Chapter 3
On the Origins of Imperfection and Apparent Non-Rationality

Miroslav Kárný and Tatiana V. Guy

Abstract

Decision making (DM) is a preferences-driven choice among available actions. Under uncertainty, Savage's axiomatisation singles out Bayesian DM as the adequate normative framework. It constructs strategies generating the optimal actions, while assuming that the decision maker rationally tries to meet her preferences.

Descriptive DM theories have observed numerous deviations of the real DM from normative recommendations. The explanation of decision-makers' imperfection or non-rationality, possibly followed by rectification, is the focal point of contemporary DM research. This chapter falls into this stream and claims that the neglecting a part of the behaviour of the closed DM loop is the major cause of these deviations. It inspects DM subtasks in which this claim matters and where its consideration may practically help. It deals with: i) the preference elicitation; ii) the "non-rationality" caused by the difference of preferences declared and preferences followed; iii) the choice of proximity measures in knowledge and preferences fusion; iv) ways to a systematic design of approximate DM; and v) the control of the deliberation effort spent on a DM task via sequential DM.

The extent of the above list indicates that the discussion offers more open questions than answers, however, their consideration is the key element of this chapter. Their presentation is an important chapter's ingredient.

Miroslav Kárný
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodáreno věží 4, 182 08 Prague 8, Czech Republic, e-mail: school@utia.cas.cz

Tatiana V. Guy
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodáreno věží 4, 182 08 Prague 8, Czech Republic, e-mail: guy@utia.cas.cz

3.1 Introduction

The chapter considers decision making (DM) as a direct choice among available actions, which is driven by the wish to meet DM preferences. The main concern of the reported research is the repeatedly observed discrepancies between real decision making and recommendations or predictions of normative theories. The discrepancies include: i) framing effect – for instance, the chosen action depends on whether identical DM consequences are presented as a gain or loss, [23]; ii) bounded rationality – for instance, the chosen action differs from the optimal one due to inherent constraints on the effort spent on solving a specific DM task, [31]; iii) violation of game-theory predictions – for instance, players often use apparently non-optimal strategies even in very simple games, [48,64]; iv) intransitivity of preferences – they violate logically appealing linear order, [66]; and many others.

Any theory is a meta-model of reality, while models are its inputs to applications. Insufficiency of the theory or of these inputs may result in unpredicted or bad outcomes. This chapter inspects DM theory from this perspective, providing a unifying view on the roots of the above-mentioned discrepancies.

DM concerns interactions of the decision maker with her environment. An environment[1] considered during DM is a part of the real world specified by the decision maker for each DM task. The decision maker expends her intellectual and technical resources to: i) delimit informally the addressed DM task; ii) select theoretical and technical tools for solving the DM task; iii) formalise the DM task and use the selected tools; iv) apply the resulting sequence of actions; v) handle (accumulate, aggregate, forget, etc.) the knowledge contained in the closed loop formed by the decision maker and her environment. Complexity of these activities causes some parts of the closed-loop behaviour are unintentionally neglected during DM. This chapter claims that such neglect is the dominating cause of the above-mentioned discrepancies. It shows that the consideration of the neglected parts has practically significant consequences and opens interesting research problems.

Control theory, pattern recognition, fault detection, medical and technologic diagnosis, machine learning, statistics, signal processing are examples of fields *de-facto* addressing DM. This indicates the broadness of the inspected topic and explains the proliferation of exploited formal tools as well as of the terminology.

Sections 3.1.1 and 3.1.2 recall the normative DM theory we rely on, namely, the fully probabilistic design (FPD) of decision strategies, [33, 41, 45]. Section 3.1.3 specifies the DM aspects discussed within this chapter. Then, Section 3.1.4 presents the layout of the remainder of the chapter.

[1] [68] calls it a small world. Alternative terms like system, plant, object are used.

3.1.1 Fully Probabilistic Design of Decision Strategies

The Bayesian framework confers a key normative theory of DM under uncertainty, [68]. Its internal consistency and advanced computing technology have made realistic the insightful prediction: "in 2020, we all will be Bayesians", [57].

Bayesian theory recommends a DM strategy that minimises an expected loss (or maximises an expected reward). The loss expresses preferences among possible closed-loop behaviours. Hereafter, the term behaviour means the collection of all actions, observed and considered but unobserved variables. The expectation is taken over all uncertain – yet or never unobserved – parts of the behaviour.

The decision maker expresses her wishes as preferences among possible behaviours. The loss quantifies the preferences among behaviours and expectation transforms them into quantified preferences among decision strategies, which map the decision-maker's knowledge onto actions. The strategy choice influences the strategy-dependent probability density (pd)[2] describing the closed-loop behaviour and determining the expectation. This pd is the basic formal object of the fully probabilistic design of DM strategies, [26, 33, 38, 40–42], that the chapter relies on.

FPD extends Bayesian DM theory by allowing the loss to depend on the strategy. FPD adopts the logarithmic score of the closed-loop-describing pd to its ideal counterpart as the universal loss, [45]. The ideal pd specified by the decision maker, acting also on possible behaviours, expresses the DM preferences among these. The value of the ideal pd is high for desirable behaviours and small for undesirable ones. For the loss constructed in this way, FPD minimises the Kullback-Leibler divergence (KLD, [54])[3] of these pds. Note that FPD formulations contains a dense set within the set of all Bayesian DM formulations, [45], and thus our discussion does not neglect any Bayesian DM task.

3.1.2 Formal Description of Fully Probabilistic Design

Agreement 1 (Fonts in Notation, Behaviour, Time) *The symbol $\boldsymbol{x}$ is the set of x values.* C*apitals in sans serif denote mappings and thus* $\boldsymbol{\mathsf{C}}$ *is the set of* C *instances.* $\mathscr{C}$*aligraphic letters are reserved for functionals.*

The dominating measure with respect to which a probability density (pd) is defined [65] is denoted $\mathrm{d}\bullet$*. Here, it is either the Lebesgue or counting measure.*

The closed-loop behaviour $b \in \boldsymbol{b}$ is formalised as a collection of random variables considered by the decision maker. They are ordered according to time, labelled by discrete time $t \in \boldsymbol{t} = \{1, 2, \ldots, T\}$, at which the optional actions $a_t \in \boldsymbol{a}$ are chosen. The time extent is delimited by a decision horizon $T \leq \infty$.

All functions having time-dependent arguments are generally time-dependent. Exceptions are explicitly pointed out.

[2] A pd is the Radon-Nikodým derivative of a probabilistic, randomness-modelling measure.

[3] The KLD has many names. Relative entropy and cross entropy [70] are the most common.

Throughout, the following partitioning of the behaviour, b, is exploited.

Agreement 2 (Time of the Action Choice Splits Behaviour) *When considering an action $a_t \in \boldsymbol{a}$ at time $t \in \boldsymbol{t}$, any possible behaviour $b \in \boldsymbol{b}$ splits into*

- *action $a_t \in \boldsymbol{a}$, which is chosen by the decision maker;*
- *knowledge k_{t-1} available for choosing a_t: it includes prior knowledge k_0;*
- *ignorance containing variables considered by the decision maker but unavailable for choosing action a_t, i.e. considered but never observed internal variables and yet unused observations o_τ, $\tau \geq t$, and actions a_τ, $\tau > t$.*

The term "internals" covers notions like hidden or latent variables, internal states, and an unknown (multivariate) parameter $\Theta \in \boldsymbol{\Theta}$, which is uninfluenced by the action and knowledge. This chapter predominantly considers parameter Θ as the only internal[4].

As time evolves, the knowledge is enriched by the observations $o_t \in \boldsymbol{o}$ and by the chosen action $a_t \in \boldsymbol{a}$, i.e. $\boldsymbol{k}_t = \boldsymbol{k}_{t-1} \cup \boldsymbol{o} \cup \boldsymbol{a}$, and the ignorance shrinks correspondingly. The unknown parameter is a permanent part of the ignorance.

Closed-loop behaviours $b \in \boldsymbol{b}$ are described by a closed-loop model, which is a pd $\mathsf{C}_\mathsf{S}(b)$ on $\boldsymbol{b}$ depending on the inspected strategy $\mathsf{S} \in \boldsymbol{\mathsf{S}}$. In harmony with Agreement 2, the chain rule for pds, [63], factorises the closed-loop model in a well-interpretable way. For simplicity, the factorisation is made for ignorance containing only unused observations, actions and the unknown parameter. The factorisation has the form

$$\underset{\text{closed-loop model}}{\mathsf{C}_\mathsf{S}(b)} = \underset{\text{evironment model}}{\prod_{t \in \boldsymbol{t}} \mathsf{M}(o_t|a_t,k_{t-1},\Theta)} \times \underset{\text{parameter model}}{\prod_{t \in \boldsymbol{t}} \mathsf{P}(\Theta|k_{t-1})} \times \underset{\text{strategy model}}{\prod_{t \in \boldsymbol{t}} \mathsf{S}(a_t|k_{t-1})} \quad (3.1)$$

Remarks

- All factors are generally time-variant.
- The environment model relates observations to internals, here, to the unknown parameter. The terms "parametric environment model" or, briefly, "parametric model" are used. The parameter model is traditionally called the posterior pd.
- Within a single DM problem, the parametric environment model is assumed to be common to all strategies $\mathsf{S} \in \boldsymbol{\mathsf{S}}$. This does not restrict the presentation.
- The posterior pd evolves according to Bayes' rule, valid under adopted natural conditions of control [63], stating that Θ is unknown to decision maker:

$$\mathsf{P}(\Theta|k_t) = \frac{\mathsf{M}(o_t|a_t,k_{t-1},\Theta)\mathsf{P}(\Theta|k_{t-1})}{\mathsf{M}(o_t|a_t,k_{t-1})}, \text{ starting from a prior pd } \mathsf{P}(\Theta|k_0),$$

$$\mathsf{M}(o_t|a_t,k_{t-1}) = \int_{\boldsymbol{\Theta}} \mathsf{M}(o_t|a_t,k_{t-1},\Theta)\mathsf{P}(\Theta|k_{t-1})\mathrm{d}\Theta. \quad (3.2)$$

[4] When ignorance includes non-constant internals, Bayesian learning used below becomes stochastic filtering, [30]. If moreover, the decision maker's preferences depend on an action-dependent internal state, the stochastic control problem arises [41]. This general case is not treated here as it complicates explanations without offering any conceptual shift.

- The predictive pds $(\mathsf{M}(o_t|a_t,k_{t-1}))_{t\in\boldsymbol{t}}$ (3.2) used in the denominator of Bayes' rule form an (external) environment model. The parametric and environment models coincide whenever ignorance contains no unknown parameter. This has motivated the use of the same letter M for these different pds.
- The factorisation (3.1) and Bayes' rule (3.2) are correct if the learnt parameter $\Theta \in \boldsymbol{\Theta}$ and actions a_t are conditionally independent
$$\mathsf{S}(a_t|k_{t-1},\Theta) = \mathsf{S}(a_t|k_{t-1}),\ t \in \boldsymbol{t}. \tag{3.3}$$
The assumption (3.3) expresses natural conditions of control, [63], which are met for the optimised strategies.
- The strategy model $\mathsf{S} = (\mathsf{S}(a_t|k_{t-1}))_{t\in\boldsymbol{t}}$ is composed of decision rules $\mathsf{S}(a_t|k_{t-1}), t \in \boldsymbol{t}$. □

FPD assumes that the decision maker expresses her preferences between a pair of behaviours $b_1, b_2 \in \boldsymbol{b}$ by an ideal closed-loop model, which is a pd[5] $\mathsf{C}^\star(b)$ defined on $\boldsymbol{b}$. By definition[6]

$$\begin{aligned} &b_1 \preceq_{\boldsymbol{b}} b_2 \text{ means: } b_1 \text{ is preferred against } b_2 && \text{iff } \mathsf{C}^\star(b_1) \geq \mathsf{C}^\star(b_2) \\ &b_1 \prec_{\boldsymbol{b}} b_2 \text{ means: } b_1 \text{ is strictly preferred against } b_2 && \text{iff } \mathsf{C}^\star(b_1) > \mathsf{C}^\star(b_2). \end{aligned} \tag{3.4}$$

FPD also orders strategies $\mathsf{S}_1, \mathsf{S}_2 \in \boldsymbol{\mathsf{S}}$ via the same pd $\mathsf{C}^\star = (\mathsf{C}^\star(b))_{b\in\boldsymbol{b}}$ by comparing closed-loop models $\mathsf{C}_1 = \mathsf{C}_{\mathsf{S}_1}$, $\mathsf{C}_2 = \mathsf{C}_{\mathsf{S}_2}$ connected with them

$$\begin{aligned} &\mathsf{S}_1 \preceq_{\boldsymbol{\mathsf{S}}} \mathsf{S}_2 \text{ means: } \mathsf{S}_1 \text{ is preferred against } \mathsf{S}_2 && \text{iff } \mathscr{D}(\mathsf{C}_1||\mathsf{C}^\star) \leq \mathscr{D}(\mathsf{C}_2||\mathsf{C}^\star) \\ &\mathsf{S}_1 \prec_{\boldsymbol{\mathsf{S}}} \mathsf{S}_2 \text{ means: } \mathsf{S}_1 \text{ is strictly preferred against } \mathsf{S}_2 && \text{iff } \mathscr{D}(\mathsf{C}_1||\mathsf{C}^\star) < \mathscr{D}(\mathsf{C}_2||\mathsf{C}^\star). \end{aligned} \tag{3.5}$$

The functional $\mathscr{D}(\mathsf{C}_\mathsf{S}||\mathsf{C}^\star)$ used in (3.5) is the Kullback-Leibler divergence

$$\mathscr{D}(\mathsf{C}_\mathsf{S}||\mathsf{C}^\star) = \mathscr{E}_\mathsf{S}\left[\ln\left(\frac{\mathsf{C}_\mathsf{S}}{\mathsf{C}^\star}\right)\right] = \int_{\boldsymbol{b}} \ln\left(\frac{\mathsf{C}_\mathsf{S}(b)}{\mathsf{C}^\star(b)}\right)\mathsf{C}_\mathsf{S}(b)\mathrm{d}b. \tag{3.6}$$

Hence, the KLD is the S-dependent expectation $\mathscr{E}_\mathsf{S}[\bullet] = \int_{\boldsymbol{b}} \bullet \mathsf{C}_\mathsf{S}(b)\mathrm{d}b$ of the S-dependent loss, L_S,

$$\mathsf{L}_\mathsf{S}(b) = \ln\left(\frac{\mathsf{C}_\mathsf{S}(b)}{\mathsf{C}^\star(b)}\right),\ b \in \boldsymbol{b}. \tag{3.7}$$

The ordering (3.5) reflects the fact that the strategy S_1 is better than S_2 as it provides the closed-loop model $\mathsf{C}_1 = \mathsf{C}_{\mathsf{S}_1}$, which is closer to the ideal closed-loop model $\mathsf{C}^\star$ than the closed-loop model $\mathsf{C}_2 = \mathsf{C}_{\mathsf{S}_2}$ with the strategy S_2.

