic fonts mean functionals. Superscripts identify specific problem elements. Each probability distribution is assumed to have *probability density* (*pd*). When operating on pds, $\propto$ means proportionality. The conditioning by a prior knowledge is mostly implicit. The proof presentation indicates the novelty of the proved proposition.

1.2. A Guide Through the Presented Theory

This part aims to simplify orientation within the paper.

Sec. 2 summarises the design of the optimal strategy solving DM under uncertainty. It lists the adopted assumptions and generalises DM based on SEU [26]. The assumptions require the existence of complete preferential ordering of possible behaviours of closed DM loop and a complete ordering of compared DM strategies. The demanded existence of their numerical representations and Prop. 1 restrict us to DM on ordered sets with a countable dense topology. Practically, it allows us to derive various results for a finite number of behaviour realisations. The use of Prop. 2 leads to SEU with *random* utility.

Sec. 3 adds assumptions under which the DM strategy choice becomes the fully probabilistic design. It focuses on utilities, which assign the same values to equally probable behaviours having the equal losses, Assu. 6. It shows that the choice among equivalent utilities, leading to the same optimal closed DM loop, guarantees Pareto's optimality, Prop. 3. It characterises representatives of equivalent utilities, Prop. 4. It singles out CE as the only representative, which avoids the design-induced dependence between a pair independent DM tasks artificially connected into single tasks, Prop. 5. The corresponding design is called FPD. Props. 6, 7 show that SEU is a limiting and a special case FPD. Props. 8 and Prop. 9 recall the solutions of FPD and the needed Bayesian (stochastic) filtering.

Sec. 4 uses FPD for solving the key supporting DM tasks: an approximation of a known pd, Prop. 10, and an extension of a partial specification to the joint pd Prop. 11. Sec. 5 shows the use of these tools for an approximate Kalman filtering, Prop. 13, an extension of deterministic models to probabilistic ones, Prop. 14, the extension of marginal or conditional pds to a joint pd, Prop. 15, and the combination of several pds, Prop. 16. Sec. 6 provides concluding remarks.

1.3. FPD-Related Applications

This part samples the published applications and numerical examples resulting from the presented DM theory. The features brought by FPD are stressed and the exploited propositions are referred. The samples concern the works we have been directly involved in. A full coverage would need a survey paper.

The probabilistic advisory system [15] rely on FPD, Props. 7, 8. Handling of universally-approximating dynamic mixtures was made feasible via recursively applied projections, Prop. 10. The universal system was used for guiding operators of rolling mils, medical doctors curing a thyroid-gland cancer and traffic-control operators. Essentially, the advisory system serves as an upper-level, dynamic, adaptive feedback recommending the adequate actions to a human in a complex decision loop. The unified probabilistic description of the environment, of the DM aims and of the designed decision rules, Prop. 5, allowed to create a universal interface. It was well understood by rolling-mill operators (having just a basic education) as well as by highly qualified medical doctors (their expertise is out of a formalised and computerised DM). Moreover, the quantification of DM aims, expressed by the ideal probability within FPD, was made adaptive by deriving it from the learnt model [11]. A good mutual weighting of entries of a 12-dimensional state was achieved as full-scale industrial experiments at the rolling mill confirmed.

Extensive tests with FPD-based, automated knowledge elicitation serving to a widely-applied autoregressive-regressive model confirmed a significant improvements of critical transient behaviour of the gained adaptive predictors [14]. The elicitation respects incompleteness and uncertainty of the processed knowledge via Prop. 11.

The approximation of pds according to Prop. 10 forms the core of approximate Bayesian recursive estimation [12]. Its excellent properties were demonstrated on recursive estimation of the regression model with Cauchy noise, which well models time-series with heavy outliers. The application to high-order Markov chains [13] led to an efficient predictor whose performance was illustrated on predicting pewee song phrases and a prediction of sales demands.

Extensive encouraging simulation results of fully-scalable distributed, FPD-based, control and cooperation are in [18], Props. 15, 16.

Other tests concerned of softly cooperating heating systems as well as of Kalman filters, of balancing exploration with exploitation, of a predictor-based feeding external knowledge into learning. Sec. 5 offers other direct contributions to applications, namely, approximate Kalman filtering, Prop. 13 and to an extension of deterministic models to probabilistic ones, Prop. 14.

1.4. Related Works

References to related works are mostly given during the presentation. Thus, it makes sense just to mention the key clusters of the related research.

Seminal works connected with SEU are [8, 26]. All fall under the top down view of cybernetics and artificial intelligence. Excellent works reflecting backbone of the current main research stream are [3, 22].

FPD related theories can be tracked back to the first systematic relation of entropy (CE to uniform pd) and control (dynamic DM) [25]. Probabilistic modelling of closed DM loops origins in [29]. The closed-loop models are also called strategic measures and studied in connection with Markov decision processes. A use of CE for supporting DM tasks can be tracked to [2, 27]. FPD was proposed in [11], elaborated to general case in and axiomatised in [19]. An independently derived KL[1] control [10] or KL-constrained optimisation [5] can be seen as FPD versions.

The related theory of rational inattention has a strong impact in economy [28]. Probabilistic description of aims overcomes problem of deductive combination of multiple attributes of SEU [30]. Moreover, it lets multiple agents cooperate [17] without falling into the complexity trap of the systematic methodology of Bayesian games.

[1]Kullback-Leibler divergence is one of many alternative names for CE, [20].

2. Prescriptive DM Theory

The presented SEU generalisation makes the involved utility random.

2.1. DM and Quantified Ordering of Closed-Loop Behaviours

DM selects actions $a^{|\boldsymbol{t}|} = (a_t \in \boldsymbol{a}_t)_{t\in\boldsymbol{t}}$. They realise at discrete time moments $t \in \boldsymbol{t} = \{1, \ldots, |\boldsymbol{t}|\}$, $|\boldsymbol{t}| < \infty$. The actions influence the closed-loop states $(x_t \in \boldsymbol{x}_t)_{t\in\boldsymbol{t}}$. The actions and states constitute the closed-loop behaviour

$$b = \left(x^{|\boldsymbol{t}|}, a^{|\boldsymbol{t}|}\right) = (x_t, a_t)_{t\in\boldsymbol{t}} \in \boldsymbol{b}. \tag{1}$$

The agent generates its actions by an optional decision strategy $\mathsf{s} \in \boldsymbol{\mathsf{s}}$ consisting of decision rules s_t, $t \in \boldsymbol{t}$. The rule s_t maps the knowledge k^{t-1}, the accessible part of the behaviour enriched by a prior knowledge k^0, on the action a_t

$$\mathsf{s} = (\mathsf{s}_t : \boldsymbol{k}^{t-1} \to \boldsymbol{a}_t)_{t\in\boldsymbol{t}} \in \boldsymbol{\mathsf{s}}. \tag{2}$$

The actions are chosen in order to reach agent's aims while respecting its constraints. The agent's aims and constraints provide a preferential ordering $\preccurlyeq^{\boldsymbol{b}}$ of behaviours $b^\alpha, b^\beta \in \boldsymbol{b}$ (1). The adopted interpretation of this relation is

$$b^\alpha \preccurlyeq^{\boldsymbol{b}} b^\beta \ \text{ means that } \ b^\alpha \text{ is preferred against } b^\beta. \tag{3}$$

The agent may specify its preferences only on some behaviour pairs. The specification can always be completed. It suffices to include a "pointer" to all meaningful completions of the preferential ordering into the unseen state part.

The reviewed prescriptive DM theory supports consistent agents with transitive preferential orderings. This is general enough. Even the behaviourally observed intransitivity can be respected by considering an appropriate unseen agent's state.

Assumption 1 (Transitivity of Complete Ordering $\preccurlyeq^{\boldsymbol{b}}$). *The preferential ordering $\preccurlyeq^{\boldsymbol{b}}$ (3) is: a) non-empty; b) complete; c) transitive. The transitivity means that for $b^\alpha, b^\beta, b^\gamma \in \boldsymbol{b}$*

$$(b^\alpha \preccurlyeq^{\boldsymbol{b}} b^\beta \ \text{ and } \ b^\beta \preccurlyeq^{\boldsymbol{b}} b^\gamma) \ \Rightarrow \ b^\alpha \preccurlyeq^{\boldsymbol{b}} b^\gamma. \qquad \blacksquare$$

The complete preferential ordering $\preccurlyeq^{\boldsymbol{b}}$ induces the indifference $\approx^{\boldsymbol{b}}$ and the strict preference $\prec^{\boldsymbol{b}}$

$$\begin{aligned} b^\alpha &\approx^{\boldsymbol{b}} b^\beta \Leftrightarrow (b^\alpha \preccurlyeq^{\boldsymbol{b}} b^\beta \text{ and } b^\beta \preccurlyeq^{\boldsymbol{b}} b^\alpha) \\ b^\alpha &\prec^{\boldsymbol{b}} b^\beta \Leftrightarrow (b^\alpha \preccurlyeq^{\boldsymbol{b}} b^\beta \text{ and } b^\alpha \not\approx^{\boldsymbol{b}} b^\alpha). \end{aligned} \tag{4}$$

The complete ordering $\preccurlyeq^{\boldsymbol{b}}$ determines open intervals (b^α, b^β)

$$(b^\alpha, b^\beta) = \{b \in \boldsymbol{b} : b^\alpha \prec^{\boldsymbol{b}} b \prec^{\boldsymbol{b}} b^\beta\}. \tag{5}$$

They serve as neighbourhoods of bs in them and they enter the continuity definition of numerical functions defined on $\boldsymbol{b}$.

The targeted algorithmic DM requires a quantification of $\preccurlyeq^{\boldsymbol{b}}$.

Assumption 2 (Ordering Quantification). *A continuous quantification* $\mathsf{q}^{\boldsymbol{b}} : \boldsymbol{b} \rightarrow \mathbb{R} = [-\infty, \infty]$ *exists for which*

$$\begin{aligned} b^{\alpha} \prec^{\boldsymbol{b}} b^{\beta} \quad &\Leftrightarrow \quad \mathsf{q}^{\boldsymbol{b}}(b^{\alpha}) < \mathsf{q}^{\boldsymbol{b}}(b^{\beta}) \\ b^{\alpha} \approx^{\boldsymbol{b}} b^{\beta} \quad &\Leftrightarrow \quad \mathsf{q}^{\boldsymbol{b}}(b^{\alpha}) = \mathsf{q}^{\boldsymbol{b}}(b^{\beta}). \end{aligned} \tag{6}$$

■

The value of $\mathsf{q}^{\boldsymbol{b}}$ is known when knowing its argument. This trivial, but methodologically important, fact makes $\mathsf{q}^{\boldsymbol{b}}$ suitable only for *a posteriori* evaluation of the behaviour preferability.

Assu. 2 restricts us to orderings $\preccurlyeq^{\boldsymbol{b}}$ with a *countable dense subset of open intervals.* It is seen from the next rewording of Theorem II from [6].

