Fusion of Probability Density Functions

Günther Koliander*, Yousef El-Laham†, Petar M. Djurić†, and Franz Hlawatsch‡

*Acoustics Research Institute, Austrian Academy of Sciences (gkoliander@kfs.oeaw.ac.at)

†Department of Electrical and Computer Engineering, Stony Brook University ({petar.djuric, yousef.ellaham}@stonybrook.edu)

‡Institute of Telecommunications, TU Wien (franz.hlawatsch@tuwien.ac.at)

***Abstract*—Fusing probabilistic information is a fundamental task in signal and data processing with relevance to many fields of technology and science. In this work, we investigate the fusion of multiple probability density functions (pdfs) of a continuous random variable or vector. Although the case of continuous random variables and the problem of pdf fusion frequently arise in multisensor signal processing, statistical inference, and machine learning, a universally accepted method for pdf fusion does not exist. The diversity of approaches, perspectives, and solutions related to pdf fusion motivates a unified presentation of the theory and methodology of the field. We discuss three different approaches to fusing pdfs. In the axiomatic approach, the fusion rule is defined indirectly by a set of properties (axioms). In the optimization approach, it is the result of minimizing an objective function that involves an information-theoretic divergence or a distance measure. In the supra-Bayesian approach, the fusion center interprets the pdfs to be fused as random observations. Our work is partly a survey, reviewing in a structured and coherent fashion many of the concepts and methods that have been developed in the literature. In addition, we present new results for each of the three approaches. Our original contributions include new fusion rules, axioms, and axiomatic and optimization-based characterizations; a new formulation of supra-Bayesian fusion in terms of finite-dimensional parametrizations; and a study of supra-Bayesian fusion of posterior pdfs for linear Gaussian models.**

***Index Terms*—Information fusion, probabilistic opinion pooling, pooling function, multisensor signal processing, sensor network, model averaging, supra-Bayesian fusion, Kullback-Leibler divergence, Chernoff fusion, α-divergence, Hölder mean, linear Gaussian model, covariance intersection.**

I. Introduction

The fusion of multiple probabilistic descriptions of a random quantity is a fundamental task with applications in many fields including multisensor signal processing [1]–[8], machine learning [9]–[12], robotics [4], smart environments [13], medicine [14], transportation [15], precision agriculture [16], pharmacology [17], weather forecasting [18], [19], economics [20], [21], and financial engineering [22]. While this task has been studied for several decades, an in-depth treatment with a focus on continuous random variables and, accordingly, on the fusion of probability density functions (pdfs) appears to be lacking. The present paper attempts to fill this gap. Our focus on continuous random variables is motivated by the fact that continuously distributed quantities are the primary object of interest in many applications.

The fusion of pdfs can be considered in different contexts, and several different techniques for this task have been proposed in the literature. Our treatment is partly a survey of existing concepts and techniques, with an emphasis on a structured and coherent presentation. In addition, we present numerous original contributions related to axiomatic, optimization-based, and Bayesian approaches to pdf fusion.

This work was supported in part by the Vienna Science and Technology Fund (WWTF) under grant MA16-053, by the Austrian Science Fund (FWF) under grants P32055-N31 and Y1199, and by the Growing Convergence Research Program of the National Science Foundation under grant 2021002.

A. Motivation

The field of pdf fusion is multifaceted and somewhat fuzzy: there are many possible approaches to the problem of finding a pdf fusion rule, and there is no universally accepted measure of performance [23], [24]. An appropriate fusion rule and performance measure depend on the scenario and application. This situation can be aggravated by the fact that different fusion rules can lead to very different results.

Although in specific applications certain pdf fusion rules have been established and found to be useful, the rationales of these rules and their possible alternatives are not always obvious. Thus, it is both theoretically interesting and practically relevant to study the problem of pdf fusion and the existing viewpoints and solutions in a general way that abstracts from specific applications, and to put these viewpoints and solutions into a higher-level perspective. Our hope is that this analysis will support an informed choice of a pdf fusion rule for specific scenarios and applications. Accordingly, rather than considering a single framework or method for pdf fusion, this paper reviews the different approaches that have been developed over several decades in different disciplines and by different communities. In addition, these approaches are categorized into three fundamental approaches to principled pdf fusion, which we term the axiomatic, optimization, and supra-Bayesian approaches.

Fusing pdfs is a special variant of the general task of "data fusion" or "information fusion," and one may ask why it can be advantageous to perform data/information fusion at the level of pdfs. Possible answers include the following [20], [24], [25]:

- A pdf constitutes a *complete* probabilistic description of a continuous random variable or random vector. In addition to its mean or its mode (which can be used as point estimates of the random variable or vector), this description includes further important information such as effective support, multimodality, tail decay, and a detailed characterization of the "dispersion" around the mean. Moreover, it enables the calculation of quantitative measures of the accuracy of point estimates.

- A pdf provides a standardized and "genesis-agnostic" representation of the state of information of an agent or sensor, i.e., it abstracts from the intricacies of the processing employed by the agent or sensor to obtain it from the raw data. This "no questions asked" characteristic enables or facilitates an information fusion even between heterogeneous agents, which employ different sensing modalities and/or different types of data preprocessing. Furthermore, the lack of a transparent relation to the raw data is a desirable feature in privacy-sensitive applications.
- Because a pdf provides a standardized, genesis-agnostic representation, pdf fusion is well suited to a decentralized (peer-to-peer) network topology. In decentralized, possibly ad-hoc networks, a distributed in-network type of processing is used where each agent communicates with a limited set of neighboring agents and, typically, little or no information about the characteristics of far-away agents is available locally. The pdf format here facilitates the dissemination of information through the network.
- Computationally efficient pdf fusion algorithms based on parameteric pdf representations are available. For example, the fusion of Gaussian pdfs reduces to fusing the corresponding means and variances or covariance matrices. More generally, there are efficient algorithms for fusing Gaussian mixture pdfs. In distributed implementations, parametric pdf representations enable pdf fusion with low or moderate communication cost. Thus, pdf fusion is attractive because detailed probabilistic information can be fused with moderate complexity in terms of computation and communication.

B. Probabilistic Opinion Pooling

Consider K "agents," "experts," or "models," each providing an "opinion" about an unknown random object that may be a scalar or vector. In the probabilistic setting studied in this work, the opinions provided by the agents are not point estimates of the random object but probability distributions. More specifically, we focus on the case of a continuous random variable or vector $\boldsymbol{\theta}$, where the opinion of agent k is expressed by a pdf $q_k(\boldsymbol{\theta})$.

The problem studied in this paper is to combine, or fuse, the pdfs of the K agents, $q_k(\boldsymbol{\theta})$ for $k = 1, 2, \ldots, K$, into an aggregate pdf $q(\boldsymbol{\theta})$. This problem is traditionally referred to as *probabilistic opinion pooling*, although that term is also used for the fusion of discrete (categorical) distributions. We assume that the combination of the agent pdfs $q_k(\boldsymbol{\theta})$ is done by a central agent or unit, termed a "fusion center," which has access to all the agent pdfs. The function employed by the fusion center to map the $q_k(\boldsymbol{\theta})$ into the aggregate (fused) pdf $q(\boldsymbol{\theta})$ is termed a *fusion rule* or a *pooling function*. Many different pooling functions have been proposed in the literature, based on various models and considerations. Important examples include the linear pooling function (a weighted arithmetic mean, also known as arithmetic mean density) [26], [27] and the log-linear pooling function (a weighted geometric mean, also referred to as Chernoff fusion or geometric mean density) [27]–[31]. For Gaussian pdfs, the covariance intersection technique [29], [32] is an instance of a log-linear pooling function. These and several other pooling functions will be discussed in later sections.