The most preferred strategy, $\mathsf{S}^o \preceq_{\boldsymbol{\mathsf{S}}} \mathsf{S},\ \forall \mathsf{S} \in \boldsymbol{\mathsf{S}}$, see (3.5), is called the optimal strategy (in the FPD sense). In order to describe its construction, it is useful to factorise the ideal closed-loop model $\mathsf{C}^\star$ in a manner similar to (3.1)

[5] Further on, the superscript * marks pds and actions arising from this ideal closed-loop model.

[6] The quest for simple final formulas has motivated a slightly non-standard choice of the "directions" of the ordering operators $\preceq$, $\geq$ and $\prec$, $>$.

$$\underset{\substack{\text{ideal}\\ \text{closed-loop model}}}{\mathsf{C}^{\star}_{\mathsf{S}}(b)} = \underset{\substack{\text{ideal}\\ \text{environment model}}}{\prod_{t\in\boldsymbol{t}}\mathsf{M}^{\star}(o_t|a_t,k_{t-1},\Theta)} \times \underset{\substack{\text{ideal}\\ \text{parameter model}}}{\prod_{t\in\boldsymbol{t}}\mathsf{P}^{\star}(\Theta|k_{t-1})} \times \underset{\substack{\text{ideal}\\ \text{strategy model}}}{\prod_{t\in\boldsymbol{t}}\mathsf{S}^{\star}(a_t|k_{t-1},\Theta)}$$

Remark

- The names "ideal environment model, ideal strategy model and ideal knowledge model" are to be understood as mnemonic analogies. For instance, the ideal strategy model $\mathsf{S}^{\star}$ may depend on the unknown parameter. As such, it is necessarily an element of the set S of strategies, which can be used by the decision maker. Even when $\mathsf{S}^{\star} \in \mathsf{S}$ then $\mathscr{D}(\mathsf{C}_{\mathsf{S}^o}||\mathsf{C}^{\star}) < \mathscr{D}(\mathsf{C}_{\mathsf{S}^{\star}}||\mathsf{C}^{\star})$ since the ideal strategy is close to itself but does not make the environment model close to the ideal environment model, in general. □

The following results, used later on, are proved in [41,45].

Proposition 1 (Solution of FPD; Relation to Bayesian DM) *Let the parameter-independent environment model* M*, its ideal counterpart* $\mathsf{M}^{\star}$ *and the ideal strategy* $\mathsf{S}^{\star}$ *be given (they only operate on observations, actions and prior knowledge). With the ordering (3.5), the optimal (*$\preceq_{\mathsf{S}}$*-most preferred) randomised strategy* S^o *is described by the pd normalised by* $\gamma(k_{t-1})$

$$\mathsf{S}^o(a_t|k_{t-1}) = \frac{\mathsf{S}^{\star}(a_t|k_{t-1})\exp[-\omega(a_t,k_{t-1})]}{\gamma(k_{t-1})} \tag{3.8}$$
$$\omega(a_t,k_{t-1}) = \int_{\boldsymbol{o}} \mathsf{M}(o_t|a_t,k_{t-1})\ln\left(\frac{\mathsf{M}(o_t|a_t,k_{t-1})}{\mathsf{M}^{\star}(o_t|a_t,k_{t-1})}\right)\mathrm{d}o_t - \mathscr{E}[\ln(\gamma(k_t))|a_t,k_{t-1}],$$

with $\mathscr{E}[\ln(\gamma(k_t))|a_t,k_{t-1}] = \int_{\boldsymbol{o}} \mathsf{M}(o_t|a_t,k_{t-1})\ln(\gamma(k_t))\mathrm{d}o_t$.

The evaluations (3.8) run backward for $t \in \boldsymbol{t}$ *and the value function* $-\ln(\gamma(k_t))$*, [5], is zero at the decision horizon* $t = T$.

For any strategy-independent loss $\mathsf{L} : \boldsymbol{b} \to (-\infty,\infty]$ *and the ideal pd*

$$\mathsf{C}^{\star}(b) \propto \exp[-\mathsf{L}(b)/\xi]\prod_{t\in\boldsymbol{t}}\mathsf{M}(o_t|a_t,k_{t-1}),\ \propto\ \textit{means proportionality}\ ,\ \xi > 0,$$

$$\mathscr{D}(\mathsf{C}_{\mathsf{S}}||\mathsf{C}^{\star}) = \frac{1}{\xi}\mathscr{E}_{\mathsf{S}}[\mathsf{L}] + \mathscr{E}_{\mathsf{S}}[\ln(\mathsf{S})] + \textit{strategy-independent function of}\ \xi. \tag{3.9}$$

If $\xi \to 0^+$ *then FPD with this* $\mathsf{C}^{\star}$ *solves with arbitrary precision the Bayesian DM task, given by the same behaviour set* $\boldsymbol{b}$*, the environment model* M*, and the loss* L.

Open Problem

Bayesian DM with non-parametric learning, [18], or particle filtering, [10], provide environment models of practical importance but with Dirac δ-function constituents. For such pds, the KLD is infinite and thus unsuitable for optimisation. To our best knowledge, a complete and correct treatment of this case is unavailable.

3.1.3 Discussed Decision-Making Aspects

The development of the considered normative DM theory began with the static DM task, [68]. It continued over multi-step but open-loop problems, and arrived at strategy design considering closed-loop behaviours with a finite or infinite decision horizon, [5, 17]. It has been extended to FPD and closely related, but independently developed and exploited DM methodologies [25,76,78].

The persisting discrepancies between normative and descriptive DM mentioned above, [27,49,56], have motivated us to join the search for their causes.

This chapter claims that the normative theory and its use do not sufficiently respect that the closed-loop behaviour matters. This wide-spread phenomenon has an obvious common source: the decision maker delimits the behaviour with respect to the environment to be influenced. She naturally includes in the behaviour actions and observations, as well as the environment internals, that she considers important. The *neglected part of the behaviour concerns the decision maker*. Her internals, reflecting directly unobserved preferences, emotional states, degree of "economic" rationality, the decision-maker's role in DM and cognitive effort, are rarely included into the considered behaviour, [73,75,80].

In order to overcome this, it is necessary to treat DM as a process proceeding from an informal specification of the DM problem up to the final use of the strategy. The chapter follows this course, identifies the important omissions, discusses them and searches for solutions.

3.1.4 Layout

The extent of the considered problem prevents its complete coverage. The presented sub-selection of existing tasks is inevitably subjective and reflects the both authors' knowledge and the subjectively perceived importance of the existing sub-problems.

Section 3.2 focuses on the completion and conceptual quantification of preferences driving DM. This still insufficiently developed part of DM theory is vital for converting the informally-specified DM problem into an algorithmically amenable task. The section also recalls the pathway to FPD.

Section 3.3 points to the main sources of discrepancies between the current normative and practical DM. The discussed difference between declared and real preferences originating within the supported decision maker. DM theory is blamed for the lack of systematic support of (the inevitably approximate) learning of an environment model and a design of an (approximately) optimal DM strategy.

Section 3.4 deals with controlling the deliberation effort expended on DM and with the influence of the decision-maker's role within DM on the formalisation of the DM problem. Section 3.5 provides concluding remarks.

Open Problem

The normative DM theory still does not serve its purpose sufficiently well. The addressed weak points were selected subjectively. Even the list of known theoretical bottlenecks may not be complete. For instance, the presented DM theory heavily relies on Kolmogorov's probability theory while the successful use of quantum probability in explaining cognitive processes, [64], indicates that a more general normative DM theory warrants consideration.

3.2 Quantitative Description of Preferences

The specification of the ideal closed-loop model $\mathsf{C}^{\star}$ is the crucial and poorly supported step of the conversion of an informally specified DM task into the form required by the normative theory. The need to quantify complete preferences both in behaviour and strategy spaces, cf. (3.4), (3.5), is the key difficulty. This section: i) discusses completeness of the preferences; ii) shows how an extension of the behaviour leads to completeness; iii) outlines the ways in which preferences may be learnt.

3.2.1 Fully Probabilistic Design of DM Strategies

FPD is the normative theory inspected in this chapter. The inspection is supported by the sketch of its axiomatic basis, [45].

FPD has the ambition to serve all decision makers who may differ in their preferences and available strategies. For this, FPD needs to specify a priori preferences among all behaviours $b \in \boldsymbol{b}$ in spite of the fact that the vast majority of them will never be realised. Subsection 3.2.2 shows how to extend the decision maker's partial preferences among behaviours into a complete ordering $\preceq_{\boldsymbol{b}}$. The completeness of prior ordering of all strategies $\mathsf{S} \in \boldsymbol{\mathsf{S}}$ follows from the freedom of decision makers to select the optimal strategy S^{o} in an arbitrarily chosen subset of $\boldsymbol{\mathsf{S}}$. For this, they need the freedom to restrict the strategy ordering $\preceq_{\boldsymbol{\mathsf{S}}}$ to the ordering on the considered subset of compared strategies. Typically, limited deliberation or technical resources enforce this restriction. Since the decision maker may consider the subset containing an arbitrary strategy pair from $\boldsymbol{\mathsf{S}}$, the ordering $\preceq_{\boldsymbol{\mathsf{S}}}$ has to be complete. This explains why the following outline considers the complete orderings $\preceq_{\boldsymbol{b}}$, $\preceq_{\boldsymbol{\mathsf{S}}}$.

The Way to FPD

- The complete preference ordering of behaviours $\preceq_{\boldsymbol{b}}$ defines a non-empty collection of open intervals $(b_1, b_2) = \{b \in \boldsymbol{b} : b_1 \prec_{\boldsymbol{b}} b \prec_{\boldsymbol{b}} b_2\}$ given by behaviours

$b_1, b_2 \in \boldsymbol{b}$ with a strict preference[7] $b_1 \prec_{\boldsymbol{b}} b_2$. [13, 20] show that there is loss $\mathsf{L} : \boldsymbol{b} \rightarrow (-\infty, \infty]$ quantifying the ordering $\preceq_{\boldsymbol{b}}$ in the sense

$$\Big(\big(b_1 \preceq_{\boldsymbol{b}} b_2 \,\wedge\, b_2 \preceq_{\boldsymbol{b}} b_1\big) \Leftrightarrow \mathsf{L}(b_1) = \mathsf{L}(b_2)\Big) \text{ and } \Big(b_1 \prec_{\boldsymbol{b}} b_2 \Leftrightarrow \mathsf{L}(b_1) < \mathsf{L}(b_2)\Big)$$

iff from any collection of open intervals $\{(b_\alpha, b_\beta)\}$ covering $\boldsymbol{b}$ it is possible to select at most a countable sub-collection covering $\boldsymbol{b}$. Briefly, it means that the quantification by a real-valued loss exists iff the topology of the open intervals on $\boldsymbol{b}$ is not richer than the topology of open intervals on the real line.

- The inspected DM under uncertainty means that the behaviour is not uniquely determined by the used strategy $\mathsf{S} \in \mathbf{S}$. To model this, an additional variable, called uncertainty, $u \in \boldsymbol{u} \neq \emptyset$, and a strategy-dependent mapping $\mathsf{R}_\mathsf{S} : \boldsymbol{u} \rightarrow \boldsymbol{b}$ are introduced[8]. To each loss L the composition $\Lambda_\mathsf{S} = \mathsf{L} \circ \mathsf{R}_\mathsf{S}$ of L and R_S, defining $\Lambda_\mathsf{S}(u) = \mathsf{L}(\mathsf{R}_\mathsf{S}(u))$, is assigned. These strategy-dependent "auxiliary" functions of uncertainty $\Lambda_\mathsf{S} \in \boldsymbol{\Lambda}_\mathsf{S} = \{\mathsf{L} \circ \mathsf{R}_\mathsf{S}\}_{\mathsf{L} \in \mathbf{L}}$ serving for further steps.
- The ordering $\preceq_{\mathbf{S}}$ completely orders the functions $\Lambda_\mathsf{S} \in \boldsymbol{\Lambda} = \cup_{\mathsf{S} \in \mathbf{S}} \boldsymbol{\Lambda}_\mathsf{S}$

$$\Lambda_{\mathsf{S}_1} \preceq_{\boldsymbol{\Lambda}} \Lambda_{\mathsf{S}_2} \text{ holds by definition iff } \mathsf{S}_1 \preceq_{\mathbf{S}} \mathsf{S}_2, \quad \mathsf{S}_1, \mathsf{S}_2 \in \mathbf{S}. \tag{3.10}$$

Assuming the countability of the open intervals defined by $\preceq_{\mathbf{S}}$, and thus by $\preceq_{\boldsymbol{\Lambda}}$, the strategy ordering $\preceq_{\mathbf{S}}$ can be (numerically) quantified. Due to (3.10), it can be quantified via a functional $\mathscr{L}$ acting on $\boldsymbol{\Lambda}$. A sufficiently smooth local functional[9] has an integral representation determined by a function U and a probabilistic measure $\mathsf{C}(u)\mathrm{d}u$. It has the form, [65],

$$\mathscr{L}(\Lambda) = \int_{\boldsymbol{u}} \mathsf{U}(\Lambda(u), u)\mathsf{C}(u)\mathrm{d}u \quad \text{and} \quad \mathsf{S}_1 \preceq_{\mathbf{S}} \mathsf{S}_2 \;\Leftrightarrow\; \mathscr{L}(\Lambda_{\mathsf{S}_1}) \leq \mathscr{L}(\Lambda_{\mathsf{S}_2}). \tag{3.11}$$

The probabilistic measure is assumed for simplicity to be given by the pd $\mathsf{C}(u)$. The function $\mathsf{U}(\Lambda(u), u)$, fulfilling $\mathsf{U}(0, u) = 0$, scales the values $\Lambda(u)$ according to the uncertainty value[10].