Proposition 1 (Existence of Quantification). *Let* $\preccurlyeq^{\boldsymbol{b}}$ *meet Assu. 1. Then, Assu. 2 is met if a* countable *subset of open intervals (5) exists such that any open interval is a union of intervals from this subset.* ■

The work [6] also provides an example of $\preccurlyeq^{\boldsymbol{b}}$, which lacks the mentioned countable subsets and has no quantification. Prop. 1 supports the conjecture exploited when dealing with supporting DM tasks in Sec. 4:

The results obtained for the countable amount of behaviour realisations continuously extend to all preferential orderings meeting Assu. 2.	(7)

2.2. Quantified Strategy Ordering

The agent wants to select the best strategy $\mathsf{s}^{opt} \in \mathbf{s}$. Thus, it deals with an ordering $\preccurlyeq^{\mathbf{s}}$ of compared strategies $\mathsf{s}^{\alpha}, \mathsf{s}^{\beta} \in \mathbf{s}$

$$\mathsf{s}^{\alpha} \preccurlyeq^{\mathbf{s}} \mathsf{s}^{\beta} \text{ means that } \mathsf{s}^{\alpha} \text{ is better than } \mathsf{s}^{\beta}. \tag{8}$$

The constructed universal ordering $\preccurlyeq^{\mathbf{s}}$ is to be complete. It has to admit the restriction of the strategy set (2) to the subsets of compared strategies containing strategy pairs. This completeness justification is the only difference from the preferential-ordering discussion. The exposition from Assu. 1 up to Prop. 1 applies with the superscript $^{\mathbf{s}}$ replacing $^{\boldsymbol{b}}$. The next assumption covers and extends Assus. 1, 2.

Assumption 3 (Transitivity and Quantification of $\preccurlyeq^{\mathbf{s}}$). *The strategy ordering* $\preccurlyeq^{\mathbf{s}}$ *(8) is: a) non-empty; b) complete; c) transitive; d) continuously quantifiable by* $\mathsf{q}^{\mathbf{s}} : \mathbf{s} \rightarrow \mathbb{R}$.

Moreover, the optimal strategy $\mathsf{s}^{opt} \in \mathbf{s}$ *exists*

$$\mathsf{s}^{opt} \preccurlyeq^{\mathbf{s}} \mathsf{s}, \ \forall \mathsf{s} \in \mathbf{s}. \tag{9}$$

This relaxable request avoids an ε-optimality machinery. ■

2.3. DM Under Uncertainty

DM is hard due to the always present *uncertainty.* It prevents the assignment of a unique behaviour to a given strategy. Its DM consequences are inspected here without caring about its causes. They are immaterial for DM. The uncertainty primarily makes hard the "harmonisation" of the strategy ordering $\preccurlyeq^{\mathbf{s}}$ with the given preferential ordering $\preccurlyeq^{\boldsymbol{b}}$.

Assumption 4 (Harmonised[2] Orderings $\preccurlyeq^{\mathsf{s}}$ and $\preccurlyeq^{\boldsymbol{b}}$). *No strategy* $\mathsf{s} \in \mathbf{s}$ *exists leading,* irrespectively of uncertainty*, to behaviours preferred against those resulting from the optimal strategy* s^{opt} *(9) .* ■

All u*n*certainty models take behaviours as images of strategy-dependent mappings b^{s} of the uncertainty $n \in \boldsymbol{n} \neq \emptyset$

$$\mathsf{b}^{\mathsf{s}} : \boldsymbol{n} \to \boldsymbol{b}, \quad \mathsf{s} \in \mathbf{s}. \tag{10}$$

The ordered set $\boldsymbol{b}$ then consists of the union of b^{s}-images

$$\boldsymbol{b} = \{b : \exists n \in \boldsymbol{n} \text{ so that } b = \mathsf{b}^{\mathsf{s}}(n)\}_{\mathsf{s}\in\mathbf{s}}. \tag{11}$$

The uncertainties $\boldsymbol{n}$ and the mappings b^{s}, $\mathsf{s} \in \mathbf{s}$, have to respect key properties of $\boldsymbol{b}$ even when seen as the set (11). Primarily, they should preserve the lattice structure of $(\boldsymbol{b}, \preccurlyeq^{\boldsymbol{b}})$. Thus, b^{s} are quantum-mechanical observables, see [7] and the sketch in Appendix 7. The further treatment takes b^{s} as a less general $\boldsymbol{b}$-valued random variable. It maps $\sigma^{\boldsymbol{n}}$-algebra on uncertainties $\boldsymbol{n}$ to Borel's $\sigma^{\boldsymbol{b}}$-algebra constructed on $\boldsymbol{b}$.

The mapping b^{s} (10) transforms the quantification $\mathsf{q}^{\boldsymbol{b}}$ (6) into the real-valued function of uncertainties $\mathsf{q}^{\boldsymbol{b}} \circ \mathsf{b}^{\mathsf{s}} : \boldsymbol{n} \to \mathbb{R}$. Through this, the compared strategies $\mathsf{s} \in \mathbf{s}$ and the quantification $\mathsf{q}^{\boldsymbol{b}}$ generate the set $\mathbf{q}^{\boldsymbol{n}\mathsf{q}^{\boldsymbol{b}}}$ of uncertainty functions

$$\mathbf{q}^{\boldsymbol{n}\mathsf{q}^{\boldsymbol{b}}} = \left\{\mathsf{q}^{\boldsymbol{n}} : \exists \mathsf{s} \in \mathbf{s} \text{ such that } \mathsf{q}^{\boldsymbol{n}}(n) = \mathsf{q}^{\boldsymbol{b}}(\mathsf{b}^{\mathsf{s}}(n)), \forall n \in \boldsymbol{n}\right\}. \tag{12}$$

For a fixed quantification $\mathsf{q}^{\boldsymbol{b}}$, the of strategy ordering $\preccurlyeq^{\mathsf{s}}$ (8) defines the equivalent ordering $\preccurlyeq^{\mathsf{q}^{\boldsymbol{n}}}$ of functions in $\mathbf{q}^{\boldsymbol{n}\mathsf{q}^{\boldsymbol{b}}}$

$$\mathsf{q}^{\boldsymbol{b}} \circ \mathsf{b}^{\mathsf{s}^{\alpha}} \preccurlyeq^{\mathsf{q}^{\boldsymbol{n}}} \mathsf{q}^{\boldsymbol{b}} \circ \mathsf{b}^{\mathsf{s}^{\beta}} \quad \Leftrightarrow \quad \mathsf{s}^{\alpha} \preccurlyeq^{\mathsf{s}} \mathsf{s}^{\beta}. \tag{13}$$

The definition (13) makes the quantification q^{s} of a strategy ordering $\preccurlyeq^{\mathsf{s}}$ equivalent to the quantification $\mathcal{Q}$ of the ordering of $\preccurlyeq^{\mathsf{q}^{\boldsymbol{n}}}$. It orders *real-valued uncertainty functions*. The quantifying functional $\mathcal{Q}$ acts on $\mathbf{q}^{\boldsymbol{n}\mathsf{q}^{\boldsymbol{b}}}$. The *universal functional* $\mathcal{Q}$ acts on the set $\mathbf{q}^{\boldsymbol{n}}$ covering all sets (12)

$$\begin{aligned}\mathbf{q}^{\boldsymbol{n}} \quad = \quad & \cup_{\mathsf{q}^{b}\in\mathbf{q}^{b}}\mathbf{q}^{\boldsymbol{n}\mathsf{q}^{\boldsymbol{b}}} = \left\{\mathsf{q}^{\boldsymbol{n}} : \boldsymbol{n} \to \mathbb{R}, \exists(\mathsf{s} \in \mathbf{s}, \mathsf{q}^{\boldsymbol{b}} \in \mathbf{q}^{\boldsymbol{b}})\right. \\ & \left.\text{such that } \mathsf{q}^{\boldsymbol{n}}(n) = \mathsf{q}^{\boldsymbol{b}}(\mathsf{b}^{\mathsf{s}}(n)), \ \forall n \in \boldsymbol{n}\right\}.\end{aligned} \tag{14}$$

The set $\mathbf{q}^{\boldsymbol{n}}$ (14) includes all uncertainty functions arising in DMs dealing with the behaviours $b = \mathsf{b}^{\mathsf{s}}(n)$, generated by common uncertainties $n \in \boldsymbol{n}$ and by strategies $\mathsf{s} \in \mathbf{s}$ from the widest set $\mathbf{s}$ of compared strategies. Individually, DM tasks consider their specific subsets of compared strategies and their quantifications $\mathsf{q}^{\boldsymbol{b}}$ of preferential orderings.

After specifying the functional $\mathcal{Q} : \mathbf{q}^{\boldsymbol{n}} \to \mathbb{R}$, its restriction to the set $\mathbf{q}^{\boldsymbol{n}\mathsf{q}^{\boldsymbol{b}}}$ (12), given by a specific quantification $\mathsf{q}^{\boldsymbol{b}}$ (6) of preferential ordering $\preccurlyeq^{\boldsymbol{b}}$ (3), quantifies the strategy ordering $\preccurlyeq^{\mathsf{s}}$ (8) via the equivalence (13).

Assumption 5 (Smooth Local Harmonised Functionals $\mathcal{Q}$). *The considered universal quantifying functionals* $\mathcal{Q} \in \boldsymbol{\mathcal{Q}}$*:*

(a) are sequentially and boundedly uniformly continuous[3]*,*
(b) are locally additive

$$(\mathsf{q}^{\boldsymbol{n}\alpha}\mathsf{q}^{\boldsymbol{n}\beta} = 0 \text{ on } \boldsymbol{n}) \Rightarrow \mathcal{Q}(\mathsf{q}^{\boldsymbol{n}\alpha} + \mathsf{q}^{\boldsymbol{n}\beta}) = \mathcal{Q}(\mathsf{q}^{\boldsymbol{n}\alpha}) + \mathcal{Q}(\mathsf{q}^{\boldsymbol{n}\beta}), \quad (15)$$

(c) harmonise $\preccurlyeq^{\mathsf{s}}$ *with* $\preccurlyeq^{\boldsymbol{b}}$*, i.e. meet Assu. 4.* ■

Proposition 2 (Representation of Local Functionals). *The universal functional* $\mathcal{Q} : \mathbf{q}^{\boldsymbol{n}} \to \mathbb{R}$*, meeting Assu. 5 and restricted to functions in* $\mathbf{q}^{\boldsymbol{n}\mathsf{q}^{\boldsymbol{b}}}$ *(12) generated by a quantification* $\mathsf{q}^{\boldsymbol{b}}$ *of the preferential ordering* $\preccurlyeq^{\boldsymbol{b}}$*, has the form*

$$\mathcal{Q}(\mathsf{q}^{\boldsymbol{n}}) = \int_{\boldsymbol{n}} \kappa(\mathsf{q}^{\boldsymbol{n}}(n), n)\mu(\mathrm{d}n). \quad (16)$$

There μ *is a probabilistic measure on* $(\boldsymbol{n}, \sigma^{\boldsymbol{n}})$ *while the measurable* kernel[4] κ *increases in* $\mathsf{q}^{\boldsymbol{n}}(n)$ *and* $\kappa(0, n) = 0$*.* ■