An alternative to the centralized setting for probabilistic opinion pooling described above would be a decentralized network of agents without a dedicated fusion center [8], [30], [33], [34]. Here, the agents communicate their pdfs only locally, i.e., to neighboring agents, and each agent can be considered to act as a local fusion center. In this "in-network" or "network-centric" type of probabilistic opinion pooling, the agents use a distributed communication-and-fusion protocol, such as flooding, consensus, gossip, or diffusion, to disseminate their local pdfs through the network and emulate a given overall pooling function. This relies on a suitable pdf representation such as a Gaussian, Gaussian mixture, or particle representation. The fusion methods we discuss in this work are also relevant to decentralized probabilistic opinion pooling. We note, however, that there are numerous methods for in-network signal and information processing in which the local processing results that are being combined are not pdfs. For example, some methods combine local likelihood functions [1], [35], [36] or messages within a message passing algorithm such as belief propagation [7], [37], or certain iterated quantities within a networkwide adaptation-diffusion procedure [38], to name a few.

C. Relevance and Applications

Probabilistic opinion pooling is a fundamental and elementary functionality with widespread applications. Historically, the first motivation was to combine expert opinions into an aggregate opinion [26]. Nowadays it is more likely that the different probability distributions do not represent the opinions of multiple experts but originate from the use of multiple sensors, models, or data sets. In particular, probabilistic opinion pooling is often formulated in a Bayesian setting as the fusion of local *posterior* pdfs that are produced by multiple agents using local implementations of Bayesian inference [39]. The ideal aggregate pdf here is the global posterior pdf, which takes into account all the data available to the agents. However, the calculation of the global posterior pdf generally requires additional knowledge besides the local posterior pdfs, such as the local likelihood functions, the prior pdfs used by the agents, and possible statistical dependencies between the agents. By contrast, probabilistic opinion pooling requires only the local posterior pdfs. In many settings, it is easily and widely applicable because it does not make any assumptions about the local inference methods, the types of the sensors, or the nature of the local data, which can all be different at different agents.

From the viewpoint of the processed data, there is a wide range of scenarios for probabilistic opinion pooling. Two extreme cases are particularly important: all the agents process different data, or they process exactly the same data. Furthermore, the processing may be carried out with completely unrelated models but with the same objective (e.g., predicting future observations or classifying observations).

Current applications of probabilistic opinion pooling include, but are not limited to, the following selection:

- In multisensor signal processing applications of probabilistic opinion pooling, multiple sensors derive local pdfs based on local observations and either submit these pdfs (or finite-dimensional representations thereof) to a fusion center or fuse them in a distributed, peer-to-peer manner [1], [3], [27], [30]–[33], [39]–[42]. In particular, probabilistic opinion pooling plays an important role in multisensor target tracking [2], [3], [5], [6], [8], [30], [43]–[46]. For tracking an unknown number of targets, probabilistic opinion pooling has recently also been applied to the "multiobject" pdfs or to the probability hypothesis densities (i.e., the "densities" of the first moment measures) of finite point processes, also known as random finite sets [2], [5], [6], [8], [43]–[50]. Although in this work we do not consider finite point processes, much of our discussion is also relevant in that domain. The application of probabilistic opinion pooling to multisensor target tracking will be discussed in more detail in Section II-A.
- In probabilistic machine learning, several scenarios suggest the combination of probability distributions. For example, the concept of ensemble learning [9] is based on applying multiple learning algorithms whose outputs are combined to obtain an aggregate result that is more accurate than that of any of the individual learning algorithms in the ensemble. Furthermore, in federated learning [12], [51], [52], multiple edge devices learn statistical models individually from their local data sets without explicitly exchanging these data sets, and a fusion center aggregates the learned models without having access to the original data. This is attractive for privacy-sensitive applications, since no private data have to be shared. More details on probabilistic machine learning are provided in Section II-B.
- The main goal in the combination of forecasts [53], [54] is the estimation of a parameter by combining several different models. To this end, certain methods perform a fusion of pdfs and usually refer to it as "combining density forecasts" [20], [55], [56]. This application will be addressed in more detail in Section II-C.
- In Bayesian model averaging, several different models are used to derive different posterior pdfs based on the same data [57], [58]. An aggregate pdf is derived as a weighted average of the individual pdfs, where the weights are given by the posterior probabilities of the models. Bayesian model averaging has been widely used in phylogenetics [59], [60], economics [21], [61], ecology [62], and many other fields [63], [64].
- Traditional implementations of Monte Carlo-based inference schemes do not easily scale to large data sets ("big data"). A common expedient then is to partition the data set into subsets and obtain a partial posterior pdf approximation for each subset. The partial approximations are subsequently fused into an approximation of the overall posterior pdf, which, thereby, takes into account the full data set [65]–[67]. More details on this application are given in Section II-B. Another approach [68], [69] directly fuses sample representations of distributions by interpreting these samples as a weighted sum of Dirac measures.

To focus the scope of the present work, we assume for the most part that the fusion center does not have any additional data about the random vector $\boldsymbol{\theta}$ beyond the pdfs provided by the agents. (Here, an exception is given by the supra-Bayesian setting studied in Sections VIII and IX, where we assume that the fusion center knows a statistical model related to $\boldsymbol{\theta}$.) In particular, the fusion center cannot access any training data that were used by the agents, e.g., to derive a global posterior pdf, and it does not have any validation data that it could use to validate the agents' pdfs. Thus, although the fundamental problems are similar, we will not consider several ensemble learning methods such as stacking [70], [71] or many other machine learning settings related to probabilistic fusion [72]–[74]. Furthermore, given our focus on pdfs rather than discrete probability distributions, we will not touch upon methods tailored to the combination of classifiers, another large and growing field [75]. Finally, we are interested in obtaining a pdf and not merely a point estimate of $\boldsymbol{\theta}$. This is motivated by the fact that the pdf of $\boldsymbol{\theta}$ contains all the probabilistic information about $\boldsymbol{\theta}$ and can thus be used to obtain point estimates or other types of statistics. Hence, certain works on multimodel inference [76] and the combination of forecasts [53], [54], [77] share some ideas with the present work but ultimately have a different focus.

D. Approaches to Probabilistic Opinion Pooling

Although the probabilistic opinion pooling problem may appear simple and elementary, no single pooling function is universally accepted or uniformly best. Generally speaking, we would like the pooling function to involve the agent pdfs $q_k(\boldsymbol{\theta})$ in a way that follows some rationale. This rationale and the resulting choice of a pooling function may depend on the overall problem setting, application-specific aspects, side constraints, additional information available to the fusion center, and other considerations. The probabilistic opinion pooling problem has been studied for many decades, and substantial research efforts have been dedicated to the definition or derivation of pooling functions. One of the earliest works is [26], where the linear pooling function was introduced. Several survey articles on probabilistic opinion pooling with detailed literature reviews have been published [23], [78]–[80], however often with a focus on discrete random variables.