- The substitution of the behaviour $b = \mathsf{R}_\mathsf{S}(u)$ into the integral representation (3.11) transforms the pd $\mathsf{C}(u)$ into the strategy-dependent closed-loop model $\mathsf{C}_\mathsf{S}(b)$ (3.1). The composition $\mathsf{U} \circ \Lambda_\mathsf{S}$ transforms – via the substitution $b = \mathsf{R}_\mathsf{S}(u)$

[7] The existence of such pairs can be assumed without loss of generality. Indeed, no non-trivial decision task arises if all comparable pairs of behaviours in the original decision-maker-specified partial ordering are equivalent.

[8] The mapping R_S is common to decision makers differing only in preferences among behaviours.

[9] The functional is local if its value on Λ, artificially written as the sum $\Lambda_1 + \Lambda_2$ of functions Λ_1, Λ_2 fulfilling $\Lambda_1 \Lambda_2 = 0$, is the sum of its values on Λ_1 and Λ_2.

[10] The measure serves to all DM tasks facing the same uncertainty. The function U models risk awareness, neutrality or proneness. The function U, C-almost surely increasing in its first argument, guarantees that the optimal strategy S^o selected from the considered subset of $\mathbf{S}$ is not dominated. It means that it cannot happen that within this subset there is a strategy S^d such that $\Lambda_{\mathsf{S}^d}(u) \leq \Lambda_{\mathsf{S}^o}(u)$ on $\boldsymbol{u}$ with the sharp inequality on a subset of $\boldsymbol{u}$ of a positive C measure.

– to an S-dependent performance index $\mathsf{I}_{\mathsf{S}} : \boldsymbol{b} \to (-\infty,\infty]$. Its expectation $\mathscr{E}_{\mathsf{S}}[\bullet] = \int_{\boldsymbol{b}} \bullet \mathsf{C}_{\mathsf{S}}(b)\mathrm{d}b$ is taken in (3.11).

- The optimal strategy S^o with respect to $\preceq_{\mathsf{S}}$ on the full $\mathbf{S}$ determines the ideal closed-loop model $\mathsf{C}^\star$ as the closed-loop model with this strategy, $\mathsf{C}^\star = \mathsf{C}_{\mathsf{S}^o}$.
- Many expected performance indices $\mathscr{E}_{\mathsf{S}}[\mathsf{I}_{\mathsf{S}}]$ lead to the same ideal pd $\mathsf{C}^\star$; they are equivalent. The performance index $\mathsf{I}_{\mathsf{S}} = \ln(\mathsf{C}_{\mathsf{S}}/\mathsf{C}^\star)$ represents all performance indices: i) leading to the same ideal pd $\mathsf{C}^\star$; ii) depending smoothly on the optimised strategy entering I_{S} via the pd C_{S}, and iii) being independent of the realised behaviour for $\mathsf{C}_{\mathsf{S}} = \mathsf{C}^\star$. FPD simply uses this representative of equivalent performance indices, cf. (3.5) (3.6), (3.7).

Open Problem

The implicitly adopted handling of uncertainties $u \in \boldsymbol{u}$ together with σ-algebra of events makes $\mathsf{C}(u)\mathrm{d}u$ a Kolmogorov probability measure, [65]. This restricts the generality of FPD. There are strong indicators that "non-commutative probability", [15], widely used in quantum mechanics, is more adequate and can improve modelling of the "macroscopic" DM environment. A systematic development of this direction is open and quite challenging.

3.2.2 Completion of Preference Ordering

The existence of preferences $\preceq_{\boldsymbol{b}}$ with a non-empty strict part $\prec_{\boldsymbol{b}}$ makes any DM meaningful. Section 3.2.1 exploits its completeness. Everyday experience confirms that a human decision maker cannot provide the complete ordering $\preceq_{\boldsymbol{b}}$ for difficult cases calling for the use of the normative theory. Thus, there is a need for a systematic, automatically performed, completion.

Primarily, it has to be clear how to construct conceptually such a completion. It suffices to consider closed-loop behaviours $b \in \boldsymbol{b}$ having at most a countable number of realisations, cf. the conditions for the loss existence in Subsection 3.2.1.

To any pair of behaviours $b_1, b_2 \in \boldsymbol{b}$ that are un-compared by the decision maker, there are variants of the preference ordering, denoted $\preceq_{\boldsymbol{b}|\theta}$, such that $b_1 \preceq_{\boldsymbol{b}|\underline{\theta}} b_2$ and $b_2 \preceq_{\boldsymbol{b}|\overline{\theta}} b_1$ for "pointers" $\underline{\theta}$, $\overline{\theta}$. The list of all distinct alternative preference orderings has at most a countable number of different entries labelled by pointers $\theta \in \boldsymbol{\theta} \subseteqq \{1,\ldots,\infty\}$. For each fixed pointer $\theta \in \boldsymbol{\theta}$, the corresponding complete ordering of behaviours is quantified by an ideal pd $\mathsf{C}^\star$ conditioned on this pointer:

$$b_1 \preceq_{\boldsymbol{b}|\theta} b_2 \Leftrightarrow \mathsf{C}^\star(b_1|\theta) \geq \mathsf{C}^\star(b_2|\theta). \tag{3.12}$$

Any ordering of the countable set $\boldsymbol{\theta}$ of pointers can be quantified by a positive pd, say $\mathsf{C}^\star(\theta)$. Multiplying (3.12) by the pd $\mathsf{C}^\star(\theta)$ and using the chain rule for pds, a complete ordering $\preceq_{(\boldsymbol{b},\boldsymbol{\theta})}$ is obtained. It acts on behaviours with the ignorance part extended by the unknown constant pointer $\theta \in \boldsymbol{\theta}$

$$(b_1,\theta,) \preceq_{(\boldsymbol{b},\boldsymbol{\theta})} (b_2,\theta) \Leftrightarrow \mathsf{C}^{\star}(b_1,\theta) \geq \mathsf{C}^{\star}(b_2,\theta),\ \ \mathsf{C}^{\star}(b,\theta) = \mathsf{C}^{\star}(b|\theta)\mathsf{C}^{\star}(\theta). \quad (3.13)$$

The ideal pd acts on an additional unknown parameter (pointer) θ characterising the completion of the decision maker's preferences with respect to the original behaviours $b \in \boldsymbol{b}$. The completion formally compares only the behaviour pairs of the form $(\bullet,\theta)$ and $(\star,\theta)$, i.e. having the same value of the pointer θ. The complete ordering on such a set suffices for DM.

Open Problem

The countability of $\boldsymbol{b}$ avoids use of the axiom of choice, [58], and there is a need for a non-trivial check as to whether the extension meets the conditions for the existence of an ordering-quantifying loss L, [13, 20]. It is desirable and non-trivial to remove the countability assumption.

3.2.3 Ways to Preference Elicitation

The unknown pointer θ to alternative behaviour orderings enters the ideal pd, see Subsection 3.2.2. It can be learnt similarly to any parameter belonging to ignorance. This transforms the difficult preference completion problem, known as preference elicitation, [7,8,11,34], into the Bayesian-learning framework. This significantly extends the applicability of the normative DM theory. Bayesian learning of the pointer θ, which is part of the general unknown parameter $\Theta \in \boldsymbol{\Theta}$, is possible if its influence on observations can be established. Otherwise, the minimum Kullback-Leibler divergence principle is available, [70]. Both possibilities are discussed below.

3.2.3.1 Bayesian Learning in Preference Elicitation

With inclusion of the pointer θ into Θ, the ideal closed-loop model $\mathsf{C}^{\star}$ and thus the ideal environment model $\mathsf{M}^{\star}(o_t|a_t,k_{t-1},\Theta)$ as well as the ideal strategy model $\mathsf{S}^{\star}(a_t|k_{t-1},\Theta)$ depend on the Θ. The factorisations of the closed-loop model (3.1) and of its ideal counterpart, chain rules for conditional expectations, their linearity and the definition of the KLD (3.6) imply the following form of the KLD to be minimised over the admissible strategies $\mathsf{S} = (\mathsf{S}(a_t|k_{t-1}))_{t\in\boldsymbol{t}}$

$$\mathscr{D}(\mathsf{C}_{\mathsf{S}}||\mathsf{C}^{\star}) = \mathscr{E}_{\mathsf{S}}\Big\{\sum_{t\in\boldsymbol{t}}\mathscr{E}_{\mathsf{S}}\Big[\ln\Big(\frac{\mathsf{M}(o_t|a_t,k_{t-1},\Theta)\mathsf{S}(a_t|k_{t-1})}{\mathsf{M}^{\star}(o_t|a_t,k_{t-1},\Theta)\mathsf{S}^{\star}(a_t|k_{t-1},\Theta)}\Big)\Big|k_{t-1}\Big]\Big\}. \quad (3.14)$$

As it is inherent to the Bayesian paradigm, there is no need to select a unique $\Theta \in \boldsymbol{\Theta}$ and handle it as the correct one. All possibilities are admitted but within the outer expectation in (3.14) they are weighted by the posterior pd $\mathsf{P}(\Theta|k_{t-1})$ evolving according to Bayes' rule (3.2). This applies to the pointer θ distinguishing alternative ordering, which is a part of Θ. Thus, it is unnecessary to select a unique preference ordering as the correct one[11].

The use of Bayesian learning (3.2) assumes availability of the environment model $\mathsf{M}(o_t|a_t,k_{t-1},\Theta)$ relating the observations to an unknown parameter. In the preference-elicitation context, it is relatively simple to construct the parametric environment model if the observations explicitly reflect decision-maker's satisfaction with the DM course. This situation is, for instance, "natural" in various service-oriented enterprises. They systematically collect data directly reflecting satisfaction of their customers, [24]. They care about the design of questionnaires to be filled in by, say, patients in health care. Typically, they jointly consider relations of the sale levels to positions of goods in super-markets, analyse positions of clicks within lists retrieved as the answer to a customer query, etc. Then, black-box modelling, say via neural networks, [28] or finite mixtures, [77], or their discrete-valued versions, [38], suffice to relate the abundant data to the level of satisfaction. Black-box models rely on "universal approximation property", [28], and can easily be created but their learning heavily depend on data informativeness[12].

A deeper-rooted modelling is needed when the observations have only an indirect connection with the decision-maker's satisfaction. Fields that study decision makers, like behavioural economics, [73], neuro-economics, [19], or psychology of DM, [32], have to provide grey-box models, [6], relating the observations to the satisfaction and consequently to the pointer θ. The difficulty with this problem stems from the fact that satisfaction is strongly influenced by the decision maker's non-quantified experience, limited ability to grasp relations between many variables, personality and even emotions. Adequate modelling needs cognitive sciences as well as aforementioned research branches. This is quite demanding, but experience from technological applications confirms that extremely simplified models often suffice for excellent DM[13]. The key point is that they do not ignore the influence of the internals, which are related to observations and determine the degree of satisfaction.

Open Problem

The frequently observed non-transitive preferences of real decision makers can be interpreted as varying preferences. They can be modelled by the time-dependent

[11] [22] represents non-Bayesian set-ups dealing with sets of orderings without a quest for a unique completion.

[12] A decision maker interacts with customers in order to influence them in a desirable direction, for instance, to buy a specific product or services. However, even the form of the questionnaire influences the customers: typically, two different ways of posing logically the same question often provide quite different answers. This quantum-mechanics-like effect should be properly modelled.

[13] The vast majority of complex technological processes, which should be modelled by high-dimensional nonlinear stochastic partial differential equations with non-smooth boundary conditions, are controlled by proportional-integral-derivative controllers corresponding to simple linear, second order difference equations used as the environment model.

pointer θ_t used for the ordering completion, see Subsection 3.2.2. Bayesian filtering [30] can cope with this case. It needs similar but more difficult modelling of time-evolution of this internal variable. This demanding task is worth of addressing in important application domains.

3.2.3.2 Learning from Decision the Maker's Actions

The acceptance of the assumption that the decision maker is rational and selects her actions with the aim of optimising her unrevealed preferences is a specific but rich opportunity to elicit the decision-maker's preferences. In this case, the decision-maker's actions depend on the parameter to be learnt. This means that the natural conditions of control (3.3) are violated and actions have to be treated as observations. This part outlines the related modelling on a simple example of the Ultimatum Game.

The Ultimatum Game, described, for instance, in [19], models human, typically economical, DM. The game structure allows interesting conclusions to be reached by using quite simple means.

According to the game rules, the proposer offers to the responder (decision maker) a part $o_{t-1;1} > 0$ of a fixed budget $q > 0$. The responder may accept or reject the offer. The acceptance $a_t = 1$ increases the rewards $o_{t;2}$ and $o_{t;3}$ of the proposer and the responder accordingly. The rejection, $a_t = 0$, leaves both rewards unchanged.

The game was studied under the assumption that both players try to maximise their rewards. The proposer, rational in this sense, always offers the smallest positive amount and the responder accepts any positive offer. Experiments confirm that almost no responder accepts low offers and proposers respect this. The paper [48] experimentally investigated the hypothesis that the decision maker balances her personal reward with a term comparing the accumulated rewards of both players and reflecting the feeling of "self-fairness". To outline the adopted approach, the proposer is assumed to generate offers $(o_{t-1;1})_{t\in\boldsymbol{t}}$ independently according to a fixed pd and the self-fairness is reflected by the loss

$$\mathsf{L}(o_{t-1}, a_t) = -o_{t;2} + \theta_t o_{t;3} = -o_{t-1;2} + \theta_t o_{t-1;3} + a_t[-o_{t-1;1} + \theta_t(q - o_{t-1;1})]. \tag{3.15}$$

The fairness weight $\theta_t \geq 0$ is (subconsciously) known only to the responder, who varies it independently between the game repetitions around a constant expected value θ. The assumed rational responder minimises the loss (3.15) by choosing

$$a_t = \chi(-o_{t-1;1} + \theta_t(q - o_{t-1;1}) \leq 0), \text{ where } \chi \text{ is the set indicator.}$$

This description can be extended via the minimum KLD principle (see the next subsection) to the complete parametric model of the responder's strategy

$$\mathsf{S}(a_t = 0|k_{t-1}, \theta) = \exp\left[-\frac{o_{t-1;1}}{\theta(q - o_{t-1;1})}\right]. \tag{3.16}$$

The proposer can estimate the responder's preference-determining unknown parameter θ by observing $(a_t, o_{t-1;1})_{t\in \boldsymbol{t}}$ and using them as data in Bayes' rule (3.2) with (3.16) serving as the model of the responder forming the proposer's environment. The estimate of the degree of responder's self-fairness may serve the proposer for predicting future actions of the responder.