The measure μ is assumed to have the *probability density (pd)* with respect to a dominating measure denoted d•. Then, the back-substitution of (10) within the integral (16) gives

$$\mathcal{Q}(\mathsf{q}^{\boldsymbol{n}}) = \mathcal{Q}(\mathsf{q}^{\boldsymbol{b}}) = \int_{\boldsymbol{b}} \kappa(\mathsf{q}^{\boldsymbol{b}}(b), (\mathsf{b}^{\mathsf{s}})^{-1}(b))\mathsf{c}^{\mathsf{s}}(b)\ \mathrm{d}b = \mathcal{E}^{\mathsf{s}}(\mathsf{i}^{\mathsf{s}})$$

$$\mathsf{i}^{\mathsf{s}}(b) = \kappa(\mathsf{q}^{\boldsymbol{b}}(b), (\mathsf{b}^{\mathsf{s}})^{-1}(b)) \text{ is called perfomance index.} \quad (17)$$

This quantifies the strategy ordering in a way harmonised with the agent's aims and defines the optimal strategy (9)

$$\mathsf{s}^{opt} \in \mathrm{Arg}\min_{\mathsf{s}\in\mathbf{s}} \mathcal{E}^{\mathsf{s}}(\mathsf{i}^{\mathsf{s}}). \quad (18)$$

Unlike SEU [26], the quantification of the strategy ordering is non-linear in the optimised strategy. The non-standard dependence of the performance index i^{s} (17) on the strategy s comes from the second argument of the kernel in (16). It appears due to weakening of the usually assumed additivity of the functional $\mathcal{Q}$ to its *local* additivity (15). Importantly, the measure μ, and thus the pd c^{s}, serves to all agents facing the same uncertainties. In this sense, the pd c^{s} is the *objective model of the closed decision loop.* The chain rule for pds [22], together with the state and knowledge definitions, implies, $\forall b \in \boldsymbol{b}$ (1),

$$\mathsf{c}^{\mathsf{s}}(b) = \overbrace{\prod_{t\in\boldsymbol{t}} \mathsf{m}(x_t|a_t, x_{t-1})}^{\mathsf{m}(b)} \overbrace{\prod_{t\in\boldsymbol{t}} \mathsf{s}(a_t|k^{t-1})}^{\mathsf{s}(b)} = \mathsf{m}(b)\mathsf{s}(b). \quad (19)$$

The conditional pds $\mathsf{m}(x_t|a_t, x_{t-1})$, $t \in \boldsymbol{t}$, describe the transitions from the state x_{t-1} to the state x_t for the action a_t. They model the agent's environment[5]. The pds $\mathsf{s}(a_t|k^{t-1})$, $t \in \boldsymbol{t}$, assign the probability to the action a_t under the knowledge

$$k^{t-1} = (a_\tau, x_\tau^s)_{\tau\le t-1} \text{ enriching a prior knowledge } k_0. \quad (20)$$

[3] [24] defines exactly these notions intuitively understandable for the considered functions of finite uniform norm. The used representation Theorem 9.3-5 in [24] operates on such functions but it holds for more general function spaces.
[4] It acts as the randomised utility function.
[5] The agent is a part of its environment and x_t also models its state.

There, x_t^s is the seen part of the closed-loop state, i.e.

$$x_t = (x_t^s, x_t^u) = (\text{seen}, \text{unseen}) \text{ state parts}. \tag{21}$$

The pds $\mathsf{s}(a_t|k^{t-1}),\ t \in \boldsymbol{t}$, model the strategy.

3. Fully Probabilistic Design

Assumptions formulated here single out FPD as the relevant DM theory.

3.1. FPD as Prescriptive DM Theory

Performance indices $\mathsf{i}^{\mathsf{s}^\alpha}, \mathsf{i}^{\mathsf{s}^\beta}$ (17) resulting into the optimal strategies $\mathsf{s}^{opt\alpha}, \mathsf{s}^{opt\beta}$ (18), which give the same closed-loop model $\mathsf{c}^{\mathsf{i}}(b),\ \forall b \in \boldsymbol{b}$,

$$\mathsf{c}^{\mathsf{s}^{opt\alpha}} = \mathsf{c}^{\mathsf{s}^{opt\beta}} = \mathsf{c}^{\mathsf{i}} \qquad \text{are equivalent}. \tag{22}$$

The choice of the equivalent performance indices can be replaced by the specification of c^{i} (22) referred as the *ideal closed-loop model.* Any representant of the corresponding equivalence class of performance indices then serves for the strategy design. Such a representant is here constructed under additional, broadly acceptable, assumptions.

Assumption 6 (Equally Probable Indifferent Behaviours). *For a strategy $\mathsf{s} \in \boldsymbol{\mathsf{s}}$, let the behaviours $b^\alpha, b^\beta \in \boldsymbol{b}$ be equally probable $\mathsf{c}^{\mathsf{s}}(b^\alpha) = \mathsf{c}^{\mathsf{s}}(b^\beta)$ and let the agent be indifferent to them $b^\alpha \approx^{\boldsymbol{b}} b^\beta \Leftrightarrow \mathsf{q}^{\boldsymbol{b}}(b^\alpha) = \mathsf{q}^{\boldsymbol{b}}(b^\beta)$. Then, the performance index assigns them the same values $\mathsf{i}^{\mathsf{s}}(b^\alpha) = \mathsf{i}^{\mathsf{s}}(b^\beta)$.* ■

Assu. 6 restricts performance indices (17) to those dependent on the function $\mathsf{q}^{\boldsymbol{b}}$ and the closed-loop model c^{s}

$$\mathsf{i}^{\mathsf{s}}(b) = \kappa(\mathsf{q}^{\boldsymbol{b}}(b), (\mathsf{b}^{\mathsf{s}})^{-1}(b)) = \kappa(\mathsf{q}^{\boldsymbol{b}}(b), \mathsf{c}^{\mathsf{s}}(b)). \tag{23}$$

The next proposition confirms that under Assu. 6 the use of a representant of equivalent performance indices does not violate the harmonisation Assu. 4.

Proposition 3 (Equivalent Indices Harmonise $\preccurlyeq^{\mathsf{s}}$ with $\preccurlyeq^{\boldsymbol{b}}$). *Optimal strategies of performance indices equivalent with respect to an ideal closed-loop model (22) meet Assu. 4.*

Proof Let c^{i} be a fixed ideal closed-loop model (22). It delimits equivalent performance indices (23) that use preferential ordering quantified by $\mathsf{q}^\alpha, \mathsf{q}^\beta$. Let $\mathsf{s}^{opt\alpha}$, $\mathsf{s}^{opt\beta} \in \boldsymbol{\mathsf{s}}$ be the corresponding optimal strategies (18). The mapping (10) $\mathsf{b}^{\mathsf{s}} : \boldsymbol{n} \to \boldsymbol{b}$ allows us to express the harmonisation violation

$$\mathsf{q}^\alpha(\mathsf{b}^{\mathsf{s}^{opt\beta}}(n)) \leq \mathsf{q}^\alpha(\mathsf{b}^{\mathsf{s}^{opt\alpha}}(n)),\ \forall n \in \boldsymbol{n}, \tag{24}$$

with a sharp inequality on a subset of $\boldsymbol{n}$ having a positive probability given by the universal measure μ (16). Such a violation, the universality of the kernel κ (16), the assumed equivalence $\mathsf{c}^{\mathsf{s}^{opt\alpha}} = \mathsf{c}^{\mathsf{s}^{opt\beta}} = \mathsf{c}^{\mathsf{i}}$ and the substitutions $b =$

$\mathsf{b}^{\mathsf{s}^{opt\alpha}}(n),\ b = \mathsf{b}^{\mathsf{s}^{opt\beta}}(n)$ give the next inequality, which contradicts the optimality of $\mathsf{s}^{opt\alpha}$ with respect to the preferential ordering quantified by q^{α},

$$\int_{\boldsymbol{b}} \kappa(\mathsf{q}^{\alpha}(b), \underbrace{\mathsf{c}^{\mathsf{s}^{opt\alpha}}(b)}_{=\mathsf{c}^{\mathsf{i}}})\mathsf{c}^{\mathsf{s}^{opt\alpha}}(b)\mathrm{d}b = \int_{\boldsymbol{b}} \kappa(\mathsf{q}^{\alpha}(b), \underbrace{\mathsf{c}^{\mathsf{s}^{opt\beta}}(b)}_{=\mathsf{c}^{\mathsf{i}}})$$
$$\times \mathsf{c}^{\mathsf{s}^{opt\alpha}}(b)\mathrm{d}b \underbrace{<}_{(24)} \int_{\boldsymbol{b}} \kappa(\mathsf{q}^{\alpha}(b), \mathsf{c}^{\mathsf{s}^{opt\alpha}}(b))\mathsf{c}^{\mathsf{s}^{opt\alpha}}(b)\mathrm{d}b. \qquad \blacksquare$$

Assumption 7 (Support of the Ideal Closed-Loop Model). *The support* $\mathrm{supp}[\mathsf{c}^{\mathsf{i}}]$ *of the ideal closed-loop model* c^{i} *meets*

$$\mathrm{supp}[\mathsf{c}^{\mathsf{i}}] = \{b \in \boldsymbol{b} : \mathsf{c}^{\mathsf{i}}(b) > 0\} \supseteqq \cup_{\mathsf{s}\in\mathbf{s}}\mathrm{supp}[\mathsf{c}^{\mathsf{s}}]. \qquad \blacksquare$$

Proposition 4 (Jensen's Representant). *Let us consider performance indices (17) meeting Assus. 6, 7. Then,*

$$\mathsf{i}^{\mathsf{s}}(b) = \mathsf{w}(\rho^{\mathsf{s}}(b)), \quad \textit{with} \quad \rho^{\mathsf{s}}(b) = \frac{\mathsf{c}^{\mathsf{s}}(b)}{\mathsf{c}^{\mathsf{i}}(b)}, \tag{25}$$

where the function $\rho\mathsf{w}(\rho)$ *is strictly convex for* $\rho > 0$*, represents the equivalent performance indices leading to the given ideal closed-loop model* c^{i} *(22).*

Proof The form of i^{s} in (25) meets Assu. 6, see (23), as Assu. 7 allows to express it as a function of the ratio ρ^{s}. By definition, the optimal strategies, assigned to the equivalent performance indices, give the same closed-loop model c^{i}. Their addition to (25) with a positive weight does not change the optimisation result. Jensen's inequality for convex $\rho\mathsf{w}(\rho)$, [24], implies, for any $\mathsf{s} \in \mathbf{s}$,

$$\begin{aligned} \mathcal{E}^{\mathsf{s}}[\mathsf{i}^{\mathsf{s}}] &= \int_{\boldsymbol{b}} \rho^{\mathsf{s}}(b)\mathsf{w}(\rho^{\mathsf{s}}(b))\mathsf{c}^{\mathsf{i}}(b)\ \mathrm{d}b \\ &\geq \underbrace{\int_{\boldsymbol{b}} \rho^{\mathsf{s}}(b)\mathsf{c}^{\mathsf{i}}(b)\ \mathrm{d}b}_{=1} \times \mathsf{w}\Bigg(\underbrace{\int_{\boldsymbol{b}} \rho^{\mathsf{s}}(b)\mathsf{c}^{\mathsf{i}}(b)\ \mathrm{d}b}_{=1}\Bigg) = \mathsf{w}(1). \end{aligned} \tag{26}$$

The minimiser of the expected performance index (26) leads to $\rho^{\mathsf{s}^{opt}} = 1 \Leftrightarrow \mathsf{c}^{\mathsf{s}^{opt}} = \mathsf{c}^{\mathsf{i}}$. $\blacksquare$

Assumption 8 (Avoiding a Design-Induced Dependence). *The optimal strategy gained for concatenated but independent DM problems consists of the optimal strategies obtained for the individual DM problems.* $\blacksquare$

This assumption selects cross entropy (CE, [20]) among I-divergences determined by functions w (25), [19].