In this work, we consider three principled approaches to defining a pooling function for pdfs. In what we call the *axiomatic approach*, the pooling function is defined indirectly by a set of properties (axioms) that it is required to satisfy. For example, it may be reasonable to require that the aggregate pdf $q(\boldsymbol{\theta})$ does not depend on the indexing order of the agent pdfs $q_k(\boldsymbol{\theta})$, or that for equal $q_k(\boldsymbol{\theta})$—i.e., unanimity among all the agents—the aggregate pdf $q(\boldsymbol{\theta})$ conforms to that unanimous opinion. Most of the early literature in the field was dedicated to the axiomatic approach [23], [80]. An axiomatic approach is also adopted in the literature based on *imprecise probabilities* [81]. There, the idea is to define pooling operators that map from the agents' probability mass functions (pmfs) to a set of pmfs rather than a single pmf. To the best of our knowledge, the concept of imprecise probabilities has so far been considered only for discrete probability spaces [81], [82].

In the *optimization approach*, the pooling function is the result of an optimization, i.e., the minimization or maximization of an objective function. Usually, the idea is that the aggregate pdf $q(\boldsymbol{\theta})$ should be as close as possible to all the agent pdfs $q_k(\boldsymbol{\theta})$ simultaneously. This can be formulated as a minimization involving an information-theoretic divergence [83]–[85] or a distance measure [84], [86]. The resulting optimum $q(\boldsymbol{\theta})$ can typically be interpreted as an "average" of the $q_k(\boldsymbol{\theta})$.

Finally, the *supra-Bayesian approach* considers the fusion center as a Bayesian observer that interprets the agent pdfs $q_k(\boldsymbol{\theta})$ as random observations. This Bayesian observer builds on additional information about the dependence of these pdfs on $\boldsymbol{\theta}$ (represented by the conditional probability distribution $p(q_1, \ldots, q_K \,|\, \boldsymbol{\theta})$ of the random functions $q_1, \ldots, q_K$ given $\boldsymbol{\theta}$) to calculate a posterior pdf, which then constitutes the fusion result [87], [88]. Most of the early literature [89]–[91] describes $p(q_1, \ldots, q_K \,|\, \boldsymbol{\theta})$ implicitly by assuming that the joint distribution of the errors $\boldsymbol{\mu}_k - \boldsymbol{\theta}$ (where $\boldsymbol{\mu}_k$ is the expectation of $\boldsymbol{\theta}$ induced by the pdf q_k) is multivariate Gaussian. This reduces the fusion problem to the calculation of the posterior pdf for a simple Bayesian linear Gaussian model where the $\boldsymbol{\mu}_k$ are treated as observations at the fusion center and the covariance structure is known. The practically most important scenario in the supra-Bayesian approach is where each agent has access to certain random observations that are statistically dependent on the random vector $\boldsymbol{\theta}$, and both the agents and the fusion center have knowledge of a prior distribution of $\boldsymbol{\theta}$ and of the local likelihood functions of the agents. The agent pdf $q_k(\boldsymbol{\theta})$ is here given by the agent's local posterior pdf. The fusion center is also aware of any statistical dependencies between the observations of different agents, which are described by a global likelihood function.

E. Contributions and Paper Organization

The diversity of approaches, perspectives, and solutions related to probabilistic opinion pooling motivates a survey that presents the theory and methodology of the field in a coherent manner. The present paper attempts to answer this call. In addition, it provides a number of original contributions and results, including the following:

- A rigorous and coherent treatment of probabilistic opinion pooling for a *continuous* random vector $\boldsymbol{\theta}$ and, accordingly, for the fusion of pdfs. In particular, for the first time, the axiomatic approach is rigorously and thoroughly discussed for pdfs (Section IV). So far, the focus in the literature has mostly been on discrete probability distributions, and it has been claimed that analogous results hold for pdfs. Although this is indeed often the case, the non-atomic structure of the pdf setting sometimes allows for stronger or different results.
- The definition of a new pooling function, referred to as "generalized multiplicative pooling function" (Section III-B8).
- Two new axioms for pooling functions, referred to as "factorization preservation" and "generalized Bayesianity" (Axioms 9 and 12 in Section IV-A).
- Several new theorems presenting axiomatic characterizations of pooling functions for pdfs and related results (Theorems 1, 2, and 6–11 in Section IV-B and Appendices A–F). These theorems are partly adaptations of existing results formulated for discrete probability distributions and partly entirely new results.
- Proofs of the following results: the pooling function minimizing the weighted sum of α-divergences is given by the weighted Hölder mean; the pooling function minimizing the weighted sum of Pearson χ^2-divergences is given by the weighted harmonic mean; the pooling function minimizing the weighted sum of L_2 distances is given by the weighted arithmetic mean (Theorems 14 and 16 in Sections V-C and V-E and Appendices G and I). Furthermore, we derive the solution to the problem of minimizing a general class of weighted symmetric distance functions (Theorem 17 in Section V-E and Appendix J).
- A new framework of supra-Bayesian fusion of posterior pdfs in terms of finite-dimensional "local statistics" (Sections VIII-B through VIII-D). This includes an explicit pooling function for the case of agents collecting conditionally independent observations (Theorem 18 in Section VIII-A), a formal definition of and result for finite-dimensional supra-Bayesian fusion (Definition 1 and Theorem 19 in Section VIII-B), and a general procedure for establishing a fusion rule for the case of agents collecting conditionally dependent observations (Section VIII-D).
- A detailed study of supra-Bayesian fusion of posterior pdfs for linear Gaussian models (Section IX), including the derivation of explicit pooling functions and fusion rules (Sections IX-C and IX-D, Appendices K and L).

The paper's structure is as follows. In Section II, we illustrate the applicability and relevance of probabilistic opinion pooling by discussing three specific example applications. In Section III, we formulate the probabilistic opinion pooling problem for pdfs and present a collection of specific pooling functions. Section IV discusses the axiomatic approach to opinion pooling and provides several new characterization theorems. In Section V, we consider the optimization approach to opinion pooling. We describe various optimization criteria and show that they partly lead to the same pooling functions as the axiomatic approach and partly to different pooling functions such as the family of Hölder means. The fusion of Gaussian distributions using the pooling functions from Sections III and V is considered in Section VI. Section VII addresses the choice of the weights involved in the two most prominent and popular pooling functions, namely, the linear and log-linear pooling functions, as well as the choice of the parameter involved in the Hölder pooling function. In Section VIII, we present a new view of the supra-Bayesian pooling approach using finite-dimensional parametrizations. The results of Section VIII are specialized to linear Gaussian models in Section IX. The model of Section IX includes as a special case the supra-Bayesian setting presented in [89]–[91]. We broaden this setting significantly and present detailed fusion rules. In Section X, we provide suggestions for future research, and in Section XI, a summary of our main insights and results. Detailed proofs of our main results are provided in several appendices.

F. Notation

We will use the following basic notation. Vectors are denoted by boldface lower-case letters (e.g., $\mathbf{t}$ and $\boldsymbol{\theta}$), matrices by boldface upper-case letters (e.g., $\mathbf{H}$ and $\boldsymbol{\Sigma}$), and sets and events by calligraphic letters (e.g., $\mathcal{A}$). The transpose is written as $(\cdot)^{\intercal}$. We write $\mathbf{I}_d$ for the identity matrix of dimension d, $\mathbf{0}_{d_1 \times d_2}$ for the $d_1 \times d_2$ zero matrix, $\mathbf{1}_d$ for the all-one vector of dimension d, and $\otimes$ for the Kronecker product. The symbol $\mathcal{P}$ denotes the set of all pdfs, and $\mathcal{S}_K$ denotes the probability simplex on $[0,1]^K$, i.e., the set of all $(w_1,\ldots,w_K) \in [0,1]^K$ with $\sum_{k=1}^{K} w_k = 1$. For a set or event $\mathcal{A}$, we denote the complement as $\mathcal{A}^c$, the indicator function as $\mathbb{1}_{\mathcal{A}}$, and the Lebesgue measure as $|\mathcal{A}|$. Further notation is listed in Table I.