Open Problem

The presented approach is applicable to a range of DM tasks. The example indicates that it is indeed possible to learn the decision maker's preferences by assuming her rationality. The experimental results in [48] show that modelling of this type is surprisingly efficient. They confirm that very simplified models suffice for describing complex objects such as decision makers. It is not known to what extent the simplified modelling suffices. It is has to be studied experimentally under more realistic set-ups.

3.2.3.3 Minimum Kullback-Leibler Divergence Principle

Bayesian learning (3.2) accumulates the knowledge about an unknown parameter by inserting new observations into the condition of the posterior pd.

There is a broad class of problems in which the knowledge about the constructed pd is specified by a list of features that it should have. Then, the minimum KLD principle is applied. It recommends selection of the pd

$$\mathsf{F}^o \in \arg\min_{\mathsf{F}\in\mathbf{F}} \mathscr{D}(\mathsf{F}||\mathsf{F}_0)$$

as an extension of the partial knowledge specified by

$$\begin{aligned} \mathsf{F}_0 &: \text{ a pd interpreted as a prior guess of the constructed pd } \mathsf{F}^o \qquad (3.17)\\ \mathbf{F} &: \text{ a set of pds with the listed features of the constructed pd.} \end{aligned}$$

This principle is axiomatically justified in [70] and generalised in [40] as the FPD solution of a DM task selecting the pd partially delimited by the knowledge (3.17). The work [9] relates this principle to conditioning, finding them equivalent in many cases.

The minimum KLD principle provides a straightforward construction of the ideal closed-loop pd $\mathsf{C}^\star$, starting from a partial description of preferences, which it quantifies, [34]. Subsection 3.2.3.4 provides a possible specification of the set $\mathbf{F}$ (3.17) in the elicitation context.

Open Problem

Bayesian learning and the minimum KLD principle appear to exhaust all universal, theoretically justified, approaches to knowledge elicitation, and, within the FDP framework, preference elicitation. Challenging this conjecture is methodologically desirable. For instance, strong and extensive results connected with fuzzy ways of knowledge processing offer methodology and algorithms worth considering. It seems to be possible whenever the membership functions amenable to probabilistic interpretation.

3.2.3.4 Minimum KLD Principle in Preference Elicitation

We claim above that the minimum KLD principle has to be applied whenever the processed information about preferences has forms other than observations. Here, a specific but widely encountered regulation task is considered to illustrate how it can be done. The treatment is a continuous-valued counterpart of the discussion in [34], which focused on discrete-valued observations and actions.

The regulation [60] is the DM task in which the decision maker selects actions $a_t \in \boldsymbol{a}$ making the observations $o_t \in \boldsymbol{o}$ as close as possible to a given reference $r_t \in \boldsymbol{o}$, $t \in \boldsymbol{t}$. The inspected reference elicitation should construct an ideal closed-loop pd $\mathsf{C}^\star$, which: i) reflects the verbally and incompletely specified regulation preferences; ii) is ambitious but potentially attainable.

The construction of $\mathsf{C}^\star$ starts with the insight that the attaining of the reference r_t is most probable with the action[14]

$$a_t^\star(k_{t-1}) \in \operatorname{Arg}\max_{a_t \in \boldsymbol{a}} \mathsf{M}(r_t|a_t,k_{t-1}) \tag{3.18}$$

where the environment model M and the set $\boldsymbol{a}$ enter the DM formalisation anyway.

The action $a_t^\star(k_{t-1})$ specifies the factor $\mathsf{M}^\star(o_t|k_{t-1}) = \mathsf{M}(o_t|a_t^\star(k_{t-1}),k_{t-1})$ of the constructed ideal pd, which properly expresses the regulation objective in an ambitious, and, – given fortunate circumstances – attainable way. What remains is to select $\mathsf{S}^\star(a_t|o_t,k_{t-1})$ in order to get complete $\mathsf{C}^\star(o_t,a_t|k_{t-1}) = \mathsf{M}^\star(o_t|k_{t-1})\mathsf{S}^\star(a_t|o_t,k_{t-1})$.

A pd $\mathsf{S}_0^\star(a_t|o_t,k_{t-1})$ with its support on $\boldsymbol{a}$, which is either flat or expresses preferences for less costly actions serves well as a first guess of the constructed $\mathsf{S}^\star(a_t|o_t,k_{t-1})$. However, the chain-rule composition $\mathsf{M}^\star(o_t|k_{t-1})\mathsf{S}_0^\star(a_t|o_t,k_{t-1})$ is an unsuitable candidate for $\mathsf{C}^\star(o_t,a_t|k_{t-1})$ because an adequate joint pd should prefer actions around $a_t^\star(k_{t-1})$ defining $\mathsf{M}^\star(o_t|k_{t-1})$. The restriction on the ideal decision rules in the set

$$\left\{\mathsf{S}^\star(a_t|o_t,k_{t-1}) : \int_{\boldsymbol{a}} a_t \mathsf{S}^\star(a_t|o_t,k_{t-1}) \mathrm{d}a_t = a_t^\star(k_{t-1})\right\}. \tag{3.19}$$

[14] The adopted notation $a^\star$ stresses that this action value serves for the construction of $\mathsf{C}^\star$.

is the simplest soft expression of the wish to be around $a_t^{\star}(k_{t-1})$.

The incomplete knowledge of preferences on actions, delimited by the set (3.19) and by the prior guess $\mathsf{S}_0^{\star}(a_t|o_t,k_{t-1})$, has the form (3.17). Thus, the minimum KLD principle can be directly used for its completion. It provides the following ideal strategy

$$\mathsf{S}^{\star}(a_t|o_t,k_{t-1}) \propto \mathsf{S}_0^{\star}(a_t|o_t,k_{t-1}) \exp\left\langle \zeta(k_{t-1}), a_t \right\rangle, \tag{3.20}$$

with the real-valued vector $\zeta(k_{t-1})$, making the scalar product $\langle \zeta(k_{t-1}), a_t\rangle$ meaningful, chosen so that equality in (3.19) is met. This implies that

$$\mathsf{C}^{\star}(o_t,a_t|k_{t-1}) \propto \mathsf{M}(a_t|a_{t-1}^{\star}(k_{t-1}),k_{t-1})\mathsf{S}_0^{\star}(a_t|o_t,k_{t-1}) \exp\left\langle \zeta(k_{t-1}), a_t \right\rangle \tag{3.21}$$

is the ambitious realistic ideal pd searched for.

Remarks

- The environment model M is generically obtained as the predictive pd arising from Bayesian learning (3.2). This answers why the action $a_t^{\star}(k_{t-1})$ (3.18) is not directly applied instead of the above complex indirect construction: the action $a_t^{\star}(k_{t-1})$ is exploitive and FPD adds the needed explorative character via the optimal randomised strategy arising from it.
- The dynamics of DM answers the question why the ideal decision rule (3.20) is not directly used as a part of the optimal strategy. Satisfying the local aim at time t can lead to bad initial conditions for the subsequent steps. Consequently, a repetitive use of one-step-ahead-looking (myopic, greedy) rules may be far from the optimal strategy, [5] and may even make the closed loop unstable, [43].
- The simple linear-Gaussian case offers an insight into (3.21). In this case[15],

$$\mathsf{M}(o_t|a_t,k_{t-1}) = \mathsf{N}_{o_t}(Ao_{t-1}+Ba_t,Q),\ \mathsf{S}_0^{\star}(a_t|k_{t-1}) = \mathsf{N}_{a_t}(Co_{t-1},R), \tag{3.22}$$
$$\text{where}\quad \mathsf{N}_x(\mu,\rho) = |2\pi\rho|^{-0.5} \exp\left[-0.5(x-\mu)'\rho^{-1}(x-\mu)\right],\ '\text{ is transposition,}$$

  and where the matrices A, B, and $Q > 0$ (positive definite), of dimensions compatible with the vector observable state o_t, are known, either from modelling or, possibly recursive, learning. The matrices C, and $R > 0$ are chosen by the decision maker to have the majority of the probabilistic mass in the desirable action set $\boldsymbol{a}$. It is usually delimited by technological or economical constraints. Assuming for simplicity the reference $r_t = 0$ and observing that the delayed observation o_{t-1} coincides with the knowledge k_{t-1}, the proposed way gives

$$\mathsf{S}^{\star}(a_t|o_{t-1}) = \mathsf{N}_{a_t}(a_t^{\star},R),\ \ a_t^{\star}(o_{t-1}) = -(B'Q^{-1}B)^{-1}B'Q^{-1}Ao_{t-1} \tag{3.23}$$
$$\mathsf{M}^{\star}(o_t|o_{t-1}) = \mathsf{N}_{o_t}(Fo_{t-1},Q),\ \text{ with } F = (I - B(B'Q^{-1}B)^{-1}B'Q^{-1})A.$$

  FPD with the Gaussian environment model and the ideal pd is a randomised version of the widespread design dealing a with linear environment (system) and quadratic loss, [33]. This loss is

[15] $\mathsf{S}_0^{\star}(a_t|o_t,k_{t-1})$ and $a_t^{\star}(k_{t-1})$ are independent of o_t, i.e. $\mathsf{S}^{\star}(a_t|o_t,k_{t-1}) = \mathsf{S}^{\star}(a_t|k_{t-1})$, see (3.21).

$$\sum_{t \in \boldsymbol{t}} (o_t - r_t)' Q_{\boldsymbol{o}} (o_t - r_t) + (a_t - r_{t;a})' Q_{\boldsymbol{a}} (a_t - r_{t;a})$$

with a given r_t, $r_{t;a}$, and $Q_{\boldsymbol{o}}$, $Q_{\boldsymbol{a}} > 0$. While the choice of references r_t, $r_{t;a}$ is well understood and mechanised, the choice of penalisation matrices $Q_{\boldsymbol{o}}$, $Q_{\boldsymbol{a}}$ represents a non-trivial, repetitively solved, problem. The above preference elicitation solves this problem (almost) completely. The matrix $Q_{\boldsymbol{o}} = Q^{-1}$ comes from the learnt environment model. The matrix $Q_{\boldsymbol{a}} = R^{-1}$ is (almost) determined by the set $\boldsymbol{a}$, which has to be probabilistically "covered" by $\mathsf{N}_{a_t}(Co_{t-1}, R)$. This makes the result implicitly dependent on the matrix C.

The proposed solution extends the line presented in [2, 44, 61]. □

Open Problem

The term "almost" used in connection with the linear Gaussian case reflects the more general fact that the resulting, preference expressing, ideal pd depends on the prior, poorly guided, choice of $\mathsf{S}_0^{\star}(a_t|o_t, k_{t-1})$. Decision maker is to be supported even in this respect.

3.3 On Imperfect Use of Imperfect Normative Theory

The imperfect use and inherent imperfections of the adopted normative theory strongly influence the quality of the resulting DM. This section focuses on these sources of deviations revealed by descriptive DM theories.

3.3.1 Rationality Behind Non-Rational Decision Making

The discussed preference elicitation is an example of where the behaviour delimited by the decision maker needs to be extended. This is not a unique case. Considering a richer behaviour often offers an explanation of why real actions deviate from recommendations of the normative theory. Neglecting an important part of the closed-loop behaviour (e.g. emotional state of the decision maker) during the normative-theory-based strategy optimisation is a significant source of apparent irrationality.

The following formalisation of the above statement assumes for simplicity no internals (see Agreement 2) and thus Proposition 1 describes the optimal strategy.

In the discussion, the observation o_t splits into the formally non-empty non-optimised part n_t and the non-empty optimised part p_t. Then, the function $\omega(a_t, k_{t-1})$, determining the optimal strategy (3.8), can be given the form

$$\begin{aligned}\omega(a_t,k_{t-1}) = &\int_{(\boldsymbol{n},\boldsymbol{p})} \mathsf{M}(n_t,p_t|a_t,k_{t-1}) \ln\left(\frac{\mathsf{M}(n_t|p_t,a_t,k_{t-1})}{\mathsf{M}^{\star}(n_t|p_t,a_t,k_{t-1})}\right) \mathrm{d}(n_t,p_t) \\ &+ \int_{\boldsymbol{p}} \mathsf{M}(p_t|a_t,k_{t-1}) \ln\left(\frac{\mathsf{M}(p_t|a_t,k_{t-1})}{\mathsf{M}^{\star}(p_t|a_t,k_{t-1})}\right) \mathrm{d}p_t - \mathscr{E}[\ln(\gamma(k_t))|a_t,k_{t-1}],\end{aligned} \tag{3.24}$$

which exploits the fact that $\mathsf{M}(p_t|a_t,k_{t-1})/\mathsf{M}^{\star}(p_t|a_t,k_{t-1})$ does not depend on n_t.

Proposition 1 implies that the value function $-\ln(\gamma(k_t)) = 0$ at the horizon $t = T$,. The expression (3.24) implies that the function $\omega(a_t,k_{t-1})$ does not depend on the factor $\mathsf{M}(n_t|p_t,a_t,k_{t-1})$ of the environment model $\mathsf{M}(n_t,p_t|a_t,k_{t-1})$ iff this factor is equal to its ideal counterpart $\mathsf{M}^{\star}(n_t|p_t,a_t,k_{t-1})$. The function $\gamma(k_{t-1})$ is then uninfluenced by it, too, as follows from (3.8) and (3.24). This implies that the factors of the ideal environment model describing the non-optimised part of the behaviour n_t should be "left to their fate", [38],

$$\mathsf{M}^{\star}(n_t|p_t,a_t,k_{t-1}) = \mathsf{M}(n_t|p_t,a_t,k_{t-1}),\ t \in \boldsymbol{t}. \tag{3.25}$$

This makes the optimal strategy S^{o} independent of the n_t-related environment model factors. At the same time, the variables $n_\tau,\ \tau \leq t-1$ form a part of the knowledge k_{t-1} and thus their realisations influence the action a_t, $t \in \boldsymbol{t}$.

In real DM, the decision maker, possibly subconsciously and informally, optimises the behaviour she considers and thus she does not leave n_t to its fate. In this case, she designs a strategy, which differs from the theoretical one, and appears to be a non-rational person.

Open Problem

This discussed discrepancy between the normatively and practically optimised behaviours is widespread. Countermeasures require one to: i) admit that something has to be added into the closed-loop behaviour; and ii) model the influence of this addition on the DM task. The discussed modelling of the Ultimatum Game, Subsection 3.2.3.2, and of the deliberation effort, Subsection 3.4.1, suggest that improvements can be achieved by revising the behaviour specification. The extent to which the unifying view is useful for practice is unclear and specific cases have to be elaborated.