Proposition 5 (FPD). *The function* $\mathsf{w}(\rho) = \ln(\rho)$ *(25) meets Assu. 8. It defines the optimal strategy* s^{opt} *as the minimiser of the cross entropy* $\mathcal{D}(\mathsf{c}^{\mathsf{s}}||\mathsf{c}^{\mathsf{i}}) = \mathcal{E}^{\mathsf{s}}[\ln(\mathsf{c}^{\mathsf{s}}/\mathsf{c}^{\mathsf{i}})]$

$$\mathsf{s}^{opt} \in \mathrm{Arg}\min_{\mathsf{s}\in\mathbf{s}} \int_{\boldsymbol{b}} \mathsf{c}^{\mathsf{s}}(b) \ln\left(\frac{\mathsf{c}^{\mathsf{s}}(b)}{\mathsf{c}^{\mathsf{i}}(b)}\right) \mathrm{d}b = \mathrm{Arg}\min_{\mathsf{s}\in\mathbf{s}} \mathcal{D}(\mathsf{c}^{\mathsf{s}}||\mathsf{c}^{\mathsf{i}}). \tag{27}$$

The optimisation (27) is called fully probabilistic design of decision strategies*.*

Proof An artificially connected pair of independent DM tasks operates on behaviours $b = (b^\alpha, b^\beta) \in \boldsymbol{b} = \boldsymbol{b}^\alpha \times \boldsymbol{b}^\beta$, $\boldsymbol{b}^\alpha \cap \boldsymbol{b}^\beta = \emptyset$, and employs the ideal closed-loop model $\mathsf{c}^{\mathsf{i}}(b) = \mathsf{c}^{\mathsf{i}\alpha}(b^\alpha)\mathsf{c}^{\mathsf{i}\beta}(b^\beta)$. The optimised functional is to be the sum of functionals corresponding to the involved particular DM tasks. It gives

$$\begin{aligned} 0 &= \int_{\boldsymbol{b}^\alpha}\int_{\boldsymbol{b}^s} \mathsf{c}^{\mathsf{s}^\alpha}(b^\alpha)\mathsf{c}^{\mathsf{s}^\beta}(b^\beta) \\ &\times \Big[\mathsf{w}(\rho^{\mathsf{s}^\alpha}(b^\alpha)\rho^{\mathsf{s}^\beta}(b^\beta)) - \mathsf{w}(\rho^{\mathsf{s}^\alpha}(b^\alpha)) - \mathsf{w}(\rho^{\mathsf{s}^\beta}(b^\beta))\Big]\, \mathrm{d}b^\alpha \mathrm{d}b^\beta. \end{aligned}$$

This implies the functional equation $\mathsf{w}(\rho^{\mathsf{s}^\alpha}\rho^{\mathsf{s}^\beta}) = \mathsf{w}(\rho^{\mathsf{s}^\alpha}) + \mathsf{w}(\rho^{\mathsf{s}^\beta})$ for w, operating on $\rho^{\mathsf{s}} > 0$. It has $\mathsf{w}(\bullet) = \ln(\bullet)$ as its only smooth solution. ■

Assu. 7 is restrictive. It excludes ideal closed-loop models resulting from generic *deterministic* SEU strategies. The next proposition, however, shows that any such DM, delimited by a *strategy-independent* performance index i, can always be arbitrarily well approximated by FPD, [19].

Proposition 6 (SEU as an FPD Limit). *Let the performance index* i *be strategy-independent. Let us consider FPD with the ideal closed-loop model of the next form, meeting Assu. 7,*

$$\mathsf{c}^{\mathsf{i}\lambda} = \frac{\underline{\mathsf{c}}\exp(-\mathsf{i}/\lambda)}{\zeta^\lambda}. \tag{28}$$

There, $\lambda > 0$ *and the optional pd* $\underline{\mathsf{c}}$ *with* $\mathrm{supp}[\underline{\mathsf{c}}] = \boldsymbol{b}$ *makes the pd* $\mathsf{c}^{\mathsf{i}\lambda}$ *normalisable,* $\zeta^\lambda < \infty$*. Then, the corresponding FPD-optimal strategy* $\mathsf{s}^{opt\lambda}$ *converges for* $\lambda \to 0^+$ *to the SEU-optimal strategy* $\mathsf{s}^{opt\mathsf{i}} \in \mathrm{Arg}\min_{\mathsf{s}\in\boldsymbol{\mathsf{s}}} \mathcal{E}^{\mathsf{s}}[\mathsf{i}]$.

Proof It holds

$\mathsf{s}^{opt\lambda} \in \mathrm{Arg}\min_{\mathsf{s}\in\boldsymbol{\mathsf{s}}} \left[\mathcal{E}^{\mathsf{s}}[\mathsf{i}] + \lambda\mathcal{D}(\mathsf{c}^s||\underline{\mathsf{c}})\right] = \mathrm{Arg}\min_{\mathsf{s}\in\boldsymbol{\mathsf{s}}} \mathcal{D}\Big(\mathsf{c}^s\Big|\Big|\frac{\underline{\mathsf{c}}\exp\left(-\mathsf{i}/\lambda\right)}{\zeta_\lambda}\Big)$.

The choice $\mathsf{s}^{opt\mathsf{i}}, \mathsf{s}^{opt\lambda}$ gives

$$\begin{aligned} 0 &\overbrace{\leq}^{\text{def. } \mathsf{s}^{opt\mathsf{i}}} \mathcal{E}^{\mathsf{s}^{opt\lambda}}[\mathsf{i}] - \mathcal{E}^{\mathsf{s}^{opt\mathsf{i}}}[\mathsf{i}] \overbrace{\leq}^{\lambda\mathcal{D}\left(\mathsf{c}^{s^{opt\lambda}}||\underline{\mathsf{c}}\right)\geq 0} \mathcal{E}^{\mathsf{s}^{opt\lambda}}[\mathsf{i}] + \lambda\mathcal{D}\big(\mathsf{c}^{\mathsf{s}^{opt\lambda}}||\underline{\mathsf{c}}\big) - \mathcal{E}^{\mathsf{s}^{opt\mathsf{i}}}[\mathsf{i}] \\ &\overbrace{\leq}^{\text{def. } \mathsf{s}^{opt\lambda}} \lambda\mathcal{D}\big(\mathsf{c}^{s^{opt\mathsf{i}}}||\underline{\mathsf{c}}\big) \to_{\lambda\to 0^+} 0. \end{aligned}$$

■

3.2. FPD with Leave to the Fate Option (LFO)

This part serves to supporting DM tasks discussed in Sec. 4 and connects SEU and FPD more tightly than Prop. 6.

Often, a decisive subpart b^d of b delimits the preferential ordering $\preccurlyeq^{\boldsymbol{b}} = (\preccurlyeq^{\boldsymbol{b}^d|b^r})_{b^r\in\boldsymbol{b}^r}$. The rest b^r complements b^d to b

$$b = (b^d, b^r) \in \boldsymbol{b}^d \times \boldsymbol{b}^r. \tag{29}$$

FPD (27) deals with the ideal closed-loop model on whole $\boldsymbol{b}$

$$\mathsf{c}^{\mathsf{i}}(b) = \mathsf{m}^{\mathsf{i}}(b)\mathsf{s}^{\mathsf{i}}(b) = \mathsf{c}^{\mathsf{i}}(b^d, b^r) = \mathsf{c}^{\mathsf{i}}(b^d|b^r)\mathsf{c}^{\mathsf{i}}(b^r). \tag{30}$$

The first factor of the last chain-rule factorisation is implied by agent's preferences and constraints on $\boldsymbol{b}^d$, both possibly dependent on b^r. The rest $\boldsymbol{d}^r$ just informs the agent about its environment. Thus, it makes sense to apply the following *leave to the fate option (LFO)*

$$\mathsf{c}^{\mathsf{i}}(b^r) = \mathsf{c}^{\mathsf{s}}(b^r). \tag{31}$$

It simply takes the optimised factor $\mathsf{c}^{\mathsf{s}}(b^r)$ as the ideal one. Under it, a tighter connection of FPD and SEU is seen.

Proposition 7 (FPD with LFO and Direct Relation to SEU). *Let $\boldsymbol{b}^r \neq \emptyset$. Then, FPD with LFO (31) reduces to*

$$\mathsf{s}^{opt} \in \operatorname{Arg}\min_{\mathsf{s}\in\mathbf{s}} \int_{\boldsymbol{b}^r} \mathsf{c}^{\mathsf{s}}(b^r) \left(\int_{\boldsymbol{b}^d} \mathsf{c}^{\mathsf{s}}(b^d|b^r) \ln\left(\frac{\mathsf{c}^{\mathsf{s}}(b^d|b^r)}{\mathsf{c}^{\mathsf{i}}(b^d|b^r)}\right) \mathrm{d}b^d \right) \mathrm{d}b^r \tag{32}$$

For LFO with $b^r = a^{|\boldsymbol{t}|} \Leftrightarrow \mathsf{s}^{\mathsf{i}} = \mathsf{s}$, FPD is SEU with the performance index $\mathsf{i} = \ln\left(\frac{\mathsf{m}}{\mathsf{m}^{\mathsf{i}}}\right)$ and $\boldsymbol{a}^{|\boldsymbol{t}|} = \{a^{|\boldsymbol{t}|} : \mathsf{i}(b) < \infty\}$.

Proof The common factor in the ratio $\mathsf{c}^{\mathsf{s}}/\mathsf{c}^{\mathsf{i}}$ cancels. It gives (32). The next statement is obvious from (32), (19), (30). ■

3.3. Solution of FPD

The next factorisation of the ideal closed-loop model, mimicking (19), serves to the solution of the general FPD[6]

$$\mathsf{c}^{\mathsf{i}}(b) = \overbrace{\prod_{t\in\boldsymbol{t}} \mathsf{m}^{\mathsf{i}}(x_t|a_t, x_{t-1})}^{\mathsf{m}^{\mathsf{i}}(b)} \overbrace{\prod_{t\in\boldsymbol{t}} \mathsf{s}^{\mathsf{i}}(a_t|k^{t-1})}^{\mathsf{s}^{\mathsf{i}}(b)}. \tag{33}$$

Proposition 8 (The FPD Solution). *For c^{s} (19) and c^{i} (33), the FPD-optimal strategy (27) has the rules, $t \in \boldsymbol{t}$,*

$$\begin{aligned} \mathsf{s}_t^{opt}(a_t|k^{t-1}) &= \frac{\mathsf{s}_t^{\mathsf{i}}(a_t|k^{t-1}) \exp[-\omega(a_t, k^{t-1})]}{\gamma(k^{t-1})} \\ \gamma(k^{t-1}) &= \int_{\boldsymbol{a}_t} \mathsf{s}_t^{\mathsf{i}}(a_t|k^{t-1}) \exp[-\omega(a_t, k^{t-1})] \, \mathrm{d}a_t \\ \omega(a_t, k^{t-1}) &= \mathcal{E}\left[\ln\left(\frac{\mathsf{m}(x_t|a_t, x_{t-1})}{\gamma(k_t)\mathsf{m}^{\mathsf{i}}(x_t|a_t, x_{t-1})}\right) \middle| a_t, k^{t-1} \right] \end{aligned} \tag{34}$$

This backward run starts with $\gamma(k^{|\boldsymbol{t}|}) = 1$, $\forall k^{|\boldsymbol{t}|} \in \boldsymbol{k}^{|\boldsymbol{t}|}$. ■

Prop. 8 is a version of stochastic dynamic programming [3] with $-\ln(\gamma(k^t))$ being the optimal value function. The minimising *randomised* decision rules are found explicitly. The expectation $\mathcal{E}[\bullet|a_t, k^{t-1}]$ used in (34) is made over the next state x_t and the unseen x_{t-1}^u, cf. (20) and (21). The next proposition provides the estimator $\mathsf{e}(x_t, x_{t-1}^u|a_t, k^{t-1})$, the pd needed for the evaluation of $\mathcal{E}[\bullet|a_t, k^{t-1}]$, [22].