Probabilistic opinion pooling		
$q_k(\boldsymbol{\theta})$	—	pdf of agent k
$q(\boldsymbol{\theta})$	—	aggregate (fused) pdf
$Q_k(\mathcal{A})$	—	probability of event $\mathcal{A}$ according to $q_k(\boldsymbol{\theta})$
$Q(\mathcal{A})$	—	probability of event $\mathcal{A}$ according to $q(\boldsymbol{\theta})$
$\boldsymbol{\mu}_{q_k}$	—	mean associated with $q_k(\boldsymbol{\theta})$
$\boldsymbol{\mu}_q$	—	mean associated with $q(\boldsymbol{\theta})$
$\boldsymbol{\Sigma}_{q_k}$	—	covariance matrix associated with $q_k(\boldsymbol{\theta})$
$\boldsymbol{\Sigma}_q$	—	covariance matrix associated with $q(\boldsymbol{\theta})$
Supra-Bayesian framework		
$\mathbf{y}_k$	—	local observation vector of agent k
$\mathbf{y}$	—	global observation vector (stacking all $\mathbf{y}_k$)
$\mathbf{t}_k$	—	local statistic of agent k
$\mathbf{t}$	—	stacked vector of all local statistics $\mathbf{t}_k$
$p(\boldsymbol{\theta})$	—	prior pdf
$\ell_k(\boldsymbol{\theta})$	—	local observation likelihood function of agent k
$\ell(\boldsymbol{\theta})$	—	global observation likelihood function
$\lambda_k(\boldsymbol{\theta})$	—	local $\mathbf{t}_k$-likelihood function of agent k
$\lambda(\boldsymbol{\theta})$	—	global $\mathbf{t}$-likelihood function
$\pi_k(\boldsymbol{\theta})$	—	local posterior pdf of agent k
General notation		
$g[\cdot]$	—	pooling function
$g[q_1,\ldots,q_K](\boldsymbol{\theta})$	—	fused pdf resulting from application of pooling function g to pdfs $q_1(\boldsymbol{\theta}),\ldots,q_K(\boldsymbol{\theta})$
$\mathbb{E}_{\psi}[\cdot]$	—	expectation operator with respect to pdf $\psi(\boldsymbol{\theta})$
$\mathbb{E}[\cdot]$	—	expectation operator with respect to the joint pdf of all involved random variables
$\mathcal{N}(\boldsymbol{\theta};\boldsymbol{\mu},\boldsymbol{\Sigma})$	—	pdf of a Gaussian random vector $\boldsymbol{\theta}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

Table I: Notation

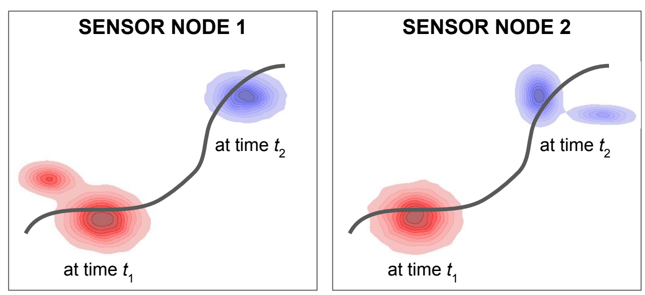


Figure 1: Schematic illustration of the state trajectory of a target and the local posterior pdfs of two sensor nodes at two different time steps.

II. Illustrative Applications

To illustrate the broad applicability of probabilistic opinion pooling or, more concretely, of the fusion of pdfs, we consider three illustrative applications in more detail.

A. Target Tracking

Target tracking aims to estimate the time-varying state (e.g., position and velocity) of a “target” from a sequence of observations [92], [93]. Applications include aeronautical and maritime situational awareness, surveillance, autonomous driving, biomedical analytics, remote sensing, and robotics. The performance of target tracking can be enhanced by using multiple sensors. This can be done in an optimal manner if the multisensor observation model is completely known, including possible statistical dependencies between the observations. However, in many cases, a simplified approach to multisensor target tracking based on probabilistic opinion pooling is adopted. Each sensor node operates a Bayesian filter that, at each time step, calculates a local posterior pdf of the current state based solely on the observation of that sensor. Fig. 1 illustrates the local posterior pdfs of two sensor nodes at two different time steps. The local posterior pdfs of the various sensor nodes are then fused using, typically, log-linear pooling or its second-order version known as covariance intersection [3], [27], [29]–[33], [40], [41], [94] (see Sections III-B3 and VI-B). This approach is practically convenient because (i) the multisensor fusion is decoupled from the filtering, and (ii) it works for any choice of Bayesian filter methods used at the sensor nodes and for any sensing modalities, even when they are different at different sensor nodes. These characteristics make the probabilistic opinion pooling approach well suited to heterogeneous and/or decentralized sensor networks.

A nontrivial extension of target tracking is *multitarget* tracking, which involves an unknown time-varying number of targets and a more complicated observation model [95]–[102]. More specifically, targets can appear and disappear randomly, and there are missed detections (i.e., some sensors do not produce observations for some of the targets), clutter or false-alarm observations (which are not related to any target), and an observation-origin uncertainty (i.e., the sensor nodes do not know whether a given observation originated from a target, and from which target, or is clutter). Probabilistic opinion pooling can be used both for "vector-based" multitarget tracking methods, which describe the joint state of the targets by a random vector, and for "set-based" methods, which describe it by a random finite set or equivalently a finite point process [97], [100], [103]. In the vector-based case, the target states are fused individually using, typically, log-linear pooling or covariance intersection. This presupposes an association of the target states across the sensors [104], [105].

In set-based methods, on the other hand, probabilistic opinion pooling is applied either to the posterior *multiobject pdfs* or to the posterior *probability hypothesis densities* (PHDs) of the sensor nodes, which provide two alternative *joint* descriptions of all the target states [97], [100]. Here, both log-linear pooling—also termed geometric average fusion, exponential mixture density, generalized covariance intersection, or Kullback-Leibler averaging [2], [5], [8], [43], [50], [85], [106]–[108]—and linear pooling (see Section III-B1)—also termed arithmetic average fusion and minimum information loss fusion [5], [43]–[46], [48], [85], [108]–[111]—have been used. Log-linear pooling is more sensitive to missed detections whereas linear pooling is more sensitive to clutter. Regarding this sensitivity tradeoff, we note that pooling functions that are intermediate between the linear and log-linear ones are provided by the family of Hölder pooling functions to be presented in Section III-B5.

Finally, both log-linear and linear pooling have recently been generalized to multitarget tracking methods based on *labeled* random finite sets, which track the identities of the targets in addition to their states [6], [47], [49], [112]–[114]. Some of these methods require a label association step that is similar in spirit to the target association step required by vector-based methods [47], [49], [113], [114].

B. Probabilistic Machine Learning

Probabilistic machine learning [115], [116] has recently seen applications in many different areas including quantum molecular dynamics [117], disease detection [118], medical diagnosis [119], scene understanding [120], and geotechnical engineering [121]. In machine learning, *uncertainty quantification* for predictive models is required for problems that involve risk assessment. Unfortunately, classical machine learning models do not account for parameter uncertainty, which makes them more susceptible to failure when dealing with unseen and/or unrelated data [122]. This is a prominent issue for deep learning models [123]. One way to account for predictive uncertainty in machine learning is to adopt a Bayesian framework: using training data, a prior pdf over the model parameters is updated to obtain a posterior pdf. This posterior pdf is then used to calculate a predictive pdf for unobserved data (test data). This pdf is often represented in parametric form—e.g., a Gaussian pdf is parameterized by its mean and covariance matrix—or by a set of samples. Examples of Bayesian machine learning models include Bayesian linear regression, Bayesian neural networks [124], [125], Gaussian processes [126], and deep Gaussian processes [127].