3.3.2 Approximate Learning

Applications of Bayesian learning face the curse of dimensionality, [3]. The evolving posterior pd $\mathsf{P}(\Theta|k_t)$ (3.2) is a function on a generically high-dimensional space $\boldsymbol{\Theta}$. Its complexity grows with the amount of processed data, which calls for approximate techniques. This makes an approximation an integral part of learning. Its quality influences the quality of the learning results. The theory of stochastic approximations, [55], is a dominating tool for analysing this influence. It mainly cares about point parameter estimates and provides qualitative guidelines for the design of approximate estimators. The normative theory, however, provides no systematic methodology for how to design approximate learning. This incompleteness of the theory leaves the decision maker unsupported. The key problems that are confronted are

- How the approximate pds should be chosen?
- What proximity measure should be used?
- How to combine the approximation with recursive learning when the knowledge is being enriched continually?

The subsequent text outlines a possible way towards making the normative theory more complete with respect to approximate learning by resolving these problems. The presentation concerns a specific but widely applicable case when the observations enter the parametric environment model $\mathsf{M}(o_t|a_t,k_{t-1},\Theta) = \mathsf{M}(o_t|a_t,\phi_{t-1},\Theta)$ via a finite-dimensional state vector ϕ_{t-1}, which can be recursively updated, $o_t,a_t,\phi_{t-1} \rightarrow \phi_t$. In the presentation, the observation o_t, action a_t and the state vector ϕ_{t-1} are collected into the data vector $x_t = (o_t,a_t,\phi_{t-1}) \in \boldsymbol{x}$.

Choice of Approximate Pds

Using the Dirac δ function, Bayes' rule (3.2) applied to the considered class of parametric models can be given the form

$$\begin{aligned}
&\mathsf{P}(\Theta|k_{t+1}) \propto \mathsf{P}(\Theta|k_0)\prod_{\tau=1}^{t}\mathsf{M}(o_\tau|a_\tau,\phi_{\tau-1},\Theta)\\
&= \mathsf{P}(\Theta|k_0)\exp\left[t\frac{1}{t}\sum_{\tau=1}^{t}\ln\big(\mathsf{M}(o_\tau|a_\tau,\phi_{\tau-1},\Theta)\big)\right]\\
&= \mathsf{P}(\Theta|k_0)\exp\left[t\frac{1}{t}\sum_{\tau=1}^{t}\int_{(\boldsymbol{o},\boldsymbol{a},\boldsymbol{\phi})}\delta\big((o,a,\phi)-(o_\tau,a_\tau,\phi_{\tau-1})\big)\ln\big(\mathsf{M}(o|a,\phi,\Theta)\big)\mathrm{d}(o,a,\phi)\right]\\
&= \mathsf{P}(\Theta|k_0)\exp\left[t\int_{\boldsymbol{x}}\mathsf{F}(x|k_{t+1})\ln(\mathsf{M}(o|a,\phi,\Theta))\mathrm{d}x\right], \text{ where} \qquad (3.26)
\end{aligned}$$

$\mathsf{F}(x|k_{t+1}) = \frac{1}{t}\sum_{\tau=1}^{t}\delta(x-x_\tau)$ is the empirical pd of the data vectors $x_t = (o_t,a_t,\phi_{t-1})$.

The pd $\mathsf{P}(\Theta|k_0)$ quantifies prior knowledge k_0 and the empirical pd $\mathsf{F}(x|k_{t+1})$ cumulates knowledge brought by the observed data up to and including t. If the parametric model belongs to the exponential family, i.e. of the form $\mathsf{A}(\Theta)\exp\langle\mathsf{B}(x),\mathsf{K}(\Theta)\rangle$ determined by an x-independent function $\mathsf{A}(\Theta) \geq 0$ and by a scalar product $\langle\mathsf{B}(x),\mathsf{K}(\Theta)\rangle$ of vector functions $\mathsf{B}(x)$, $\mathsf{K}(\Theta)$, then the data-based knowledge compresses into a finite-dimensional sufficient statistic. It is the sample average of $\mathsf{B}(x)$ and the degrees of freedom counting the number of processed samples. This is essentially the only universally feasible case [1,50].

The need for approximate learning arises when the parametric model $\mathsf{M}(o_t|a_t,\phi_{t-1},\Theta)$ does not belong to the exponential family. Let us consider this case.

The parametric environment model arises from physical modelling using first principles, e.g. [6,30]. Mostly, these do not determine the model completely but the minimum KLD principle, [70], Subsection 3.2.3.3, is available for their completion, [46]. This costly compression of invaluable domain knowledge should be preserved when approximating the posterior pd (3.26). Thus, the empirical pd $\mathsf{F}(x|k_{t-1})$ is to be approximated by a pd $\mathsf{F}(x|\nu_{t-1},V_{t-1})$ determined by degrees of freedom $\nu_{t-1} > 0$ and a non-sufficient statistic $V_{t-1} \in \boldsymbol{V}$ of a fixed finite dimension. The pd $\mathsf{P}(\Theta|k_{t-1})$ is then approximated as follows

$$\mathsf{P}(\Theta|k_{t-1}) \approx \mathsf{P}(\Theta|k_0)\exp\left[\nu_{t-1}\int_{\boldsymbol{x}} \mathsf{F}(x|\nu_{t-1},V_{t-1})\ln(\mathsf{M}(o|a,\phi,\Theta))\mathrm{d}x\right]. \quad (3.27)$$

This approximation is applicable to any parametric model operating on data vectors belonging to the same set *Sx*. This allows the decision maker to focus on exploitation of the domain knowledge and then to use a single universal approximate learning algorithm (cf. the situation in designing general-purpose optimisation algorithms).

Using the approximation (3.27) in Bayes' rule, the approximate pd $\mathsf{F}(x|\nu_{t-1},V_{t-1})$ is updated in the same way as the exact empirical pd

$$\begin{aligned}\mathsf{G}(x|\nu_{t-1},V_{t-1},x_t) &= \frac{\nu_{t-1}}{\nu_{t-1}+1}\mathsf{F}(x|\nu_{t-1},V_{t-1}) + \frac{1}{\nu_{t-1}+1}\delta(x-x_t) \quad (3.28)\\ &= (1-\beta_{t-1})\mathsf{F}(x|\nu_{t-1},V_{t-1}) + \beta_{t-1}\delta(x-x_t), \;\; \beta_{t-1} = \frac{1}{\nu_{t-1}+1}.\end{aligned}$$

To keep the computational complexity under control, the pd $\mathsf{G}(x|\nu_{t-1},V_{t-1},x_t)$ has to be again approximated by a feasible pd $\mathsf{F}(x|\nu_t,V_t)$, $\nu_t > 0$, $V_t \in \boldsymbol{V}$. The approximation quality depends on the chosen functional form $\mathsf{F}(x|\nu,V)$ of the approximate pd. A specific choice is made jointly with the choice of the proximity measure.

Proximity Measure and Functional Form of $\mathsf{F}(x|\nu,V)$

The axiomatically recommended approximation of a pd $\mathsf{G}(x)$ by a pd $\mathsf{F}(x)$ minimises the KLD of the approximated pd to the approximate pd, [4,40],

$$\mathscr{D}(\mathsf{G}||\mathsf{F}) = \underbrace{\int_{\boldsymbol{x}} \mathsf{G}(x)\ln\left(\frac{\mathsf{G}(x)}{\mathsf{F}(x)}\right)\mathrm{d}x}_{\text{KLD}} = \underbrace{-\int_{\boldsymbol{x}} \mathsf{G}(x)\ln\big(\mathsf{F}(x)\big)\mathrm{d}x}_{\text{Kerridge inaccuracy, [47]}} \; + \underbrace{\int_{\boldsymbol{x}} \mathsf{G}(x)\ln\big(\mathsf{G}(x)\big)\mathrm{d}x}_{\text{neg-entropy}}.$$

The unique minimiser of the Kerridge inaccuracy with respect to the pd F coincides with the unique minimiser of the KLD and avoids the problem that the approximated pd (3.28) contains Dirac δ functions, which make the neg-entropy infinite. Thus, the best approximation among pds $\{\mathsf{F}(x|\nu,V)\}_{\nu>0,V\in\boldsymbol{V}}$ of the intermediate outcome of

Bayes' rule $\mathsf{G}(x|\nu_{t-1},V_{t-1},x_t)$ (3.28) is given by the values

$$\tilde{\nu}_t, \tilde{V}_t \in \mathrm{Arg} \min_{\nu>0, V\in \boldsymbol{V}} - \int_{\boldsymbol{x}} \mathsf{G}(x|\nu_{t-1},V_{t-1},x_t) \ln(\mathsf{F}(x|\nu,V)) \mathrm{d}x. \quad (3.29)$$

The symbol ˜ stresses the fact that minimising arguments are intermediate quantities. They will be corrected further on in order to combine learning and approximation properly into ν_t, V_t, see the next subsection.

The functional form of approximate pds $\{\mathsf{F}(x|\nu,V)\}_{\nu>0,V\in\boldsymbol{V}}$ determines the achievable quality of approximation and the computational complexity of the minimisation task (3.29). The computational complexity is low for pds conjugate to the exponential family, [1]. In the data-vectors space $\boldsymbol{x}$, they have the form

$$\mathsf{F}(x|\nu,V) = \frac{\mathsf{A}^{\nu}(x) \exp\langle \mathsf{B}(x), V\rangle}{\int_{\boldsymbol{x}} \mathsf{A}^{\nu}(x) \exp\langle \mathsf{B}(x), V\rangle \,\mathrm{d}x}, \quad (3.30)$$

where $\mathsf{A}(x)$ is a non-negative function and the vector function $\mathsf{B}(x)$ makes the scalar product $\langle \mathsf{B}(x), V\rangle$ well defined (cf. the previous subsection).

Inserting the pd of the form (3.30) into the minimised Kerridge accuracy (3.29) and taking derivatives with respect to optional ν and V give the necessary conditions for determining $\tilde{\nu}_t, \tilde{V}_t$ in (3.29)

$$\begin{aligned} \int_{\boldsymbol{x}} \ln(\mathsf{A}(x))\mathsf{G}(x|\nu_{t-1},V_{t-1},x_t)\mathrm{d}x &= \int_{\boldsymbol{x}} \ln(\mathsf{A}(x))\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t)\mathrm{d}x \\ \int_{\boldsymbol{x}} \mathsf{B}(x)\mathsf{G}(x|\nu_{t-1},V_{t-1},x_t)\mathrm{d}x &= \int_{\boldsymbol{x}} \mathsf{B}(x)\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t)\mathrm{d}x. \end{aligned} \quad (3.31)$$

Thus, the minimisation reduces to the choice of $\tilde{\nu}_t, \tilde{V}_t$ matching the expectations of $\ln(\mathsf{A}(x))$ and $\mathsf{B}(x)$. For instance, if the approximate pd $\mathsf{F}(x|\nu,V)$ is a Gaussian pd then its mean and covariance have to coincide with the mean and covariance of $\mathsf{G}(x|\nu_{t-1},V_{t-1},x_t)$. This example also shows that the class of pds (3.30) is not rich enough to be used universally. Its members are mostly uni-modal and poorly approximate pds exhibiting light and heavy tails in different parts of their multivariate domain. Luckily, finite mixtures $\mathsf{F}(x|\nu,W)$, [77], of components given by pds $\mathsf{F}(x|\nu_j,V_j)$, $j\in\boldsymbol{j}=\{1,2,\ldots,J\}$, $J<\infty$, of the form (3.30)

$$\mathsf{F}(x|\nu,W) = \sum_{j\in\boldsymbol{j}} \alpha_j \mathsf{F}(x|\nu_j,V_j),\ \nu=(\nu_j)_{j\in\boldsymbol{j}},\ W=(V_j,\alpha_j>0)_{j\in\boldsymbol{j}},\ \sum_{j\in\boldsymbol{j}}\alpha_j = 1 \quad (3.32)$$

can approximate (loosely speaking) any pd $\mathsf{G}(x)$ as they have the universal approximation property, [28].

Generally, the evaluation and minimisation of the Kerridge inaccuracy to mixtures is computationally extremely demanding as the mixture enters the logarithm. The specific form of the approximated pd, cf. (3.28), which with the introduced symbols has the form

$$\begin{aligned}\mathsf{G}(x|\nu_{t-1},W_{t-1},x_t) &= (1-\beta_{t-1})\mathsf{F}(x|\nu_{t-1},W_{t-1})+\beta_{t-1}\delta(x-x_t)\\ \beta_{t-1} &= \frac{1}{\sum_{j\in\boldsymbol{j}}\nu_{t-1;j}+1},\end{aligned} \tag{3.33}$$

allows an efficient approximate minimisation, as outlined below.

The approximate minimisation relies on the fact that the jth approximated component $\mathsf{G}(x|\nu_{t-1;j},V_{t-1;j},x_t)$ of the updated mixture $\mathsf{G}(x|\nu_{t-1},W_{t-1},x_t)$ concentrates the majority of its probabilistic mass into a bounded part of the data-vectors space. The subsequent reorganisation of the updated mixture $\mathsf{G}(x|\nu_{t-1},W_{t-1},x_t)$ (3.33) allows to delimit these parts. It exploits definitions

$$\tilde{\beta}_{t;j} = c\mathsf{P}_{t;j},\;\; \mathsf{P}_{t;j} = \frac{\alpha_{t-1;j}\mathsf{F}(x_t|\nu_{t-1;j},V_{t-1;j})}{\sum_{j\in\boldsymbol{j}}\alpha_{t-1;j}\mathsf{F}(x_t|\nu_{t-1;j},V_{t-1;j})},\;\; \tilde{\alpha}_{t;j} = \alpha_{t-1;j}\frac{1-\beta_{t-1}}{1-\tilde{\beta}_{t;j}}, \tag{3.34}$$

where β_{t-1} is defined in (3.33). The scalar c is chosen as the solution of the equation

$$\frac{\beta_{t-1}}{(1-\beta_{t-1})} = \sum_{j\in\boldsymbol{j}}\alpha_{t-1;j}\frac{c\mathsf{P}_{t;j}}{1-c\mathsf{P}_{t;j}},\; c\in\Big(0,\big(\max_{j\in\boldsymbol{j}}\mathsf{P}_{t;j}\big)^{-1}\Big). \tag{3.35}$$

The right-hand side of (3.35) is monotonic, continuous and covers all possible non-negative values of the left-hand side when varying c in the considered interval. This guarantees solvability of (3.35) and provides $\tilde{\beta}_{t;j}\in(0,1]$, $\forall j\in\boldsymbol{j}$.