[6] The assumption that the ideal decision rules uses the knowledge available to the optimised ones can be relaxed. It just simplifies the presentation.

Proposition 9 (Bayesian Filtering). *Starting from a prior pd* $\mathsf{e}(x_1|a_1,k_0) = \int_{\boldsymbol{x}_0^u} \mathsf{e}(x_1, x_0^u|a_1,k_0)\mathrm{d}x_0^u$, *Bayesian filtering recursively evaluates the estimator described by the pd*

$$\mathsf{e}(x_t, x_{t-1}^u|a_t, k^{t-1}) = \frac{\mathsf{m}(x_t|a_t, x_{t-1}) \int_{\boldsymbol{x}_{t-2}^u} \mathsf{e}(x_{t-1}, x_{t-2}^u|a_{t-1}, k_{t-2})\ \mathrm{d}x_{t-2}^u}{\int_{\boldsymbol{x}_{t-1}^u} \int_{\boldsymbol{x}_{t-2}^u} \mathsf{e}(x_{t-1}, x_{t-2}^u|a_{t-1}, k_{t-2})\ \mathrm{d}x_{t-2}^u\ \mathrm{d}x_{t-1}^u}. \tag{35}$$

Proof

$$\begin{aligned}
&\mathsf{e}(x_t, x_{t-1}^u|a_t, k^{t-1}) = \mathsf{e}(x_t|a_t, x_{t-1}^u, k^{t-1})\mathsf{e}(x_{t-1}^u|a_t, k^{t-1}) \\
&= \mathsf{m}(x_t|a_t, x_{t-1})\mathsf{e}(x_{t-1}^u|k^{t-1}) \\
&= \mathsf{m}(x_t|a_t, x_{t-1}) \frac{\mathsf{e}(x_{t-1}^s, x_{t-1}^u|a_{t-1}, k_{t-2})}{\int_{\boldsymbol{x}_{t-1}^u} \mathsf{e}(x_{t-1}^s, x_{t-1}^u|a_{t-1}, k_{t-2})\ \mathrm{d}x_{t-1}^u} \\
&= \frac{\mathsf{m}(x_t|a_t, x_{t-1}) \int_{\boldsymbol{x}_{t-2}^u} \mathsf{e}(x_{t-1}, x_{t-2}^u|a_{t-1}, k_{t-2})\ \mathrm{d}x_{t-2}^u}{\int_{\boldsymbol{x}_{t-1}^u} \int_{\boldsymbol{x}_{t-2}^u} \mathsf{e}(x_{t-1}, x_{t-2}^u|a_{t-1}, k_{t-2})\ \mathrm{d}x_{t-2}^u\ \mathrm{d}x_{t-1}^u}.
\end{aligned}$$

The derivation of (35) sequentially uses: a) the chain rule; b) the identity $k^{t-1} = (x_{t-1}^s, a_{t-1}, k_{t-2})$, the definition of the state $x_{t-1} = (x_{t-1}^s, x_{t-1}^u)$, the environment model (19) and fact that a_t is generated from k^{t-1} (2), [22]; c) conditioning by x_{t-1}^s and marginalisation over x_{t-2}^u. ■

This explicates *DM elements of FPD*, see Tab. 1. They *are probability densities*. Props. 8, 9 provide the optimal strategy s^{opt} and the needed estimator e. The remaining pds are gained by knowledge and preference elicitation. DM feasibility adds the need for systematic approximations. Sec. 4 solves, via FPD, these supporting DM tasks.

Table 1: DM Elements in FPD

DM Element	**PD: named f in Sec. 4**	**Gained by**
environment model prior pd	$(\mathsf{m}(x_t\|a, x_{t-1}))_{t\in\boldsymbol{t}}$ $\mathsf{e}(x_1\|a_1, k_0)$	knowledge elicitation and approximation
optimal strategy	$(\mathsf{s}^{opt}(a_t\|k^{t-1}))_{t\in\boldsymbol{t}}$	FPD, Prop. 8, and approximation
estimator	$\mathsf{e}(x_t, x_{t-1}^u\|a_t, k^{t-1})$ $\forall t \in \boldsymbol{t}$	filtering, Prop. 9, and approximation
ideal strategy and environment model	$(\mathsf{s}^{\mathsf{i}}(a_t\|k^{t-1}))_{t\in\boldsymbol{t}}$ $(\mathsf{m}^{\mathsf{i}}(x_t\|a, x_{t-1}))_{t\in\boldsymbol{t}}$	preference elicitation and approximation

4. Supporting DM Tasks

Tab. 1 lists the approximation, the knowledge and preference elicitation as the basic supporting DM tasks. Their solutions rely on (7) assuming $|\boldsymbol{b}| < \infty$, possibly with $|\boldsymbol{b}| \to \infty$. Static FPDs with the unit decision horizon suffice the supporting DM tasks. They use capital counterparts of DM elements of the supported DM task and operate on

$$\begin{aligned}
B &= (X, A) = (\text{an unseen state, an action}) \\
K &= \text{an explicitly expressed prior knowledge.}
\end{aligned} \tag{36}$$

Pds $\mathsf{f} \in \mathbf{f}$, related to supported DM tasks, concern *behaviour fragments*, $f \in \boldsymbol{f}$, of the supported DM task. Such pds f are referred as *fragmental pd*.

4.1. Approximation of a Known PD

Tab. 1 stresses ubiquity of approximation of a pd f, acting on a behaviour fragment $f \in \boldsymbol{f}$ of the supported DM task. The approximation, solved in [16], deals with $B = (X, A)$

$$\begin{aligned} X &= f = \text{an unseen behaviour fragment in } \boldsymbol{f} \\ A &= \hat{\mathsf{f}} = \text{a feasible approximation of } \mathsf{f} \text{ opted in } \hat{\mathbf{f}} \\ K &= \mathsf{f} = \text{a known pd on } \boldsymbol{f} \text{ to be approximated.} \end{aligned} \tag{37}$$

It has the closed-loop model (the knowledge K is explicit)

$$\mathsf{C}^{\mathsf{S}}(B|K) = \mathsf{C}^{\mathsf{S}}(X|A,K)\mathsf{C}^{\mathsf{S}}(A|K) = \mathsf{f}(f)\mathsf{S}(\hat{\mathsf{f}}|\mathsf{f}). \tag{38}$$

It follows from the chain rule and the assumption that the pd f models $f \in \boldsymbol{f}$. Its approximation $\hat{\mathsf{f}}$ is the action, with no influence on f. The optimised decision rule S chooses $\hat{\mathsf{f}} \in \hat{\mathbf{f}}$. The wish to select the pd $\hat{\mathsf{f}} \in \hat{\mathbf{f}}$, which models well $f \in \boldsymbol{f}$, and LFO (31) with $B^r = A$ determine the ideal pd

$$\mathsf{C}^{\mathsf{I}}(B|K) = \mathsf{C}^{\mathsf{I}}(X|A,K)\mathsf{C}^{\mathsf{I}}(A|K) = \hat{\mathsf{f}}(f)\mathsf{S}(\hat{\mathsf{f}}|\mathsf{f}). \tag{39}$$

An independent justification of Prop. 10 in [2] supports (7).

Proposition 10 (Approximation of a Known PD). *The FPD-optimal decision rule for the closed-loop model (38) and the ideal pd (39) is deterministic. It selects the approximating pd* $\hat{\mathsf{f}}^{opt}$ *via* the approximation principle

$$\hat{\mathsf{f}}^{opt} \in \operatorname{Arg}\min_{\hat{\mathsf{f}} \in \hat{\mathbf{f}}} \mathcal{D}(\mathsf{f}||\hat{\mathsf{f}}). \tag{40}$$

Proof For C^{S} (38), C^{I} (39), the CE-optimisation over S reads

$$\begin{aligned} \mathsf{S}^{opt} &\in \operatorname{Arg}\min_{\mathsf{S} \in \mathbf{S}} \int_{\hat{\mathbf{f}} \times \boldsymbol{f}} \mathsf{f}(f)\mathsf{S}(\hat{\mathsf{f}}|\mathsf{f}) \ln\left(\frac{\mathsf{f}(f)}{\hat{\mathsf{f}}(f)}\right) \, \mathrm{d}f \, \mathrm{d}\hat{\mathsf{f}} \\ &= \operatorname{Arg}\min_{\mathsf{S} \in \mathbf{S}} \int_{\hat{\mathbf{f}}} \mathsf{S}(\hat{\mathsf{f}}|\mathsf{f})\mathcal{D}(\mathsf{f}||\hat{\mathsf{f}}) \, \mathrm{d}\hat{\mathsf{f}} \\ &= \text{the determistic rule concentrated on } \hat{\mathsf{f}}^{opt} \text{ (40).} \end{aligned}$$

■

4.2. Minimum Cross-Entropy Principle

An analogy of the approximation task is solved here. The pd f modelling $f \in \boldsymbol{f}$ is, however, specified incompletely. Its complete specification is needed. The behaviour B and knowledge K, relevant to this supporting DM [16], read

$$\begin{aligned} X &= f = \text{an unseen behaviour fragment in } \boldsymbol{f} \\ A &= \mathsf{f} \text{ the opted pd in the set of pds } \mathbf{f} \\ K &= \mathsf{f}_0 \text{ a prior guess of the pd } \mathsf{f} \text{ and the set } \mathbf{f}. \end{aligned} \tag{41}$$

The corresponding closed-loop model is

$$\mathsf{C}^{\mathsf{S}}(B|K) = \mathsf{C}^{\mathsf{S}}(X|A,K)\mathsf{C}^{\mathsf{S}}(A|K) = \mathsf{f}(f)\mathsf{S}(\mathsf{f}|K). \tag{42}$$

This follows from the chain rule and the assumption that the pd f models $f \in \boldsymbol{f}$. The optimised decision rule is S. The guess f_0 a priori models $f \in \boldsymbol{f}$. This and LFO applied to actions A (31) give

$$\mathsf{C}^{\mathsf{I}}(B|K) = \mathsf{C}^{\mathsf{I}}(X|A,K)\mathsf{C}^{\mathsf{I}}(A|K) = \mathsf{f}_0(f)\mathsf{S}(\mathsf{f}|K). \tag{43}$$

It directly leads to the result derived in [27]. This again supports (7).