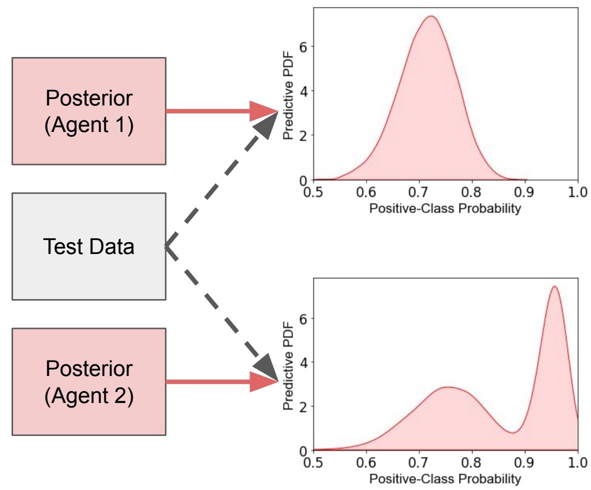


Figure 2: Bayesian machine learning in the context of binary classification with two agents. Each agent obtains a posterior pdf from training data and uses it to derive its predictive pdf of the probability that test data belong to the positive class. These predictive pdfs are subsequently combined to obtain an aggregate predictive pdf.

In certain scenarios of probabilistic machine learning, probabilistic opinion pooling can be used to resolve practical challenges. For example, the choice of a model (or an architecture, or a set of parameters) is frequently not obvious, and thus there is a model uncertainty that has to be taken into account to ensure robustness and generalization. A class of methods dealing with this issue is known as *ensemble learning*. The learning is carried out by a collection of algorithms based on different models, and the final result of classification, regression, or clustering is obtained by combining the individual results [128]–[133]. The combination of the results of individual probabilistic learning algorithms can be implemented via probabilistic opinion pooling, i.e., by fusing the predictive pdfs produced by the individual algorithms. An example in the context of binary classification is shown in Fig. 2. Probabilistic opinion pooling in ensemble learning has been successfully applied, e.g., in the context of deep ensembles [9], neural network ensembles [10], and ensemble Gaussian processes [11]. Note that in ensemble learning, unlike in multisensor signal processing and, in particular, target tracking as discussed in the previous subsection, all the algorithms may operate on the same set of data.

Another practical challenge in machine learning is posed by privacy-sensitive scenarios. Here, local (private) data observed at individual nodes may not be disseminated across the nodes or to a fusion center, and thus can be used only to train local models at the respective nodes. This framework, often referred to as *federated learning*, requires the combination of local models at a fusion center [12], [51], [52], [134]. Although in many instances of federated learning, updates are also communicated from the fusion center to the nodes, several works consider problem settings along the lines of probabilistic opinion pooling. For example, agnostic federated learning [135] combines sample representations of probability distributions trained on private data into an aggregate distribution.

Finally, the application of machine learning methods to "big data" scenarios calls for divide-and-conquer strategies that partition the data to much smaller sets, perform learning on each set, and combine the respective predictive or posterior distributions [65]–[67], [136]. Here, a focus has so far been on Markov chain Monte Carlo (MCMC) samplers for Bayesian inference [65]–[67]. For example, in [65], the idea is to generate a "subposterior" for each small dataset and combine the subposteriors using the multiplicative pooling function (see Section III-B7). Each subposterior is initially represented by a set of samples produced by an MCMC sampler but is then converted into a continuous pdf given by a kernel density estimate. The different pdfs are finally fused to form an approximation to the overall posterior pdf. This approach can be motivated by the fact, to be shown in Section VIII-A, that under a suitable conditional independence assumption a multiplicative pooling function operating on the subposteriors gives the overall posterior pdf.

The use of probabilistic machine learning has so far been restricted by the fact that many popular methods of machine learning do not provide probabilistic results. However, we expect that the outcomes of recent and ongoing research will remove this limitation and thereby increase the successful application of probabilistic opinion pooling in this field.

C. Forecasting

The goal of forecasting is to predict future values of some variable of interest based on present and past observed data [137]. An issue that may limit the performance of forecasting is a lack of confidence in the underlying model. This issue can be addressed by the *combination of forecasts*, which fuses the forecasting results obtained with several different models [53], [54], [56]. While classical work has considered point forecasts, *probabilistic forecasting* uses a description of the variable of interest in terms of probability distributions. Here, for a long time, the focus was on discrete probability distributions [18], and accordingly continuous random variables were approximated by discrete random variables through quantization. For example, in meteorology, the amount of precipitation was binned into a finite number of categories [138].

By contrast, the idea of *density forecasting* is to predict continuous random variables directly in terms of their pdfs [139]. This is visualized by Fig. 3, which shows a fan chart representation of density forecasts made by two experts. Density forecasting was suggested already more than 50 years ago [140], [141]. However, the combination of density forecasts [142]—which is a special setting of the fusion of pdfs—was considered only much later. Suggestions to combine density forecasts started with [20], [55], which discussed the optimization of the weights in the linear pooling function based on training data. At about the same time, the use of Bayesian

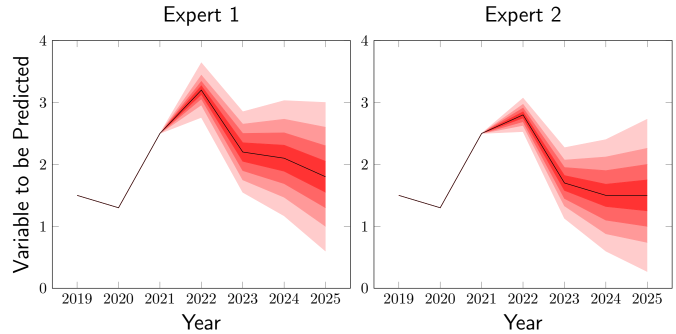


Figure 3: Density forecasts of a variable (e.g., inflation) beyond 2021 made by two experts, visualized as fan charts. The values of past years are already observed and thus fixed while predictions farther into the future become increasingly uncertain.

model averaging [57] in forecasting was proposed [143], again resulting in a linear pooling function. Also subsequent work focused on linear pooling [24]. Nonlinear pooling functions were mostly obtained by a preprocessing of the individual pdfs (e.g., in the spread-adjusted linear pool [144]) or by a postprocessing of the aggregate pdf (e.g., in the Beta-transformed linear pool [142]). Recently, the combination of density forecasts has also been studied in a nonparametric Bayesian setting based on the Beta-transformed linear pool [145].

While the combination of density forecasts has the same goal as pdf fusion—namely, to fuse pdfs from different sources—there are two distinctive features. First, realizations of the random variable to be predicted are observed on a regular basis, which enables an evaluation of density forecasts and their combinations based on new data. A significant part of the literature focuses on this aspect. Although beyond the scope of our work, such an evaluation can obviously be performed also within the general setting of pdf fusion if the required data are available. Second, forecasts usually concern one-dimensional random variables. This implies that the combination of forecasts can be formulated in terms of the one-dimensional cumulative distribution function (cdf), and more specific properties such as calibration [142] can be studied. Also the combination of forecasts—in particular, the choice of weights—is often based on new data and the evaluation of the fused one-dimensional cdf [146].