The definitions (3.34) and simple manipulations confirm that the update (3.33) can be given the form

$$\begin{aligned}\mathsf{G}(x|\nu_{t-1},W_{t-1},x_t) &= (1-\beta_{t-1})\mathsf{F}(x|\nu_{t-1},W_{t-1})+\beta_{t-1}\delta(x-x_t)\\ &= \sum_{j\in\boldsymbol{j}}\tilde{\alpha}_{t;j}\underbrace{\Big((1-\tilde{\beta}_{t;j})\mathsf{F}(x|\nu_{t-1;j},V_{t-1;j})+\tilde{\beta}_{t;j}\delta(x-x_t)\Big)}_{\mathsf{G}(x|\nu_{t-1;j},V_{t-1;j},x_t)}.\end{aligned} \tag{3.36}$$

The last expression in (3.36) interprets the Bayes update as an update of individual components with weights $\tilde{\beta}_{t;j}$, $j\in\boldsymbol{j}$ (3.34). The weight $\tilde{\beta}_{t;j}$ is chosen to be proportional to the posterior probability $\mathsf{P}_{t;j}$ that the observed data vector x_t was generated from the jth component.

Altogether, the decomposition (3.36) allows component-wise approximation, i.e. solution of a few (J) simple approximation tasks of the type (3.29).

Open Problem

Preliminary experiments confirm efficiency of the outlined mixture-based approximation. They also reveal that the computational approximation is sensitive to initialisation, as is true of any mixture estimation, [38]. A significant effort is needed to convert the idea into a reliable and feasible algorithmic solution.

Combination of the Approximation and Learning

The approximation step above prepares for the learning step, which involves the processing of new observations by Bayes' rule. Then, the projection onto the class of feasible approximate pds follows, and so forth. Here, the general warning applies that the approximate pd $\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t)$ (or their mixture-related counterparts $\mathsf{F}(x|\tilde{\nu}_{t;j},\tilde{V}_{t;j})$), given by (3.29), *should not* serve as the prior pd for the Bayesian updating step. The pd $\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t)$ is the good approximation of $\mathsf{G}(x|\nu_{t-1},V_{t-1},x_t)$ but not of the empirical pd $\mathsf{F}(x|k_t)$. Ignoring this fact may cause an accumulation of approximation errors over a sequence of combined learning and approximation steps. Gradually, this may cause a divergence of the approximate posterior pd from the best projection within the set $\{\mathsf{F}(x|\nu,V)\}_{\nu>0,V\in\boldsymbol{V}}$ or its mixture counterpart.

[51] completely characterises the class of non-sufficient statistics ν,V for which the error-accumulation problem does not arise. This class, consisting of time and data invariant finite-dimensional images of the logarithms of the parametric environment model, is, however, too narrow. To cover a richer class of problems, [36] proposes a countermeasure against error accumulation.

The proposed way assumes that (3.37) holds. It says that the best unknown approximation $\mathsf{F}(x|\nu_t,V_t)$ of the correct empirical pd $\mathsf{F}(x|k_t)$ is closer to the updated pd $\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t)$ than to the non-updated approximate pd $\mathsf{F}(x|\nu_{t-1},V_{t-1})$

$$\begin{aligned}\mathsf{F}(x|\nu_t,V_t)\in\mathbf{F} &= \left\{\mathsf{F}(x):\mathscr{D}(\mathsf{F}(x)||\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t))\le\mathscr{D}(\mathsf{F}(x)||\mathsf{F}(x|\nu_{t-1},V_{t-1}))\right\}\\ &= \left\{\mathsf{F}(x):\int_{\boldsymbol{x}}\mathsf{F}(x)\ln\left(\frac{\mathsf{F}(x|\nu_{t-1},V_{t-1})}{\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t)}\right)\mathrm{d}x\le 0.\right\}\end{aligned}\qquad(3.37)$$

This is the processed knowledge about the unknown best approximation $\mathsf{F}(x|\nu_t,V_t)$ of the exact empirical pd $\mathsf{F}(x|k_t)$. The pd $\mathsf{F}_0(x)=\mathsf{F}(x|\nu_{t-1},V_{t-1})$ is its available prior guess. The minimum KLD principle, see Subsection 3.2.3.3, extends this knowledge to the pd $\mathsf{F}(x|\nu_t,V_t)$, which should serve as the starting pd for the next learning step. Due to the linearity of the constraint (3.37) in the constructed pd $\mathsf{F}(x|\nu_t,V_t)$, the unique outcome of the minimum KLD principle can be found explicitly. It formally coincides with the outcome of stabilised forgetting, [53]. The functional form (3.30) is invariant under stabilised forgetting and the recommended pd $\mathsf{F}(x|\nu_t,V_t)$ to be used further on is given by

$$\nu_t=\lambda_t\tilde{\nu}_t+(1-\lambda_t)\nu_{t-1},\ V_t=\lambda_t\tilde{V}_t+(1-\lambda_t)V_{t-1}.\qquad(3.38)$$

The observation-dependent forgetting factor $\lambda_t\in[0,1]$ solves the equation obtained when the inequality (3.37) is replaced by equality

$$\int_{\boldsymbol{x}}\mathsf{F}(x|\lambda\tilde{\nu}_t+(1-\lambda)\nu_{t-1},\lambda\tilde{V}_t+(1-\lambda)V_{t-1})\ln\left(\frac{\mathsf{F}(x|\nu_{t-1},V_{t-1})}{\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t)}\right)\mathrm{d}x=0.$$

It has a solution in [0,1] as the left-hand side is continuous in λ, equals $\mathscr{D}(\mathsf{F}(x|\nu_{t-1},V_{t-1})||\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t))\ge 0$ for $\lambda=0$ and becomes $-\mathscr{D}(\mathsf{F}(x|\tilde{\nu}_t,\tilde{V}_t)||\mathsf{F}(x|\nu_{t-1},V_{t-1}))\le 0$ for $\lambda=1$.

Open Problem

The normative theory has still not provided to decision makers a unique, optimal theory for an unambiguous combination of learning and approximation. The above treatment has an intuitive appeal but it is not unique and as such it needs further justification or possibly modification. An extension to filtering – coping with more general state-space models – remains open.

3.3.3 Approximate Design of Strategy

The search for the optimal strategy, Proposition 1, suffers even a more pronounced curse of dimensionality than learning. The evaluated value function $-\ln(\gamma(k_t))$ (3.8) acts on the space $\boldsymbol{k}_t$, which has a high dimension. Its complexity grows with the amount of processed data, which calls for approximate techniques. This makes an approximation an integral part of the optimal strategy design. In spite of significant progress in related research, [5, 71], the normative theory still lacks a systematic methodology for approximate strategy design. This incompleteness of the theory leaves the decision maker insufficiently supported.

This section exploits the potential of FPD and outlines how the strategy design can be converted into a learning problem. This enhances the unifying features of FPD as it converts the approximate strategy design into approximate learning. The presentation is based on [35] and deals with a stationary version of FPD.

Agreement 3 (Stationary FPD and Stabilising Strategy) *The stationary FPD is delimited by the following conditions.*

- *The environment model* $\mathsf{M}(o_t|a_t,k_{t-1}) = \mathsf{M}(o_t|a_t,\phi_{t-1})$ *is a time-invariant function of the data vector* $x_t = (o_t,a_t,\phi_{t-1})$ *with a recursively updatable, finite-dimensional state vector* ϕ_{t-1} *while the updating rule* $o_t,a_t,\phi_{t-1} \to \phi_t$ *is also time invariant.*
- *The ideal environment model* $\mathsf{M}^\star(o_t|a_t,k_{t-1}) = \mathsf{M}^\star(o_t|a_t,\phi_{t-1})$ *and the ideal decisions rules* $\mathsf{S}^\star(a_t|k_{t-1}) = \mathsf{S}^\star(a_t|\phi_{t-1})$ *are time-invariant functions of the data vector and* (a_t,ϕ_{t-1})*, respectively.*
- *The decision horizon is unbounded,* $T \to \infty$*.*

Stationary FPD is meaningful if there is a stabilising strategy S_s *making* $\forall t \leq \infty$

$$\int_{\boldsymbol{o},\boldsymbol{a}} \mathsf{M}(o_t|a_t,\phi_{t-1})\mathsf{S}_s(a_t|k_{t-1}) \ln\left(\frac{\mathsf{M}(o_t|a_t,\phi_{t-1})\mathsf{S}_s(a_t|k_{t-1})}{\mathsf{M}^\star(o_t|a_t,\phi_{t-1})\mathsf{S}^\star(a_t|\phi_{t-1})}\right) \mathrm{d}o_t \mathrm{d}a_t \leq c_s < \infty,$$

for a finite constant c_s*.*

Proposition 2 (Solution of Stationary FPD) *Let the solved stationary FPD be meaningful, see Agreement 3. Then, the optimal FPD strategy* S^o *exists and is a stabilising strategy. The optimal strategy is stationary, i.e. it is formed by the time-invariant decision rules* $\mathsf{S}^o(a_t|\phi_{t-1})$*. It holds that*

$$
\begin{aligned}
\mathsf{S}^o(a_t|\phi_{t-1}) &= \frac{\mathsf{S}^\star(a_t|\phi_{t-1})\exp[-\omega(a_t,\phi_{t-1})]}{\gamma(\phi_{t-1})} \\
c+\omega(a_t,\phi_{t-1}) &= \int_{\boldsymbol{o}} \mathsf{M}(o_t|a_t,\phi_{t-1})\ln\left(\frac{\mathsf{M}(o_t|a_t,\phi_{t-1})}{\mathsf{M}^\star(o_t|a_t,\phi_{t-1})}\right)\mathrm{d}o_t \\
&\quad -\int_{\boldsymbol{o}} \mathsf{M}(o_t|a_t,\phi_{t-1})\underbrace{\ln\left(\int_{\boldsymbol{a}} \mathsf{S}^\star(a_{t+1}|\phi_t)\exp[-\omega(a_{t+1},\phi_t)]\mathrm{d}a_{t+1}\right)}_{\ln(\gamma(k_t=\phi_t))}\mathrm{d}o_t,
\end{aligned}
\tag{3.39}
$$

where $0 \le c \le c_s$.

Proof. The existence of the stationary strategy and the form of(3.39) follow from standard considerations concerning additive losses with unbounded horizon, [5]. □

Let us discuss the last term in (3.39), which makes this equation non-linear. It is the conditional expectation of the value function $-\ln(\gamma(k_t=\phi_t))$. The integral over $a_{t+1}\in\boldsymbol{a}$ defines it and can be expressed via the mean value theorem for integrals. It means that there is an $a_{t+1}(\phi_t)\in\boldsymbol{a}$ such that

$$
\int_{\boldsymbol{a}} \mathsf{S}^\star(a_{t+1}|\phi_t)\exp[-\omega(a_{t+1},\phi_t)]\mathrm{d}a_{t+1} = \exp[-\omega(a_{t+1}(\phi_t),\phi_t)]. \tag{3.40}
$$

The conditional expectation $\int_{\boldsymbol{o}} \bullet \mathsf{M}(o_t|a_t,\phi_{t-1})\mathrm{d}o_t$ is then expressed as the difference between the value of the argument and innovations $\varepsilon_t=\varepsilon_t(a_t,\phi_{t-1})$, as follows:

$$
\int_{\boldsymbol{o}} \omega(a_{t+1}(\phi_t),\phi_t)\mathsf{M}(o_t|a_t,\phi_{t-1})\mathrm{d}o_t = \omega(a_{t+1}(\phi_t),\phi_t)-\varepsilon_t. \tag{3.41}
$$

By construction, the innovations $(\varepsilon_t)_{t\in\boldsymbol{t}}$ are zero mean and mutually uncorrelated, [63]. This now permits the strategy design to be expressed as non-linear regression problem.

Proposition 3 (Conversion of Stationary FPD into Non-linear Regression) *Let the solved stationary FPD be meaningful, see Agreement 3. Let us parameterise the function* $\omega(a,\phi)$ *determining the optimal decision rule (3.39),*

$$
\omega(a,\phi) \approx \Omega(a,\phi,\Theta) \ \textit{for a finite-dimensional parameter}\ \Theta\in\boldsymbol{\Theta} \tag{3.42}
$$

and assume

$$
\omega(a_{t+1}=a,\phi_t) \approx \omega(a_{t+1}(\phi_t)=a,\phi_t) \underbrace{\approx}_{(3.42)} \Omega(a_{t+1}=a,\phi_t,\Theta). \tag{3.43}
$$

Then, the data $(a_{t+1},\phi_t)_{t\in\boldsymbol{t}}$ *and the unknown parameter* $\Theta\in\boldsymbol{\Theta}$ *are related by the following non-linear regression model*

$$\Omega(a_{t+1},\phi_t,\Theta) = c + \Omega(a_t,\phi_{t-1},\Theta) \qquad (3.44)$$
$$- \int_{\boldsymbol{o}} \mathsf{M}(o_t|a_t,\phi_{t-1}) \ln\left(\frac{\mathsf{M}(o_t|a_t,\phi_{t-1})}{\mathsf{M}^{\star}(o_t|a_t,\phi_{t-1})}\right) \mathrm{d}o_t + \varepsilon_t.$$

Proof. The (approximate) equality (3.44) follows directly by inserting (3.40), (3.41), and (3.42) into the second equality in (3.39). □

Note that the constant c and possible additional characteristics of the innovations, ε_t, e.g. their variance, are unknown and have to be estimated together with Θ.

Open Problem

The above paragraph outlines the basic idea of how to convert the equations describing the value function, and thus the optimal strategy, into a non-linear regression. The choice of the pd describing the innovations, needed for the design of the learning algorithm, should be done by the minimum KLD principle. Then, approximate learning, like that outlined in Section 3.3.2, can be used. It allows various functions $\Omega(a,\phi,\Theta)$ to be tested in parallel at relatively low computational cost. The conversion of this methodology into a full and reliable algorithm represents an open, but promising, direction, [35]. Close correspondence to established approximate techniques [71] may prove helpful in this effort.

Open Methodological Problem

A methodologically interesting question is why decision makers deviate from Bayesian DM. Various discussions, see e.g. [16], indicate that the emotionally biased attitude of DM experts in the team forming the decision maker strongly influence the choice, especially if the action can be postponed, [74]. The emotionally motivated choice of theoretical tools (Fisher vs. Bayesian statistics vs. fuzzy sets, etc.) should be avoided as much as possible as it introduces preferences unrelated to the preferences of the solved DM task. FPD tries to suppress this common, DM-quality-adverse, phenomenon by creating and offering a strong – axiomatically not emotionally – supported and widely applicable theory. However, the constraints on the overall deliberation effort spent by the decision maker on the solution of DM tasks inevitably induce a significant gap between the exploited theoretical tools and their advanced state. Thus, no attempt of this type can completely avoid personal biases of the human beings who are involved. The ongoing development of efficient ways of suppressing these biases is the challenge to be confronted.