Proposition 11 (Minimum CE Principle). *The FPD-optimal decision rule for* C^{S} *(42) and* C^{I} *(43) is deterministic. It optimally completes an incomplete specification of the fragmental pd to* f^{opt} *via* the minimum CE principle

$$\mathsf{f}^{opt} \in \operatorname{Arg}\min_{\mathsf{f} \in \mathbf{f}} \mathcal{D}(\mathsf{f}||\mathsf{f}_0). \tag{44}$$

4.3. General Minimum CE Principle

The minimum CE principle is here generalised as in [16]. The pd $\mathsf{f} \in \mathbf{f}$ of $f \in \boldsymbol{f}$ is uncertain and *the model of* f *is required*. The relevant behaviour B and knowledge K are

$$\begin{aligned}
X &= (f, \mathsf{f}) = (\text{an unseen behaviour fragment, its model}) \\
A &= \mathsf{A}(\mathsf{f}|K) \text{ the opted pd on fragmental pds } \mathsf{f} \in \mathbf{f} \\
K &= \mathsf{A}_0(\mathsf{f}|K) \text{ a prior guess of the optimal action } \mathsf{A}^{opt}(\mathsf{f}|K).
\end{aligned} \tag{45}$$

The corresponding closed-loop model is

$$\mathsf{C}^{\mathsf{S}}(B|K) = \mathsf{C}^{\mathsf{S}}(X|A,K)\mathsf{C}^{\mathsf{S}}(A|K) = \mathsf{f}(f)\mathsf{A}(f|K)\mathsf{S}(\mathsf{A}|K). \tag{46}$$

It follows from the chain rule, the assumptions that f models f, A models f and S is the optimised decision rule.

The guess A_0 a priori models $\mathsf{f} \in \mathbf{f}$. This and LFO applied to A (31) give

$$\mathsf{C}^{\mathsf{I}}(B|K) = \mathsf{C}^{\mathsf{I}}(X|A,K)\mathsf{C}^{\mathsf{I}}(A|K) = \mathsf{f}(f)\mathsf{A}_0(\mathsf{f}|K)\mathsf{S}(\mathsf{A}|K). \tag{47}$$

Proposition 12 (General Minimum CE Principle). *The FPD-optimal decision rule for the closed-loop model (46) and the ideal pd (47) is deterministic. It selects* $\mathsf{A}^{opt}(\mathsf{f}|K)$*, describing the unknown pd* f*, via* the general minimum CE principle

$$\mathsf{A}^{opt} \in \operatorname{Arg}\min_{\mathsf{A} \in \mathbf{A}} \mathcal{D}(\mathsf{A}||\mathsf{A}_0). \tag{48}$$

5. On Use of the Solved Supporting DM Tasks

This section illustrates the use of the solved supporting tasks without the ambition to cover their full applicability.

5.1. Approximation in Kalman Filtering

Computational feasibility motivates an approximation. Approximate Kalman filtering (KF) operating on real vectors $x_t = (x_t^s, x_t^u)$ (21) is the widespread case of this type. It approximates the estimator $\mathsf{e}(x_t, x_{t-1}^u|a_t, k^{t-1})$ (35) by the normal pd. The approximation principle (40) provides the approximator.

Proposition 13 (Approximate Kalman Filter). *Let us consider the Bayesian filtering at time* $t \in \boldsymbol{t}$. *The environment model* $\mathsf{m}(x_t|a_t, x_{t-1})$ *and the pd* $\mathsf{e}(x_{t-1}, x^u_{t-2}|a_{t-1}, k_{t-2})$ *are given. The best normal approximation, Prop. 10,* $\hat{\mathsf{e}}^{opt}(x_t, x^u_{t-1}|a_t, k^{t-1})$ *of the pd* $\mathsf{e}(x_t, x^u_{t-1}|a_t, k^{t-1})$*, gained by (35), preserves its 1st and 2nd moments.* ■

Prop. 13 recommends a simple moment fitting *unconditional with respect to* x^u_{t-1}. This disqualifies techniques like extended KF and supports unscented KF-type algorithms.

Prop. 13 serves well if the prior pd $\mathsf{e}(x_{t-1}, x^u_{t-2}|a_{t-1}, k_{t-2})$ is known. If an approximation replaces the prior pd the approximation errors may accumulate. It is critical when the unseen x^u is time-invariant, when an estimation is addressed. Due to previous approximations, the approximated pd $\mathsf{e}(x_t, x^u_{t-1}|a_t, k_{t-1})$ is unknown and the minimum CE principle is relevant. It leads to estimation with forgetting [12].

5.2. Knowledge Elicitation

Knowledge elicitation leads to the environment model $(\mathsf{m}(x_t|a_t, x_{t-1}))_{t\in\boldsymbol{t}}$ and the prior pd $\mathsf{e}(x_1|a_1, k_0)$. The minimum CE principles, Props. 11, 12, support it.

Proposition 14 (Completion of Deterministic Models). *Let a domain theory (say, 1st principles of physics) provide a multivariate deterministic mapping* $\mathsf{d} : \boldsymbol{a}_t, \boldsymbol{x}_{t-1} \to \boldsymbol{x}_t$*, which approximately relates triples* (x_t, a_t, x_{t-1})

$$\Delta_t = x_t - \mathsf{d}(a_t, x_{t-1}) \approx 0 \text{ seen as } \mathcal{E}[\Delta_t|a_t, x_{t-1}] = 0. \tag{49}$$

Let a rough qualitative model $(\mathsf{m}_0(x_t|a_t, x_{t-1}))_{t\in\boldsymbol{t}}$ *be chosen*[7]. *Then, the minimum CE principle, Prop. 11, and the constrained optimisation recommend the environment model*

$$\mathsf{m}(x_t|a_t, x_{t-1}) \propto \mathsf{m}_0(x_t|a_t, x_{t-1}) \exp\Big(\sum_{i\in\boldsymbol{i}} \lambda_i \Delta_{t;i}\Big). \tag{50}$$

Lagrangian multipliers $(\lambda_i = \lambda_i(a_t, x_{t-1}))_{i\in\boldsymbol{i}}$ *solve, cf. (49),*

$$\int_{\boldsymbol{x}_t} x_t \mathsf{m}_0(x_t|a_t, x_{t-1}) \exp\Big(\sum_{i\in\boldsymbol{i}} \lambda_i x_{t;i}\Big)\, \mathrm{d}x_t = \mathsf{d}(a_t, x_{t-1}). \qquad ■$$

Usually, the elicitation processes several knowledge pieces, [14]. Tab. 2 provides some examples.

A combination of pds modelling different behaviour fragments $f \in \boldsymbol{f}$ requires their extension. Each processed pd f is to be extended to a pd c of behaviours $b = (f, f^r) \in \boldsymbol{b} = \boldsymbol{f} \times \boldsymbol{f}^r$, $\boldsymbol{f}^r \neq \emptyset$. This induces the supporting DM with

$$\begin{aligned} X &= (b, \mathsf{c}(f^r|f)) = (\text{an unseen behaviour, pd on the rest } \boldsymbol{f}^r) \\ A &= \mathsf{c}(b),\ b \in \boldsymbol{b}, = \text{the opted behaviour model} \\ K &= \begin{cases} \mathsf{f} & = \text{a given pd on the behaviour fragment } f \in \boldsymbol{f} \\ \mathsf{c}_0 & = \text{a prior description of the behaviour } b \in \boldsymbol{b}. \end{cases} \end{aligned} \tag{51}$$

The next proposition applies the minimum CE principle, Prop. 11, to the extension problem structured according to (51).

[7] It typically delimits the domain $(\boldsymbol{a}, \boldsymbol{x})$ and the range $\boldsymbol{x}$ of the mapping d.

Table 2: Some Knowledge Pieces

Description	Definition	Fragmental PD
a data sample	(x_t, a_t, x_{t-1})	Dirac on the data sample
a function $\mathsf{d}(x_t, a_t) \approx 0$	$\mathcal{E}[\mathsf{d}(x_t, a_t)] = 0$	$\mathsf{f}(x_t, a_t)$
a fuzzy rule	if x_{t-1} then the membership of x_t is $\mathsf{d}_{x_{t-1}}(x_t)$	$\mathsf{f}(x_t\|x_{t-1})$ $\propto \mathsf{d}_{x_{t-1}}(x_t)$

Proposition 15 (Extension of a Fragmental PD). *The pd* $\mathsf{c}(b) = \mathsf{c}_0(b)\mathsf{f}(f)/\mathsf{c}_0(f)$ *extends the given pd* f *on* $\boldsymbol{f}$ *to* $\boldsymbol{b}$. ■

Sec. 5.6 addresses the potentially critical choice of the prior guess $\mathsf{c}_0(f|f^r) = \mathsf{c}_0(f, f^r)/\mathsf{c}_0(f)$.

Props. 14, 15 process an incomplete but granted knowledge. This is unrealistic when addressing the supporting task combining several pds into single pd c. The combined pds need not be mutually absolutely continuous (compatible), [24]. This fact drives our formulation and solution of this classical pds' *pooling*, see the next section. The solution serves both to knowledge and preference elicitation. The choice of the prior pd $\mathsf{e}(x_1|a_1, k_0)$ is then a special case of the general solution requiring extensions only to $f = (x_1, a_1, k_0)$, [14].

5.3. Pooling of PDs

The supporting pooling DM task combines given pds $(\mathsf{c}_j(b))_{j\in\boldsymbol{j}}$, $b \in \boldsymbol{b}$, $\boldsymbol{j} = \{1, \ldots, |\boldsymbol{j}|\}$, $|\boldsymbol{j}| < \infty$. Prop. 15 makes the used assumption that the combined pds $(\mathsf{c}_j)_{j\in\boldsymbol{j}}$ operate on a common behaviour $b \in \boldsymbol{b}$ unrestrictive.

Behaviour B and knowledge K of this DM are

$$\begin{aligned}
&X = (b, \mathsf{c}) = (\text{an unseen behaviour in } \boldsymbol{b}, \text{its model in } \boldsymbol{\mathsf{c}}) \\
&A = \mathsf{A}(\mathsf{c}|K) \text{ the opted pd on } \mathsf{c} \in \boldsymbol{\mathsf{c}} \text{ modelling } b \in \boldsymbol{b} \qquad (52)\\
&K = \mathsf{A}_0(\mathsf{c}|K) \text{ a prior guess of the optimal } \mathsf{A}^{opt}(\mathsf{c}|K) \\
&\text{supp}[A_0] = \big\{\mathsf{c} \text{ is a pd on } \boldsymbol{b} : \ (\mathcal{E}[\mathcal{D}(\mathsf{c}_j||\mathsf{c})|A, K] \le \nu_j)_{j\in\boldsymbol{j}}\big\}, \ \ (\nu_j > 0)_{j\in\boldsymbol{j}} \text{ given.}
\end{aligned}$$

Prop. 10 determines the adequate proximity measure and motivates the knowledge specification used in (52), which otherwise copies (45). The specification demands the probable pd c to approximate well the given pds $(\mathsf{c}_j)_{j\in\boldsymbol{j}}$.