Probabilistic forecasting has been used in the broad domains of meteorology [18] and economics [20], [61], [139], [147] and, more specifically and more recently, in many disciplines including wind forecasting [148], [149], electric load forecasting [150], electricity price forecasting [151], and solar power forecasting [152]. The combination of density forecasts has, e.g., been considered in [20], [149], [152], and we conjecture that successful deployments of this variant of pdf fusion will emerge in many further applications of probabilistic forecasting.

III. Probabilistic Opinion Pooling

A. Basic Framework

In probabilistic opinion pooling, we are interested in fusing the pdfs of K agents or "experts" into a single pdf. Let $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{d_\theta}$ be a continuous random variable or vector defined on some probability space.[1] Furthermore, let the pdf $q_k(\boldsymbol{\theta}) \in \mathcal{P}$ represent the *opinion* of the kth agent. The sequence of all opinions $(q_1, q_2, \ldots, q_K) \in \mathcal{P}^K$ is called the *opinion profile*. We consider *events* to be (measurable) subsets of Θ. The probability of an event $\mathcal{A} \subseteq \Theta$ according to the opinion of the kth agent is given by

$$Q_k(\mathcal{A}) = \int_{\mathcal{A}} q_k(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}.$$

Given an opinion profile $(q_1, q_2, \ldots, q_K)$, a *pooling function* $g \colon \mathcal{P}^K \to \mathcal{P}$ is used to fuse the agents' pdfs $q_k(\boldsymbol{\theta})$ into a single pdf

$$q(\boldsymbol{\theta}) = g[q_1, \ldots, q_K](\boldsymbol{\theta}).$$

The probability of an event $\mathcal{A} \subseteq \Theta$ according to the fused pdf $q(\boldsymbol{\theta})$ is then given by

$$Q(\mathcal{A}) = \int_{\mathcal{A}} q(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}.$$

The fused pdf $q(\boldsymbol{\theta})$ summarizes the opinions of the K agents and will be referred to as the *aggregate pdf*. The fusion of the agent opinions via the pooling function is done (at least virtually) at a *fusion center*.

B. Pooling Functions

Over the years, many different pooling functions g have been proposed. We summarize some of them in the following. These pooling functions will be reconsidered in later sections.

1) Linear Pooling: The most popular pooling function is the *linear pooling function*, which was introduced in [26]. Linear pooling aggregates the agent opinions through a weighted arithmetic average, i.e.,

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta}), \tag{1}$$

where $(w_1, \ldots, w_K) \in \mathcal{S}_K$.

One can establish a connection between linear opinion pooling and model averaging [57]. Let us consider the joint distribution $q(\boldsymbol{\theta}, M)$ of the unknown random vector $\boldsymbol{\theta}$ and a discrete "model" random variable $M \in \{M_1, \ldots, M_K\}$. Furthermore, let $q(\boldsymbol{\theta}|M_k)$ denote the pdf of $\boldsymbol{\theta}$ conditioned on model M_k and $P(M_k)$ denote the probability of M_k. Then the marginal pdf of $\boldsymbol{\theta}$ is given by

$$q(\boldsymbol{\theta}) = \sum_{k=1}^{K} P(M_k) q(\boldsymbol{\theta}|M_k). \tag{2}$$

This is equivalent to the linear pooling operation (1), wherein the agent pdf $q_k(\boldsymbol{\theta})$ is interpreted as the pdf of $\boldsymbol{\theta}$ under model M_k, the weight w_k equals the probability of M_k, and the aggregate pdf $q(\boldsymbol{\theta})$ is the marginal pdf of $\boldsymbol{\theta}$.

[1]Our results extend to arbitrary probability measures that are absolutely continuous with respect to a σ-finite non-atomic measure. However, to keep the presentation more easily accessible, we present all results in the familiar setting of pdfs on $\mathbb{R}^{d_\theta}$.

2) Generalized Linear Pooling: The *generalized linear pooling function* defined in [153] includes an arbitrary pdf q_0 in the weighted arithmetic average (1), i.e.,

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = \sum_{k=0}^{K} w_k q_k(\boldsymbol{\theta}), \tag{3}$$

where $(w_0, \ldots, w_K) \in \mathcal{S}_{K+1}$. We note that in the general, measure-theoretic formulation of generalized linear opinion pooling in [153], some weights w_i are allowed to be negative. However, in the setting of fusing pdfs, this would result in a fusion rule g that does not give a valid (nonnegative) pdf for all possible opinion profiles $(q_1, \ldots, q_K)$. Thus, we restrict to nonnegative weights. One possible interpretation of the pdf q_0 is as the opinion of the fusion center. Alternatively, q_0 can be interpreted as a regularization.

3) Log-linear Pooling: Another popular pooling function is the *log-linear pooling function* [28]. This function aggregates the agent opinions using a weighted geometric average, i.e.,

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = c \prod_{k=1}^{K} \left(q_k(\boldsymbol{\theta})\right)^{w_k}, \tag{4}$$

where c is a normalization factor given by

$$c = \frac{1}{\int_\Theta \prod_{k=1}^{K} \left(q_k(\boldsymbol{\theta})\right)^{w_k} \mathrm{d}\boldsymbol{\theta}}, \tag{5}$$

and $(w_1, \ldots, w_K) \in \mathcal{S}_K$. To avoid the possibility of the integral in (5) being zero and, thus, c being undefined, this pooling function is usually only defined for pdfs that are positive on the domain Θ. We will refer to opinion profiles $(q_1, \ldots, q_K)$ that satisfy

$$q_k(\boldsymbol{\theta}) > 0 \quad \text{for all } \boldsymbol{\theta} \in \Theta \tag{6}$$

as *positive opinion profiles*.

The pooling function is called "log-linear" because it is a linear function of the agent pdfs in the log-domain, i.e., the logarithm of the right-hand side of (4) is

$$\log\left(c \prod_{k=1}^{K} \left(q_k(\boldsymbol{\theta})\right)^{w_k}\right) = \log(c) + \sum_{k=1}^{K} w_k \log(q_k(\boldsymbol{\theta})),$$

which is a weighted arithmetic average (up to the additive constant $\log(c)$). We will therefore refer to the powers $w_1, \ldots, w_K$ as "weights."

4) Generalized Log-linear Pooling: Similar to the generalized linear pooling function, a generalization of the log-linear pooling function can be obtained by including an arbitrary function ξ_0 as an additional factor. However, in contrast to the generalized linear pooling function, ξ_0 is not necessarily a pdf. More specifically, the *generalized log-linear pooling function* [154] is defined as

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = c\, \xi_0(\boldsymbol{\theta}) \prod_{k=1}^{K} \left(q_k(\boldsymbol{\theta})\right)^{w_k}, \tag{7}$$

where

$$c = \frac{1}{\int_\Theta \xi_0(\boldsymbol{\theta}) \prod_{k=1}^{K} \left(q_k(\boldsymbol{\theta})\right)^{w_k} \mathrm{d}\boldsymbol{\theta}},$$

ξ_0 is a bounded, positive function, and $(w_1, \ldots, w_K) \in \mathcal{S}_K$. Here, we again restrict to positive opinion profiles. The function ξ_0 can be used, e.g., to include the opinion of the fusion center or to regularize the fused density.