3.4 Decision-Maker Induced Internals

This section discusses other important cases in which the decision maker directly contributes to the behaviour $b \in \boldsymbol{b}$. In particular, these cases deal with the deliberation effort needed for solving the DM task and with the role of the decision maker within a group of interacting decision makers.

3.4.1 Deliberation Effort and Sequential Decision Making

A real decision maker devotes a limited deliberation effort to any particular DM task. Taking account of the deliberation effort presents no theoretical problem if a hard bound on it is proposed and attained. Then, the attained solution is the only available option. Often, however, an additional effort can be expended in getting a higher DM quality and it is necessary to decide whether it is worthwhile to exert this effort or not.

Prominent works [68, 72, 73], and others, concluded that any attempt to include search for a compromise between the additional effort and DM quality into the optimal design leads to an infinite regress. Loosely, they claim that an extension of the loss of the solved DM task by a term penalising the deliberation effort increases the deliberation effort, which calls for an additional penalisation etc.

However, classical results on sequential DM, [81], indicate that this is generally untrue. The subsequent novel FPD version of sequential DM shows it.

In this presentation, iterative steps towards the strategy that solves the original DM task in the best way are interpreted as discrete time. Also, an additional *stopping action*

$$z_t \in \boldsymbol{z} \equiv \{1, 0\} \equiv \{\text{continue improvements, stop improvements}\}$$

is introduced. This complements the behaviour, giving $b = \big((o_t, a_t, z_t)_{t \geq 1}, k_0\big)$

$$= \big((\text{observation}_t, \text{original action}_t, \text{stopping action}_t)_{t \geq 1}, \text{prior knowledge}\big).$$

Within the original DM task, the quality of the decision strategy is evaluated by the ideal closed-loop model, which is the product of pds[16]

$$\mathsf{M}^\star(o_t|a_t, z_t = 1, k_{t-1})\mathsf{S}^\star(a_t|z_t = 1, k_{t-1}).$$

The ideal closed-loop model of the inspected DM task with stopping is specified by employing the leave-to-its-fate choice, Section 3.3.1,

[16] The condition $z_t = 1$ stresses that the optimisation *is* performed: it is not stopped.

$$\begin{aligned} \mathsf{C}^{\star}(o_t,a_t,z_t|k_{t-1}) &= \mathsf{M}^{\star}(o_t|a_t,z_t,k_{t-1})\mathsf{S}^{\star}(a_t|z_t,k_{t-1})\mathsf{S}^{\star}(z_t|k_{t-1}) \qquad (3.45)\\ \equiv \Big[\mathsf{M}^{\star}(o_t|a_t,z_t=1,k_{t-1})&\mathsf{S}^{\star}(a_t|z_t=1,k_{t-1})\mathsf{S}^{\star}(z_t=1|k_{t-1})\Big]^{z_t}\\ \times\Big[\mathsf{M}(o_t|a_t,k_{t-1})&\mathsf{S}(a_t|k_{t-1})(1-\mathsf{S}^{\star}(z_t=1|k_{t-1}))\Big]^{1-z_t}, \end{aligned}$$

where M is the considered environment model, S is the strategy optimised in the original DM task and the value $\mathsf{S}^{\star}(z_t=1|k_{t-1}) \in (0,1)$ quantifies the readiness for continuation of the search for the optimal strategy. It reflects the deliberation cost. The ideal pd (3.45) delimits a FPD counterpart of sequential DM, which often guarantees stopping in a finite time and thus avoids the infinite regress, [62].

Proposition 1 applied to the ideal pd (3.45) specialises to the next proposition.

Proposition 4 (FPD with Stopping) *With the ideal pd (3.45), the optimal randomised strategy* S^{o} *has the form*

$$\begin{aligned} \mathsf{S}^{o}(a_t|z_t=1,k_{t-1}) &= \frac{\mathsf{S}^{\star}(a_t|z_t=1,k_{t-1})\exp[-\omega(a_t,k_{t-1})]}{\rho(k_{t-1})} \qquad (3.46)\\ \rho(k_{t-1}) &= \int_{\boldsymbol{a}} \mathsf{S}^{\star}(a_t|z_t=1,k_{t-1})\exp[-\omega(a_t,k_{t-1})]\mathrm{d}a_t\\ \mathsf{S}^{o}(z_t=1|k_{t-1}) &= \mathsf{S}^{\star}(z_t=1|k_{t-1})\rho(k_{t-1})/\mathrm{e},\ \mathrm{e}=exp(1),\\ \gamma(k_{t-1}) &= \exp[-\mathsf{S}^{\star}(z_t=1|k_{t-1})\rho(k_{t-1})/\mathrm{e}]\\ \omega(a_t,k_{t-1}) &= \int_{\boldsymbol{o}} \mathsf{M}(o_t|a_t,k_{t-1})\ln\left(\frac{\mathsf{M}(o_t|a_t,k_{t-1})}{\mathsf{M}^{\star}(o_t|a_t,z_t=1,k_{t-1})\gamma(k_t)}\right)\mathrm{d}o_t. \end{aligned}$$

The evaluations (3.46) run backwards and the value function $-\ln(\gamma(k_t))$ *is zero at a priori specified hard upper bound on the decision horizon* $t=T$. *Only the values* $z_\tau=1$, $\tau<t$, *enter the knowledge* k_{t-1}.

Proof. Let us start at the last time t moment before stopping $t \le T$. The part of the KLD $\mathscr{D}(\mathsf{C}||\mathsf{C}^{\star})$ (3.6) influenced by the last optimized decision rule has the form

$$\begin{aligned} \mathsf{J}_t &= \mathsf{S}(z_t=1|k_{t-1})\\ &\times\Big\{\ln\Big(\frac{\mathsf{S}(z_t=1|k_{t-1})}{\mathsf{S}^{\star}(z_t=1|k_{t-1})}\Big)+\int_{\boldsymbol{a}}\mathsf{S}(a_t|z_t=1,k_{t-1})\times\Big[\ln\Big(\frac{\mathsf{S}(a_t|z_t=1,k_{t-1})}{\mathsf{S}^{\star}(a_t|z_t=1,k_{t-1})}\Big)\\ &+\underbrace{\int_{\boldsymbol{o}}\mathsf{M}(o_t|a_t,k_{t-1})\ln\Big(\frac{\mathsf{M}(o_t|a_t,k_{t-1})}{\mathsf{M}^{\star}(o_t|a_t,z_t=1,k_{t-1})\gamma(o_t,a_t,z_t=1,k_{t-1})},\Big)\mathrm{d}o_t\Big]}_{\omega(a_t,k_{t-1})}da_t\Big\} \end{aligned}$$

where $\gamma(k_t)=\gamma(o_t,a_t,z_t=1,k_{t-1})$ contains the knowledge realisations with $(z_\tau = 1)_{\tau\le t}$. This inductive assumption holds for the considered t as $\gamma(k_t)=1$. The rearrangement of the part depending on the optimised decision rule $\mathsf{S}(a_t|z_t=1,k_{t-1})$ and the fact that the KLD reaches its minimum for identical arguments gives the optimal factor of the decision rule

$$\mathsf{S}^o(a_t|z_t=1,k_{t-1}) = \frac{\mathsf{S}^\star(a_t|z_t=1,k_{t-1})\exp[-\omega(a_t,k_{t-1})]}{\rho(k_{t-1})}$$
$$\rho(k_{t-1}) = \int_{\boldsymbol{a}} \mathsf{S}^\star(a_t|z_t=1,k_{t-1})\exp[-\omega(a_t,k_{t-1})]\mathrm{d}a_t \in (0,1).$$

With this, it remains to minimise, over $\mathsf{S}(z_t=1|k_{t-1}) \in (0,1)$, the function

$$\min_{\{\mathsf{S}(a_t|z_t=1,k_{t-1})\}} \mathsf{J}_t = \mathsf{S}(z_t=1|k_{t-1})\ln\left(\frac{\mathsf{S}(z_t=1|k_{t-1})}{\mathsf{S}^\star(z_t=1|k_{t-1})\rho(k_{t-1})}\right).$$

Its minimiser and the reached minimum are

$$\mathsf{S}^o(z_t=1|k_{t-1}) = \mathsf{S}^\star(z_t=1|k_{t-1})\rho(k_{t-1})/\mathrm{e}$$
$$\text{and} \quad \min_{\{\mathsf{S}(a_t,z_t|k_{t-1})\}} \mathsf{J}_t = -\mathsf{S}^\star(z_t=1|k_{t-1})\rho(k_{t-1})/\mathrm{e} \equiv \ln(\gamma(k_{t-1})).$$

The last equality defines $\gamma(k_{t-1}) \leq 1$, which depends on the part of k_{t-1} entering $\mathsf{M}, \mathsf{M}^\star, \mathsf{S}^\star$, and containing only $z_\tau = 1$. The situation repeats for decreased t. □

Open Problem

The simple evaluation in the second equation in (3.46) represents the only increase of computational complexity per design step. It is compensated by an expected decrease of the number of design steps. The other (standard) computations are too complex and require approximation. The approximation that transforms the design of the strategy, Proposition 1, into a learning problem, see Subsection 3.3.3, may serve this purpose. It needs learning combined with approximation, discussed in Section 3.3.2. When it is well solved, the deliberation effort connected with optimisation will be under control. Even then, the open problem remains of how to control the deliberation effort in other, less formalised, parts of DM process.

3.4.2 The Decision-Maker's Role

DM complexity inevitable forces a division of tasks. The division requires knowledge fusion and possibly a search for a compromise between disparate individual decision-making preferences. The cooperation of affective robots [29] or the exploitation of crowd wisdom [67] are examples of this situation. The way of combining more knowledge pieces or DM preferences is strongly influenced by the purpose. Primarily, it is necessary to specify to whom the combination should serve. In other words, the role of an individual decision maker with respect to the group, within which the combination is to be performed, needs to be delimited. The influence of this specification is briefly discussed now.

Within FPD, knowledge and preferences are described by pds. This makes their combination formally similar. Essentially, a representative pd $\mathsf{H} = (\mathsf{H}(b))_{b \in \boldsymbol{b}}$ of a finite (possibly large) number of pds $\mathsf{G}_{\kappa} = (\mathsf{G}_{\kappa}(b))_{b \in \boldsymbol{b}}$, $\kappa \in \boldsymbol{\kappa}$, is to be constructed. There, the pd G_{κ} quantifies knowledge or preferences of the κth decision maker. The roles of the involved decision makers influence the choice of the representative, which describes the resulting combination. The following formalisation of role-dependent processing scenarios confirms this.

Selfish Scenario The representative is formed *for* a κth decision maker, $\kappa \in \boldsymbol{\kappa}$, offering the pd G_{κ}. The κth decision maker naturally takes her knowledge or preferences as adequate, and uses other group members, offering other G-pds, as important but complementary sources of knowledge or preferences. In harmony with the results on approximation of pds, [4, 40], the κth decision maker should use the KLD $\mathscr{D}(\mathsf{G}_{k}||\mathsf{H})$ as the proximity measure and, for instance, delimit the acceptable compromises (representatives)as being in the set

$$\{\mathsf{H} : \mathscr{D}(\mathsf{G}_{\kappa}||\mathsf{H}) \leq \text{a given, not-too-small, bound}\}\,. \qquad (3.47)$$

Then, the decision maker uses a prior guess H_0 of the compromise (representative) H and applies the minimum KLD principle.

[69] uses this scenario and arrives at a specific version of supra-Bayesian combination, [21], of given pds G_{κ}, $\kappa \in \boldsymbol{\kappa}$. The combination is a finite mixture of the combined pds G_{κ}, $\kappa \in \boldsymbol{\kappa}$, and H_0. The mixture weights are determined by conditions (3.47) and the prior guess H_0.

Group Scenario The representative H *serving the whole group* of decision makers is sought. By definition, the representative H reflects group knowledge or preferences in the best possible way. Thus, the individual pds G_{κ}, $\kappa \in \boldsymbol{\kappa}$, only approximate the group representative H. Then, the KLD $\mathscr{D}(\mathsf{H}||\mathsf{G}_{k})$ is the appropriate proximity measure and the analogy of (3.47) is

$$\{\mathsf{H} : \mathscr{D}(\mathsf{H}||\mathsf{G}_{\kappa}) \leq \text{a given, not-too-small, bound}\}\,. \qquad (3.48)$$

Then, the group uses a prior guess H_0 of the compromise H and apply the minimum KLD principle. The combination is now a weighted geometric mean of the pds G_{κ}, $\kappa \in \boldsymbol{\kappa}$, and H_0. The weights are determined by conditions (3.48) and the prior guess H_0. This variant of KLD is also used in approximate learning or it serves for extending incompletely specified G-pds, [42].

Asymmetry of the KLD well expresses asymmetry of the relation between an individual and the group in which she acts. The asymmetry implies that different representatives (compromises) are obtained in these two scenarios, even when both deal with the same pds $(\mathsf{G}_{\kappa})_{\kappa \in \boldsymbol{\kappa}}$, H_0. Thus, whenever the decision maker has freedom to delimit her role within the group, she influences closed-loop behaviour and consequently the solution of the addressed DM task.

Open Problem

The interpretation presented above reveals the controversial methodological dichotomy between the subjective and objective views of the world and DM. The group scenario appears as the objective one because it leads to a common (group) representative, while the selfish scenario (probably) fits better to DM, [68]. The situation appears to be simple, but obtaining operational guidelines is non-trivial. Crossing between dual versions of learning with forgetting, [37, 52], this illustrates practically.

3.5 Concluding Remarks

This text contributes to the applicability of fully probabilistic design of decision strategies (FPD), which is the normative decision-making theory that extends established Bayesian DM.

The chapter shows that a significant proportion of the observed discrepancies between the normative recommendations and real DM are caused by: i) neglecting an important part of the closed-loop behaviour; ii) differences between the claimed and actually respected DM preferences; and iii) incompleteness of FPD with respect to the complexity of the strategy design.