The optional non-negative bounds $\nu = (\nu_j)_{j\in\boldsymbol{j}}$ should be as small as possible while keeping supp$[A_0]$ non-empty. This choice copes with incompatibility of the combined pds $(\mathsf{c}_j)_{j\in\boldsymbol{j}}$. It, however, requires the multi-valued minimisation over $\nu \ge 0$. This minimisation is converted into the scalar minimisation guaranteeing Pareto's optimality

$$\nu^{opt} \in \operatorname{Arg}\min_{\nu\in\boldsymbol{\nu}} \sum_{j\in\boldsymbol{j}} \alpha_j \nu_j. \qquad (53)$$

The optional probabilistic weights $\alpha = (\alpha_j)_{j\in\boldsymbol{j}}$ select the specific point ν^{opt} on Pareto's frontier. A discussion of their choice is postponed to Sec. 5.6.

The need for supp$[A_0] \neq \emptyset$ and the wish to make all constraints in (52) active delimit the set $\boldsymbol{\nu}$ of bounds in (52).

Tab. 3 summarises symbols used in Prop. 16 below.

Table 3: Symbols Used in Prop. 16

Definition	Comment
$\bar{\mathsf{c}} = \sum_{j\in\boldsymbol{j}} \alpha_j \mathsf{c}_j$	see (52), (53)
$\mathsf{A} = \mathsf{Di}(\tau) \propto \prod_{b\in\boldsymbol{b}} \mathsf{c}(b)^{\tau(b)-1},\ \tau(b) > 0$	Dirichlet's pd on the joint pd $\mathbf{c}$
opted $\tau_0(b) = \lambda_0 \mathsf{c}_0(b),\ \lambda_0 = \sum_{b\in\boldsymbol{b}} \tau_0(b)$	prior pd is $\mathsf{Di}(\tau_0)$
$\Psi(z) = \frac{\mathrm{d}\ln\left(\int_0^\infty y^{z-1}\exp(-y)\mathrm{d}y\right)}{\mathrm{d}z},\ z > 0$	digamma function

Proposition 16 (Pooling of Behaviour Descriptions). *The solution of the general minimum CE principle (48), with $\mathsf{A}_0 = \mathsf{Di}(\tau_0 = \lambda_0 \mathsf{c}_0)$ and knowledge K (52) for $\nu = \nu^{opt}$ (53), is Dirichlet's pd $\mathsf{A}^\lambda = \mathsf{Di}(\tau^\lambda)$, Tab. 3. Its parameter is*

$$\tau^\lambda = (\tau^\lambda(b))_{b\in\boldsymbol{b}},\ \tau^\lambda(b) = \sum_{j\in\boldsymbol{j}} \lambda_j \mathsf{c}_j(b) + \tau_0(b) = \sum_{j\in\boldsymbol{j}\cup\{0\}} \lambda_j \mathsf{c}_j(b). \tag{54}$$

The optimal pd $\mathsf{A}^{opt} = \mathsf{A}^{\lambda^{opt}}$ models the promising pooled pd $\mathsf{c} \in \mathbf{c}$ of $(\mathsf{c}_j)_{j\in\boldsymbol{j}}$. Its parameter is

$$\lambda^{opt} \in \mathrm{Arg}\max_{\lambda>0} \Big[\int_{b\in\boldsymbol{b}} \bar{\mathsf{c}}(b) \Psi\big(\tau^\lambda(b)\big)\ \mathrm{d}b - \Psi\big(\sum_{j\in\boldsymbol{j}\cup\{0\}} \lambda_j \big) \Big]. \tag{55}$$

Proof The minimisation of the Kuhn-Tucker functional, given by multipliers $\lambda = (\lambda_j)_{j\in\boldsymbol{j}}$, respects K-constraints (52) in the general minimum CE principle (48). For a conjugated Dirichlet's prior $\mathsf{A}_0 = \mathsf{Di}(\tau_0)$, the minimisation reads

$$\begin{aligned}
\mathsf{A}^\lambda \quad &\in \quad \mathrm{Arg}\min_{\mathsf{A}\in\mathbf{A}} \mathcal{D}(\mathsf{A}||\mathsf{A}_0) + \sum_{j\in\boldsymbol{j}} \lambda_j \mathcal{E}[\mathsf{D}(\mathsf{c}_j||\mathsf{c})|\mathsf{A}, K] \\
&= \mathrm{Arg}\min_{\mathsf{A}\in\mathbf{A}} \int_{\mathbf{c}} \mathsf{A}(\mathsf{c}) \Big[\ln\left(\frac{\mathsf{A}(\mathsf{c})}{\mathsf{A}_0(\mathsf{c})}\right) - \sum_{\boldsymbol{b}} \sum_{j\in\boldsymbol{j}} \lambda_j \mathsf{c}_j(b) \ln(\mathsf{c}(b)) \Big] \mathrm{d}\mathsf{c} \\
&= \quad \mathrm{Arg}\min_{\mathsf{A}\in\mathbf{A}} \mathcal{D}(\mathsf{A}||\mathsf{A}^\lambda) \text{ with } \tau^\lambda \text{ given by (54).}
\end{aligned}$$

The choice of ν activating bounds in (52) makes the Kuhn-Tucker multipliers λ positive and relates them to ν

$$\nu_j = \mathcal{E}[\mathsf{D}(\mathsf{c}_j||\mathsf{c})|\mathsf{A}^\lambda, K],\ j \in \boldsymbol{j}.$$

This reduces (53) to minimisation

$$\begin{aligned}
\lambda^{opt} \quad &\in \quad \mathrm{Arg}\min_{\lambda>0} \sum_{j\in\boldsymbol{j}} \alpha_j \mathcal{E}[\mathsf{D}(\mathsf{c}_j||\mathsf{c})|\mathsf{A}^\lambda, K] \\
&= \quad \mathrm{Arg}\max_{\lambda>0} \sum_{\boldsymbol{b}} \bar{\mathsf{c}}(b) \mathcal{E}[\ln(\mathsf{c}(b))|\mathsf{A}^\lambda, K].
\end{aligned} \tag{56}$$

For the relevant $\mathsf{A}^\lambda = \mathsf{Di}(\tau^\lambda)$, it holds, (54), (56),

$$\begin{aligned}
&\mathcal{E}[\ln(\mathsf{c}(b))|\mathsf{A}^\lambda, K] = \Psi\big(\tau^\lambda(b)\big) - \Psi\big(\sum_{b\in\boldsymbol{b}} \tau^\lambda(b) \big) \\
\Rightarrow \quad &\lambda^{opt} \in \mathrm{Arg}\max_{\lambda>0} \sum_{b\in\boldsymbol{b}} \bar{\mathsf{c}}(b) \Big[\Psi\Big(\sum_{j\in\boldsymbol{j}\cup\{0\}} \lambda_j \mathsf{c}_j(b) \Big) - \Psi\Big(\sum_{j\in\boldsymbol{j}\cup\{0\}} \lambda_j \Big) \Big].
\end{aligned}$$

This proves (55) intentionally written in the integral form, cf. conjecture (7). ∎

5.4. Preference Elicitation

The probabilistic description of DM aims and constraints is one of the key advantages brought by FPD. Having elicited fragmental descriptions of the ideal pd, Props. 15, 16 guide how to extend and pool them. Thus, it remains to show how to construct these fragmental ideal pds in typical situations. Their samples are in Tab. 4.

Table 4: Some Descriptions of DM Aims

Definition given by	**The ideal pd on $f \in \boldsymbol{f}$**
the upper hierarchic level	relevant to the upper hierarchic level
the neighbour, Sec. 5.5	relevant to the agent's neighbour
the desired value of an attribute entering behaviour	see (57)
the set of possible actions	see (60)

The elicitation of the ideal pd is illustrated on the regulation problem [3]. It pushes the state x_t to a given ideal state x_t^{i}. With the ideal decision rule (33), it should hold

$$\begin{aligned} x_t^{\mathsf{i}} &\in \operatorname{Arg}\max_{x\in\boldsymbol{x}} \mathsf{m}^{\mathsf{i}}(x_t = x|x_{t-1}) \\ &= \operatorname{Arg}\max_{x\in\boldsymbol{x}} \int_{\boldsymbol{a}} \mathsf{m}(x|a_t = a, x_{t-1})\mathsf{r}^{\mathsf{i}}(a_t = a|x_{t-1}) \; \mathrm{d}a. \end{aligned} \tag{57}$$

Ideally, no possible action should be a priori excluded

$$\operatorname{supp}[\mathsf{r}^{\mathsf{i}}(\bullet|x_{t-1})] \supseteq \operatorname{supp}[\mathsf{m}(x_t = x^{\mathsf{i}}|a_t = \bullet, x_{t-1})]. \tag{58}$$

For $|\boldsymbol{b}| < \infty$, $||\mathsf{r}^{\mathsf{i}}(\bullet|x_{t-1})||_p = [\int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a|x_{t-1})^p \; \mathrm{d}a]^{1/p}|| < \infty \; \forall p > 1$ and $||\mathsf{m}(x_t^{\mathsf{i}}|\bullet, x_{t-1})||_q < \infty$, $q^{-1} + p^{-1} = 1$. Hölder's inequality [24] and (57) imply

$$\mathsf{m}^{\mathsf{i}}(x_t^{\mathsf{i}}|x_{t-1}) \le ||\mathsf{m}(x_t^{\mathsf{i}}|\bullet, x_{t-1})||_q ||\mathsf{r}^{\mathsf{i}}(\bullet|x_{t-1})||_p, \tag{59}$$

with equality reached for

$$\mathsf{r}^{\mathsf{i}}(a_t|x_{t-1}) \propto [\mathsf{m}(x_t^{\mathsf{i}}|a_t, x_{t-1})]^{q-1}, \tag{60}$$

which meets (58). This gives fragmental ideal pds reflecting the regulation problem. The limited space allows us just to add some comments on this open-ended problem: a) the conjecture (7) is valid here; b) the case $p = 1$ leads to CE-type treatment; c) LFO (31) applies when only a part of x_t matters; d) end-points of possible ranges form x_t^{i} when its individual entries (attributes) should be maximised; e) the construction of the ideal environment model m^{i} (57) from the environment model m cares about the best *potentially reachable* DM quality; f) (58) is vital for guaranteeing the explorative nature of the DM strategy.

5.5. Cooperation in a Flat Manner

FPD has allowed to develop a powerful concept of interactions of selfish imperfect but wise agents, [18]. The adjective "selfish" means that the agent follows its "personal" aims. The adjective "imperfect" labels the agent's limited knowledge, observation, evaluation, and acting abilities. Such an agent acts in the environment containing other imperfect selfish agents. The "wise" agent knows that others influence its success in reaching of its aims. It enhances its personal chances by (partially) publishing the information it deals. The published information is expressed in probabilistic manner understandable by any FPD-using agent. This allows other agents to modify their strategy. The individual modifications diminish clashes'probability and increase the chances of agent's to reach *its unchanged individual aims.*

An imperfect agent can reach information provided by a small number of other agents, its recognisable neighbours. Thus, it can extend and pool handled pds, Props. 15, 16, even with its limited resources. Consequently, no mediating or facilitating agent, the bottleneck of standard schemes, is needed. This makes the outlined cooperation fully scalable. Moreover, a) the common aim of respective agents make them cooperate; b) this flat cooperation imitates acting of complex well-surviving societies; c) misleading information from neighbours may spoil the reaching of personal aims. It can, however, be numerically judged and used for decreasing trust (α_j in (53)) to such adversaries.