5) Hölder Pooling: The following pooling function was apparently first suggested in [78] as a generalization of the linear and log-linear pooling functions:

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = c\left(\sum_{k=1}^{K} w_k (q_k(\boldsymbol{\theta}))^{\alpha}\right)^{1/\alpha}, \tag{8}$$

where

$$c = \frac{1}{\int_{\Theta} \left(\sum_{k=1}^{K} w_k (q_k(\boldsymbol{\theta}))^{\alpha}\right)^{1/\alpha} \mathrm{d}\boldsymbol{\theta}}$$

and $\alpha \in \mathbb{R} \setminus \{0\}$. While for $\alpha \geq 1$ it can be shown that c is defined for arbitrary opinion profiles, in the other cases we have to restrict to opinion profiles such that c is defined. Because the pooling function in (8) is the weighted Hölder mean (also called the generalized average) [155] of the agent pdfs $q_k(\boldsymbol{\theta})$, we will refer to (8) as the *Hölder pooling function*. The linear and log-linear pooling functions are special cases of the Hölder pooling function for $\alpha = 1$ and $\alpha \to 0$, respectively.

6) Inverse-linear Pooling: The *inverse-linear pooling function* (weighted harmonic average) is defined as

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = c\left(\sum_{k=1}^{K} \frac{w_k}{q_k(\boldsymbol{\theta})}\right)^{-1}, \tag{9}$$

where

$$c = \frac{1}{\int_{\Theta} \left(\sum_{k=1}^{K} \frac{w_k}{q_k(\boldsymbol{\theta})}\right)^{-1} \mathrm{d}\boldsymbol{\theta}}.$$

This is the special case of the Hölder pooling function for $\alpha = -1$.

7) Multiplicative Pooling: The *multiplicative pooling function*, proposed in [80] for pmfs, is defined as

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = c\,(q_0(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} q_k(\boldsymbol{\theta}), \tag{10}$$

where

$$c = \frac{1}{\int_{\Theta} (q_0(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} q_k(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}},$$

and q_0 is a positive pdf called the *calibrating pdf*. Here, we restrict to positive opinion profiles and further assume that $q_k(\boldsymbol{\theta})/q_0(\boldsymbol{\theta})$ is bounded for all $k = 1, \ldots, K$. These assumptions guarantee that the normalization constant c is well-defined and nonzero. In Section VIII-A, we will show that within the supra-Bayesian framework, the multiplicative pooling function is the correct fusion rule for combining posterior pdfs in the case of conditionally independent observations. In that case, the calibrating pdf q_0 is the prior pdf used by the agents to form their posterior pdfs.

8) Generalized Multiplicative Pooling: We propose another pooling function that is a generalization of both the generalized log-linear pooling function and the multiplicative pooling function. In addition to a calibrating pdf q_0, we also allow for arbitrary weights in the generalized log-linear pooling function (7). More specifically, we define the *generalized multiplicative pooling function* as

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = c\,(q_0(\boldsymbol{\theta}))^{1-\sum_{k=1}^{K} w_k} \prod_{k=1}^{K} (q_k(\boldsymbol{\theta}))^{w_k}, \tag{11}$$

where

$$c = \frac{1}{\int_{\Theta} (q_0(\boldsymbol{\theta}))^{1-\sum_{k=1}^{K} w_k} \prod_{k=1}^{K} (q_k(\boldsymbol{\theta}))^{w_k} \mathrm{d}\boldsymbol{\theta}},$$

q_0 is a positive calibrating pdf, and the weights $w_1, \ldots, w_K \in \mathbb{R}$ are arbitrary real numbers. We again restrict to positive opinion profiles and assume that $(q_k(\boldsymbol{\theta})/q_0(\boldsymbol{\theta}))^{w_k}$ is bounded for all $k = 1, \ldots, K$. In Section IX-C, we will show that within the supra-Bayesian framework with a linear Gaussian model, the generalized multiplicative pooling function is the correct fusion rule for combining posterior pdfs.

9) Dictatorship Pooling: The *dictatorship pooling function* maps the opinion profile to a single agent opinion, i.e.,

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = q_k(\boldsymbol{\theta}), \tag{12}$$

for some fixed $k \in \{1, \ldots, K\}$. Although this function is a valid pooling function, one would not normally expect it to be a good choice.

10) Dogmatic Pooling: The *dogmatic pooling function* enforces a fixed pdf q_0 independently of the opinion profile, i.e.,

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = q_0(\boldsymbol{\theta}). \tag{13}$$

Again, this pooling function will not be suitable in most applications.

IV. The Axiomatic Approach

Fundamentally, we would like the pooling function $g[q_1, \ldots, q_K]$ to depend directly on all the agent pdfs q_k in a way that follows some rationale. One principled approach to probabilistic opinion pooling is the axiomatic approach, which seeks to determine all pooling functions that satisfy a set of desirable properties (axioms). In this section, we first formulate some axioms and then rigorously analyze the relationships between these axioms and the pooling functions presented in Section III-B.

A. Axioms

To begin, one basic restriction we may impose on the pooling function is that it be a symmetric function, i.e., a function whose arguments can be interchanged without altering the output of the function. This means that there is no "natural order" of the agents, and all agents are treated equally. This is formally stated in the following axiom:

Axiom 1. (Symmetry) *For all permutations* $\beta\colon \mathcal{K} \to \mathcal{K}$ *of the set* $\mathcal{K} = \{1, \ldots, K\}$ *and all opinion profiles* $(q_1, \ldots, q_K)$*, the pooling function* g *satisfies*

$$g[q_1, \ldots, q_K](\boldsymbol{\theta}) = g[q_{\beta(1)}, \ldots, q_{\beta(K)}](\boldsymbol{\theta}).$$

A symmetric pooling function seems to be desirable and natural since it treats the pdfs of the agents equally at the fusion center. However, if certain agents are known a priori to be more "reliable" or "informative" than other agents, then it may be reasonable to emphasize them in the pooling function. For example, in the linear or log-linear pooling function, we may assign larger weights w_k. If this is done in a fixed manner, the pooling function is no longer symmetric. On the other hand, if the weights are chosen adaptively such that each weight is an explicit function of the opinion profile and this adaptation rule involves each agent in the same way, then all agents are treated equally and the resulting pooling function is still symmetric. This will be further discussed in Section VII.

Another basic property for a pooling function is the preservation of agreement among agents. For instance, if each of the agents believes that a certain event $\mathcal{A} \subset \Theta$ is a null event, i.e., the probability of $\mathcal{A}$ is 0 according to all the agents, then $\mathcal{A}$ should also be a null event according to the aggregate pdf. This property is called the *zero preservation property* (ZPP) [156]:

Axiom 2. (Zero Preservation) *For any event* $\mathcal{A} \subset \Theta$*, if* $Q_k(\mathcal{A}) = 0$ *for all* k*, then* $Q(\mathcal{A}) = 0$*.*

The next property, termed *unanimity preservation* [80], asserts that if the opinions of the agents are identical, then the aggregate pdf should conform to that unanimous opinion.

Axiom 3. (Unanimity Preservation) *If for all events* $\mathcal{A} \subseteq \Theta$*, the probabilities* $Q_k(\mathcal{A}) = p_{\mathcal{A}}$ *coincide for all* k*, then* $Q(\mathcal{A}) = p_{\mathcal{A}}$*. Equivalently, if*[2] $q_k(\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$ *for all* k *and some pdf* $q_0(\boldsymbol{\theta})$*, then* $q(\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$*.*

Another property that may be desirable in a pooling function is the *strong setwise function property* (SSFP) [156]. The SSFP states that the probability of an event $\mathcal{A} \subseteq \Theta$ according to the aggregate pdf $q(\boldsymbol{\theta})$ can be expressed as a function of the probabilities of that event according to each agent, i.e., $Q_1(\mathcal{A}), \ldots, Q_K(\mathcal{A})$.