The main results are:

ad i) The extension of the closed-loop behaviour by a pointer to adequate complete DM preferences converts the hard preference-elicitation problem into a well-supported learning of the ideal pd, which quantifies DM preferences within FPD framework. The possibility to learn preferences systematically from decision-maker's actions is especially important.

ad ii) A specific construction of the ideal pd, known as the leave-to-its-fate option, models differences between claimed and respected DM preferences well. This insight can be used for analysing these differences.

ad iii) The applicability of FPD is enhanced by further development of methodology of approximate learning and strategy design. Also, controlling of the deliberation effort spent on DM design is embedded into *sequential* FPD.

This challenging of the unitary normative DM theory with needs of practical DM has proved to be quite fruitful. In addition to the results and open problems provided above, it has opened a pathway to consideration of an efficient human DM [79], to a unifying interpretation [14] of quantum mechanics, and to its use in DM [64].

Acknowledgements

The reported research has been supported by GAČR 13-13502S. Dr. Anthony Quinn from Trinity College, Dublin, has provided us useful feedback.

References

1. Barndorff-Nielsen, O.: Information and exponential families in statistical theory. Wiley, New York (1978)
2. Belda, K.: Probabilistically tuned LQ control for mechatronic applications (paper version). AT&P J., **9**(2 (2009)), 19–24 (2009)
3. Bellman, R.: Adaptive Control Processes. Princeton U. Press, NJ (1961)
4. Bernardo, J.: Expected information as expected utility. The Annals of Statistics **7**(3), 686–690 (1979)
5. Bertsekas, D.: Dynamic Programming and Optimal Control. Athena Scientific, US (2001)
6. Bohlin, T.: Interactive System Identification: Prospects and Pitfalls. Springer, NY (1991)
7. Boutilier, B.: A POMDP formulation of preference elicitation problems. In: Proc. of the 18th National Conference on AI, AAAI-2002, pp. 239–246. Edmonton, AB (2002)
8. Boutilier, C., Drummond, J., Lu, T.: Preference elicitation for social choice: A study in stable matching and voting. In: T. Guy, M. Kárný (eds.) Proc. of the 3rd International Workshop on Scalable Decision Making, ECML/PKDD 2013. ÚTIA AVČR, Prague (2013)
9. Campenhout, J.V., Cover, T.: Maximum entropy and conditional probability. IEEE Tran. on Inf. Theory **27**(4), 483–489 (1981)
10. Cappe, O., Godsill, S., Moulines, E.: An overview of existing methods and recent advances in sequential Monte Carlo. Proc. of the IEEE **95**(5), 899–924 (2007). DOI 10.1109/JPROC.2007.893250
11. Chen, L., Pu, P.: Survey of preference elicitation methods. Tech. Rep. IC/2004/67, Human Computer Interaction Group Ecole Politechnique Federale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (2004)
12. Conlisk, J.: Why bounded rationality? Journal of Economic Behavior & Organization **34**(2), 669–700 (1996)
13. Debreu, G.: Representation of a preference ordering by a numerical function. In: R. Thrall, C. Coombs, R. Davis (eds.) Decision Processes. Wiley, New York (1954)
14. DeWitt, B., Graham, N.: The Many-Worlds Interpretation of Quantum Mechanics. Princeton University Press (1973)
15. Dvurečenskij, A.: Gleasons Theorem and Its Applications, *Mathematics and Its Applications*, vol. 60. Kluwer Academic Publishers – Ister Science Press, Bratislava, Dordrecht/Boston/London (1993)
16. Efron, B.: Why isn't everyone a Bayesian. The American Statistician **40**(1), 1–11 (1986)
17. Feldbaum, A.: Theory of dual control. Autom. Remote Control **21**(9) (1960)
18. Ferguson, T.: A Bayesian analysis of some nonparametric problems. The Annals of Statistics **1**, 209–230 (1973)
19. Fiori, M., Lintas, A., Mesrobian, S., Villa, A.: Effect of emotion and personality on deviation from purely rational decision-making. In: T. Guy, M. Kárný, D. Wolpert (eds.) Decision Making and Imperfection, vol. 28, pp. 133–164. Springer, Berlin (2013). Studies in Computation Intelligence
20. Fishburn, P.: Utility Theory for Decision Making. J. Wiley, New York, London, Sydney, Toronto (1970)
21. Genest, C., Zidek, J.: Combining probability distributions: A critique and annotated bibliography. Statistical Science **1**(1), 114–148 (1986)
22. Giarlotta, A., Greco, S.: Necessary and possible preference structures. Journal of Mathematical Economics (2013)
23. Gong, J., Zhang, Y., Yang, Z., Huang, Y., Feng, J., Zhang, W.: The framing effect in medical decision-making: a review of the literature. Psychology, Health and Medicine **18**(6), 645–653 (2013)
24. Grigoroudis, E., Siskos, Y.: Customer Satisfaction Evaluation: Methods for Measuring and Implementing Service Quality. International Series in Operations Research and Management. Springer (2010)

25. Guan, P., Raginsky, M., Willett, R.: Online Markov decision processes with Kullback-Leibler control cost. IEEE Trans. on Automatic Control (2014)
26. Guy, T.V., Böhm, J., Kárný, M.: Probabilistic mixture control with multimodal target. In: J. Andrýsek, M. Kárný, J. Kracík (eds.) Multiple Participant Decision Making, pp. 89–98. Advanced Knowledge International, Adelaide (2004)
27. H.A.Simon: Models of Bounded Rationality. MacMillan, London (1997)
28. Haykin, S.: Neural Networks: A Comprehensive Foundation. Macmillan, New York (1994)
29. Insua, D., Esteban, P.: Designing societies of robots. In: T. Guy, M. Kárný (eds.) Proceedings of the 3rd International Workshop on Scalable Decision Making held in conjunction with ECML/PKDD 2013. ÚTIA AVČR, Prague (2013)
30. Jazwinski, A.: Stochastic Processes and Filtering Theory. Academic Press, New York (1970)
31. Jones, B.: Bounded rationality. Annual Rev. Polit. Sci. **2**, 297—-321 (1999)
32. Kahneman, D., Tversky, A.: The psychology of preferences. Scientific American **246**(1), 160–173 (1982)
33. Kárný, M.: Towards fully probabilistic control design. Automatica **32**(12), 1719–1722 (1996)
34. Kárný, M.: Automated preference elicitation for decision making. In: T. Guy, M. Kárný, D. Wolpert (eds.) Decision Making and Imperfection, vol. 474, pp. 65–99. Springer, Berlin (2013)
35. Kárný, M.: On approximate fully probabilistic design of decision making strategies. In: T. Guy, M. Kárný (eds.) Proceedings of the 3rd International Workshop on Scalable Decision Making, ECML/PKDD 2013. UTIA AV ČR, Prague (2013). ISBN 978-80-903834-8-7
36. Kárný, M.: Approximate Bayesian recursive estimation. Information Sciences (2014). DOI 10.1016/j.ins.2014.01.048
37. Kárný, M., Andrýsek, J.: Use of Kullback-Leibler divergence for forgetting. Int. J. of Adaptive Control and Signal Processing **23**(1), 1–15 (2009)
38. Kárný, M., Böhm, J., Guy, T.V., Jirsa, L., Nagy, I., Nedoma, P., Tesař, L.: Optimized Bayesian Dynamic Advising: Theory and Algorithms. Springer (2006)
39. Kárný, M., Guy, T.: Preference elicitation in fully probabilistic design of decision strategies. In: Proc. of the 49th IEEE Conference on Decision and Control (2010)
40. Kárný, M., Guy, T.: On support of imperfect Bayesian participants. In: T. Guy, M. Kárný, D. Wolpert (eds.) Decision Making with Imperfect Decision Makers, vol. 28. Springer, Berlin (2012). Intelligent Systems Reference Library
41. Kárný, M., Guy, T.V.: Fully probabilistic control design. Systems & Control Letters **55**(4), 259–265 (2006)
42. Kárný, M., Guy, T.V., Bodini, A., Ruggeri, F.: Cooperation via sharing of probabilistic information. Int. J. of Computational Intelligence Studies pp. 139–162 (2009)
43. Kárný, M., Halousková, A., Böhm, J., Kulhavý, R., Nedoma, P.: Design of linear quadratic adaptive control: Theory and algorithms for practice. Kybernetika **21** (1985). Supp. Nos 3–6
44. Kárný, M., Jeníček, T., Ottenheimer, W.: Contribution to prior tuning of LQG selftuners. Kybernetika **26**(2), 107–121 (1990)
45. Kárný, M., Kroupa, T.: Axiomatisation of fully probabilistic design. Information Sciences **186**(1), 105–113 (2012)
46. Kárný, M., Nedoma, P., Böhm, J.: On completion of probabilistic models. In: L. Berec, et al (eds.) Prep. of the 2nd European IEEE Workshop on Computer Intensive Methods in Control and Signal Processing, pp. 59–64. ÚTIA AVČR, Prague (1996)
47. Kerridge, D.: Inaccuracy and inference. J. of Royal Statistical Society **B 23**, 284–294 (1961)
48. Knejflová, Z., Avanesyan, G., T.V.Guy, Kárný, M.: What lies beneath players' non-rationality in ultimatum game? In: T. Guy, M. Kárný (eds.) Proceedings of the 3rd International Workshop on Scalable Decision Making, ECML/PKDD 2013. UTIA AV ČR, Prague (2013)
49. Knoll, M.A.: The role of behavioral economics and behavioral decision making in Americans' retirement savings decisions. Social Security Bulletin **70**(4) (2010)
50. Koopman, R.: On distributions admitting a sufficient statistic. Trans. of American Mathematical Society **39**, 399 (1936)

51. Kulhavý, R.: A Bayes-closed approximation of recursive nonlinear estimation. Int. J. Adaptive Control and Signal Processing **4**, 271–285 (1990)
52. Kulhavý, R., Kraus, F.J.: On duality of regularized exponential and linear forgetting. Automatica **32**, 1403–1415 (1996)
53. Kulhavý, R., Zarrop, M.B.: On a general concept of forgetting. Int. J. of Control **58**(4), 905–924 (1993)
54. Kullback, S., Leibler, R.: On information and sufficiency. Annals of Mathematical Statistics **22**, 79–87 (1951)
55. Kushner, H.: Stochastic approximation: a survey. Wiley Interdisciplinary Reviews: Computational Statistics **2**(1), 87–96 (2010). URL `http://dx.doi.org/10.1002/wics.57`
56. Landa, J., Wang, X.: Bounded rationality of economic man: Decision making under ecological, social, and institutional constraints. J. of Bioeconomics **3**, 217–235 (2001)
57. Lindley, D.: The future of statistics – A Bayesian 21st Century. Supp. Advances of Applied Probability **7**, 106 – 115 (1975)
58. Marczewski, E.: Sur l'extension de l'ordre partiel. Fundamental Mathematicae **16**, 386–389 (1930). In French
59. McCormick, T., Raftery, A.E., Madigan, D., Burd, R.: Dynamic logistic regression and dynamic model averaging for binary classification. Tech. rep., Columbia University (2010)
60. Meditch, J.: Stochastic Optimal Linear Estimation and Control. Mc. Graw Hill (1969)
61. Novák, M., Böhm, J.: Adaptive LQG controller tuning. In: M.H. Hamza (ed.) Proc. of the 22nd IASTED Int. Conference Modelling, Identification and Control. Acta Press, Calgary (2003)
62. Novikov, A.: Optimal sequential procedures with Bayes decision rules. Kybernetika **46**(4) (2010)
63. Peterka, V.: Bayesian system identification. In: P. Eykhoff (ed.) Trends and Progress in System Identification, pp. 239–304. Pergamon Press, Oxford (1981)
64. Pothos, E., Busemeyer, J.: A quantum probability explanation for violations of 'rational' decision theory. Proceedings of The Royal Society, B pp. 2171–2178 (2009)
65. Rao, M.: Measure Theory and Integration. John Wiley, NY (1987)
66. Regenwetter, M., Dana, J., Davis-Stober, C.: Transitivity of preferences. Psychological Review **118**(1), 42–56 (2011)
67. Roberts, S.: Scalable information aggregation from weak information sources. In: T. Guy, M. Kárný (eds.) Proceedings of the 3rd International Workshop on Scalable Decision Making held in conjunction with ECML/PKDD 2013. ÚTIA AVČR, Prague (2013)
68. Savage, L.: Foundations of Statistics. Wiley, NY (1954)
69. Sečkárová, V.: On supra-Bayesian weighted combination of available data determined by Kerridge inaccuracy and entropy. Pliska Stud. Math. Bulgar. **22**, 159–168 (2013)
70. Shore, J., Johnson, R.: Axiomatic derivation of the principle of maximum entropy & the principle of minimum cross-entropy. IEEE Tran. on Inf. Th. **26**(1), 26–37 (1980)
71. Si, J., Barto, A., Powell, W., Wunsch, D. (eds.): Handbook of Learning and Approximate Dynamic Programming. Wiley-IEEE Press, Danvers (2004)
72. Simon, H.: A behavioral model of rational choice. The Quarterly J. of Economics **LXIX**, 299–310 (1955)
73. Simon, H.: Theories of decision-making in economics and behavioral science. The American Economic Review **XLIX**, 253–283 (1959)
74. Syll, L.: Dutch books, money pump and Bayesianism (2012). http://larspsyll.wordpress.com/2012/06/25/dutch-books-money-pumps-and-bayesianism. Economics, Theory of Science & Methodology
75. Tishby, N.: Predictive information and the brain's internal time. In: T. Guy, M. Kárný (eds.) Proceedings of the 3rd International Workshop on Scalable Decision Making held in conjunction with ECML/PKDD 2013. ÚTIA AVČR, Prague (2013)
76. Tishby, N., Polani, D.: Information theory of decisions and actions. In: V. Cutsuridis, A. Hussain, J. Taylor (eds.) Perception-Action Cycle, Springer Series in Cognitive and Neural Systems, pp. 601–636. Springer, New York (2011)

77. Titterington, D., Smith, A., Makov, U.: Statistical Analysis of Finite Mixtures. John Wiley, New York (1985)
78. Todorov, E.: Linearly-solvable Markov decision problems. In: B. Schölkopf, et al (eds.) Advances in Neural Inf. Processing, pp. 1369 – 1376. MIT Press, NY (2006)
79. Tordesillas, R., Chaiken, S.: Thinking too much or too little? The effects of introspection on the decision-making process. Personality and Social Psychology Bulletin **25**, 623–629 (1999)
80. Tversky, A., Kahneman, D.: Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty **5**, 297–323 (1992)
81. Wald, A.: Statistical Decision Functions. John Wiley, New York, London (1950)