5.6. Choice of Optional Parameters in Extension and Pooling

The extension and pooling serving to knowledge and preference elicitation depend on: 1) the optional prior guess $\mathsf{c}_0(b)$, $b \in \boldsymbol{b}$, (51) of the joint pd; 2) the parameter τ_0 determining the prior pd, see Tab. 3, and 3) the probabilistic weights α (53) selecting the point on Pareto's frontier.

Their universally applicable choices are presented here.

ad 1) The optimal pooling decision rule $\mathsf{A}^{\lambda^{opt}} = \mathsf{Di}(\tau^{\lambda^{opt}})$, Prop. 16, has the expectation

$$\mathsf{c}^{\lambda^{opt}}(b) = \frac{\tau^{\lambda^{opt}(b)}}{\sum_{j \in \boldsymbol{j} \cup \{0\}} \lambda_j^{opt}}. \tag{61}$$

It is an optimal point estimate of unknown $\mathsf{c}(b)$ and as such it offers the inherently implicit choice, cf. [17],

$$\mathsf{c}_0(b) = \mathsf{c}^{\lambda^{opt}}. \tag{62}$$

Existence of its solution is conjectured and verified in particular cases. Successive iterations offer finding it.

ad 2) In a particular case $\mathsf{c}_j = \mathsf{c}$, $\forall j \in \boldsymbol{j}$, it is "natural" to require the expectation (61) $\mathsf{c}^{\lambda^{opt}} = \mathsf{c}$. This can be reached iff $\tau_0 \to 0^+ \Leftrightarrow \lambda_0 \to 0^+$.

ad 3) The weights α reflect either trust into or importance of respective sources. As such, they can be learnt by observing the past performance of the knowledge source. It is straightforward but waits for an elaboration.

6. Concluding Remarks

The presented theory is relatively matured but still there is a range of open problems. Their list, mixing simple and difficult questions, should ideally stimulate readers to inspect them.

✓ When the general minimum CE principle reduces to minimum CE principle?

✓ How to formulate and solve FDP with discounting and how to choose data-dependent discounting factor?

✓ Does exist a systematic approach to prove (7)?

✓ Is it possible to relax the assumption that the behaviour set is specified beforehand? The knowledge transfer via predictors [23] seems to suit to this.

✓ How to merge systematically complexity aspects into the problem formulation and solution? This would lead to truly universal artificial intelligence.

✓ How to approach prediction of emerging behaviours in a systematic way overcoming current analysis of particular cases or relying on simulations [1]?

✓ How to systematically elicit aims-related fragmental pds?

The FPD theory is ready for extensive applications but it needs a lot for technical work. For instance:

✓ The used optimisations call for algorithmic solutions.

✓ The trust estimation is to be elaborated in detail.

✓ The conceptual solution is to be converted into reliable algorithms for universally approximating black-box type models as mixture ratios.

✓ A practical experience is to be accumulated with simulations and applications in connection with industry 4.0 [21] or various cyber-physical and social systems.

The next list gives the reasons why a DM expert could care about FPD. It

- provides a unified theory extending SEU, Props. 6, 7;
- finds minimising strategy explicitly even in general setting, which makes approximate dynamic programming simpler;
- allows to address hard DM problems [5];
- has approximation [2] and generalisation of minimum KL principles [27] as simple consequences [16, 23];
- puts KL control [9, 10] into a wider perspective;
- feeds a proper exploration into adaptive DM;
- unifies otherwise disparate languages describing environment and DM aims and strengthen the deductive machinery of DM with multiple aims;
- quantifies DM aims by the ideal probability: this allows to employ estimation and approximation for its construction;
- reveals that any realistic aim quantification is to respect the environment model (57), [11], and it allows to adapt the performance index;
- converts cooperation of agents into the pooling problem of understandable shared fragmental pds [18].

Acknowledgements This research has been supported by MŠMT LTC18075.

7. Appendix: A Way to Quantum DM

The mapping b^{s} (10), reflecting uncertainties, makes the quantification $\mathsf{q}^{\boldsymbol{b}}$ the function $\mathsf{q}^{\mathsf{s}\boldsymbol{n}} = \mathsf{q}^{\boldsymbol{b}} \circ \mathsf{b}^{\mathsf{s}} : \boldsymbol{n} \to \mathbb{R}$.

Assumption 9 (No Hidden Feedback of s to u). *Within the addressed DM, all known dependencies of the behaviour on the considered strategies are exploited. It means that the influences of strategy and uncertainty on the closed-loop behaviour are taken as independent.* ■

The conjecture (7) allows to consider a finite amount of strategies $|\boldsymbol{b}| < \infty$ and a finite amount of behaviours $|\boldsymbol{s}| < \infty$. In conjunction with (7), it suffices the select a finite number $|\boldsymbol{n}|$ of distinguishable uncertainties. This makes $\mathsf{q}^{\mathsf{s}\boldsymbol{n}}$ real-valued $(|\mathsf{s}|, |\boldsymbol{n}|)$ matrix. It can always be decomposed

$$\mathsf{q}^{\mathsf{s}\boldsymbol{n}} = \mathsf{q}^{\mathsf{s}\boldsymbol{\psi}}\mathsf{q}^{\boldsymbol{\psi}\boldsymbol{n}}. \tag{63}$$

There $\boldsymbol{\psi} = \{1, \dots, |\boldsymbol{\psi}| = |\boldsymbol{n}|\}$, $\mathsf{q}^{\mathsf{s}\boldsymbol{\psi}}$ is $(|\mathsf{s}|, |\boldsymbol{\psi}|)$ matrix and $\mathsf{q}^{\boldsymbol{\psi}\boldsymbol{n}}$ is unitary matrix. Both factors may have complex entries. The resulting $\mathsf{q}^{\mathsf{s}\boldsymbol{n}}$ is real-valued. The factor $\mathsf{q}^{\mathsf{s}\boldsymbol{\psi}}$ comprises influence of strategies and the factor $\mathsf{q}^{\boldsymbol{\psi}\boldsymbol{n}}$ the independent influence of uncertainties. Rows of the matrix $\mathsf{q}^{\boldsymbol{\psi}\boldsymbol{n}}$ span the Hilbert's space of wave functions. Invariance of (63) with respect the right unitary transformation of $\mathsf{q}^{\boldsymbol{\psi}\boldsymbol{n}}$, the isometry interpretation of the uncertainty influence, together with the fundamental Gleason's expression of all measures on Hilbert's spaces of dimension greater than 2, [7], make quantum probability calculus relevant to DM. There is a strong evidence that it improves models of human DM [4].

References

[1] S. Abar, G.K. Theodoropoulos, P. Lemarinier, and G.M.P. O'Hare. Agent based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review*, 24 (2017) 13 – 33.

[2] J.M. Bernardo. Expected information as expected utility. *The An. of Stat.*, 7(3) (1979) 686–690.

[3] D.P. Bertsekas. *Dynamic Programming and Optimal Control.* Athena Sci., (2001).

[4] J.R. Busemeyer and P.D. Bruza. *Quantum Models of Cognition and Decision.* Cambridge University Press, New York, (2012).

[5] A. Bušič and S. Meyn. Action-constrained Markov decision processes with Kullback-Leibler cost. In S. Bubeck, V. Perchet, and P. Rigollet, editors, *Proc. of Machine Learning Research*, volume 75, pages 1–14. 2018.

[6] G. Debreu. Representation of a preference ordering by a numerical function. In R.M. Thrall, C.H. Coombs, and R.L. Davis, editors, *Decision Processes*, NY, 1954. Wiley.

[7] A. Dvurečenskij. *Gleasons Theorem and Its Applications*, volume 60 of *Mathematics and Its Applications.* Kluwer Academic Publishers, Dordrecht/Boston/London, (1993).

[8] A.A. Feldbaum. Theory of dual control. *Autom. Remote Control*, 21-22(9-2)(1960-61).

[9] P. Guan, M. Raginsky, and R.M. Willett. Online Markov decision processes with Kullback Leibler control cost. *IEEE Trans. on AC*, 59(6) (2014) 1423–1438.

[10] H.J. Kappen. Linear theory for control of nonlinear stochastic systems. *Physical review letters*, 95(20) (2005) 200201.

[11] M. Kárný. Towards fully probabilistic control design. *Automatica*, 32(12) (1996) 1719–1722.

[12] M. Kárný. Approximate Bayesian recursive estimation. *Inf. Sci.*, 289 (2014) 100–111.

[13] M. Kárný. Recursive estimation of high-order Markov chains: Approximation by finite mixtures. *Inf. Sci.*, 326 (2016) 188–201.

[14] M. Kárný, A. Bodini, T.V. Guy, J. Kracík, P. Nedoma, and F. Ruggeri. Fully probabilistic knowledge expression and incorporation. *Statistics and Its Interface*, 7(4) (2014) 503–515.

[15] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař. *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. Springer, (2006).

[16] M. Kárný and T.V. Guy. On support of imperfect Bayesian participants. In T.V. Guy and et al, editors, *Decision Making with Imperfect Decision Makers*, volume 28. Springer, Int. Syst. Ref. Lib., 2012.

[17] M. Kárný, T.V. Guy, A. Bodini, and F. Ruggeri. Cooperation via sharing of probabilistic information. *Int. J. of Computational Intelligence Studies*, pages 139–162.

[18] M. Kárný and R. Herzallah. Scalable harmonization of complex networks with local adaptive controllers. *IEEE Trans. on SMC: Systems*, 47(3) (2017) 394–404.

[19] M. Kárný and T. Kroupa. Axiomatisation of fully probabilistic design. *Inf. Sci.*, 186(1) (2012) 105–113.

[20] S. Kullback and R. Leibler. On information and sufficiency. *An. of Math. Stat.*, 22 (1951) 79–87.

[21] Y. Liao, F. Deschamps, E.F.R. Loures, and L.F.P. Ramos. Past, present and future of industry 4.0 – A systematic literature review and research agenda proposal. *Int. J. of Production Res.*, 55(12) (2017) 3609–3629.

[22] V. Peterka. Bayesian system identification. In P. Eykhoff, editor, *Trends and Progress in System Identification*, pages 239–304. Perg. Press, 1981.

[23] A. Quinn, M. Kárný, and T.V. Guy. Fully probabilistic design of hierarchical Bayesian models. *Inf. Sci.*, 369 (2016) 532–547.

[24] M.M. Rao. *Measure Theory and Integration.* John Wiley, NY, (1987).

[25] G.N. Saridis. An integrated theory of intelligent machines by expressing the control performance as entropy. *Control Th. & Advanced Technology*, (2) (1985) 125–138.

[26] L.J. Savage. *Foundations of Statistics.* Wiley, (1954).

[27] J. Shore and R. Johnson. Axiomatic derivation of the principle of maximum entropy & the principle of minimum cross-entropy. *IEEE Tran. on Inf. Th.*, 26(1) (1980) 26–37.

[28] C.A. Sims. Implications of rational inattention. *J. of Monetary Economics*, 50(3) (2003) 665 – 690.

[29] M. Ullrich. Optimum control of some stochastic systems. In *Preprints of the VIII-th conference ETAN.* Beograd, 1964.

[30] J. Wallenius, J.S. Dyer, P.C. Fishburn, R.E. Steuer, S. Zionts, and K. Deb. Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Sci.*, 54(7) (2008) 1336–1349.