Axiom 4. (Strong Setwise Function Property) *There exists a function* $h\colon [0,1]^K \to [0,1]$ *such that for all opinion profiles* $(q_1, \ldots, q_K)$ *and for all events* $\mathcal{A} \subseteq \Theta$*,*

$$Q(\mathcal{A}) = h(Q_1(\mathcal{A}), \ldots, Q_K(\mathcal{A})). \tag{14}$$

We note that this axiom is in general not equivalent to the property that there exists a function $\tilde{h}\colon [0,\infty)^K \to [0,\infty)$ such that for all opinion profiles $(q_1, \ldots, q_K)$ and each point $\boldsymbol{\theta} \in \Theta$

$$q(\boldsymbol{\theta}) = \tilde{h}(q_1(\boldsymbol{\theta}), \ldots, q_K(\boldsymbol{\theta})). \tag{15}$$

In particular, for the case that Θ has finite Lebesgue measure $|\Theta|$, the dogmatic pooling function $q(\boldsymbol{\theta}) = 1/|\Theta|$ for $\boldsymbol{\theta} \in \Theta$ trivially satisfies (15) but not (14) (as a simple consequence of Theorem 1 below).

A more relaxed criterion than the SSFP is the *weak setwise function property* (WSFP) [156]. The WSFP states that the probability of an event according to the aggregate pdf is a function of the probabilities of that event according to each agent *and* the event itself.

Axiom 5. (Weak Setwise Function Property) *For all events* $\mathcal{A} \subseteq \Theta$*, there exists a generally* $\mathcal{A}$*-dependent function* $h_{\mathcal{A}}\colon [0,1]^K \to [0,1]$ *such that for all opinion profiles* $(q_1, \ldots, q_K)$

$$Q(\mathcal{A}) = h_{\mathcal{A}}(Q_1(\mathcal{A}), \ldots, Q_K(\mathcal{A})). \tag{16}$$

The WSFP is also equivalent to the so-called *marginalization property*, which states that marginalization and fusion are commutative operations. Formulating the marginalization property requires a measure-theoretic language that is beyond the scope of this paper. We thus omit a discussion of the marginalization property and refer the interested reader to [156] and [153].

Another relaxation of the SSFP is the *likelihood principle* [28]. Here, the value of the aggregate pdf $q(\boldsymbol{\theta})$ at some point $\boldsymbol{\theta}$ may only depend on the values of all $q_k(\boldsymbol{\theta})$ at the same $\boldsymbol{\theta}$ up to a normalization constant that can depend on the opinion profile.

Axiom 6. (Likelihood Principle) *There exists a function* $h\colon [0,\infty)^K \to [0,\infty)$ *such that for all opinion profiles* $(q_1, \ldots, q_K)$ *and each point* $\boldsymbol{\theta} \in \Theta$

$$q(\boldsymbol{\theta}) = \frac{h(q_1(\boldsymbol{\theta}), \ldots, q_K(\boldsymbol{\theta}))}{\int_\Theta h(q_1(\boldsymbol{\theta}'), \ldots, q_K(\boldsymbol{\theta}'))\mathrm{d}\boldsymbol{\theta}'}.$$

The name "likelihood principle" is motivated by viewing the pdfs as normalized likelihood functions: in this viewpoint, the idea is that the fused likelihood at $\boldsymbol{\theta}$ should only depend on the local likelihoods at $\boldsymbol{\theta}$ up to normalization [28]. Note that (15) is a significantly stronger assumption because the function $\tilde{h}$ in (15) has to normalize to one.

We can also formulate a weak version of the likelihood principle, where the function h may depend on $\boldsymbol{\theta}$ [28].

Axiom 7. (Weak Likelihood Principle) *For all* $\boldsymbol{\theta} \in \Theta$*, there exists a generally* $\boldsymbol{\theta}$*-dependent function* $h_{\boldsymbol{\theta}}\colon [0,\infty)^K \to [0,\infty)$ *such that for all opinion profiles* $(q_1, \ldots, q_K)$

$$q(\boldsymbol{\theta}) = \frac{h_{\boldsymbol{\theta}}(q_1(\boldsymbol{\theta}), \ldots, q_K(\boldsymbol{\theta}))}{\int_\Theta h_{\boldsymbol{\theta}}(q_1(\boldsymbol{\theta}'), \ldots, q_K(\boldsymbol{\theta}'))\mathrm{d}\boldsymbol{\theta}'}.$$

Another important axiom is *independence preservation*[3] [157]. This axiom asserts that if all the agents agree that two events $\mathcal{A}, \mathcal{B} \subseteq \Theta$ are independent, then these events should be independent also according to the aggregate pdf.

Axiom 8. (Independence Preservation) *For any events* $\mathcal{A}, \mathcal{B} \subseteq \Theta$*, if*

$$Q_k(\mathcal{A} \cap \mathcal{B}) = Q_k(\mathcal{A})Q_k(\mathcal{B})$$

[2]We consider two pdfs to be equal if they are equal almost everywhere with respect to the Lebesgue measure.

[3]Independence preservation should not be confused with the WSFP, which is sometimes referred to as the independence or eventwise independence property (e.g., [80]).

for all $k \in \{1,\ldots,K\}$*, then* $Q(\mathcal{A}\cap\mathcal{B}) = Q(\mathcal{A})Q(\mathcal{B})$.

A relaxation of independence preservation which, to the best of our knowledge, has not been considered before is to assume the preservation of a given factorization structure.

Axiom 9. (Factorization Preservation) *For any functions* $f_1\colon \Theta \to \mathbb{R}^{d_1}$ *and* $f_2\colon \Theta \to \mathbb{R}^{d_2}$*, if there exist functions* $q_{k,1}$ *and* $q_{k,2}$ *such that*

$$q_k(\boldsymbol{\theta}) = q_{k,1}(f_1(\boldsymbol{\theta}))q_{k,2}(f_2(\boldsymbol{\theta}))$$

for all $k \in \{1,\ldots,K\}$*, then there exist functions* $q_{\mathrm{a},1}$ *and* $q_{\mathrm{a},2}$ *such that*

$$q(\boldsymbol{\theta}) = q_{\mathrm{a},1}(f_1(\boldsymbol{\theta}))q_{\mathrm{a},2}(f_2(\boldsymbol{\theta}))\,.$$

This axiom expresses, in particular, preservation of the independence of components of $\boldsymbol{\theta}$. Assume that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and all agent pdfs factor according to $q_k(\boldsymbol{\theta}) = q_{k,1}(\boldsymbol{\theta}_1)q_{k,2}(\boldsymbol{\theta}_2)$. We can choose $f_1(\boldsymbol{\theta}) = \boldsymbol{\theta}_1$ and $f_2(\boldsymbol{\theta}) = \boldsymbol{\theta}_2$, and factorization preservation then implies that also the aggregate pdf preserves the independence of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, i.e., $q(\boldsymbol{\theta}) = q_{\mathrm{a},1}(\boldsymbol{\theta}_1)q_{\mathrm{a},2}(\boldsymbol{\theta}_2)$.

The final axioms we consider are motivated by Bayesian updating of probabilities. More specifically, we interpret each agent pdf $q_k(\boldsymbol{\theta})$ as the agent's belief about an unknown quantity $\boldsymbol{\theta}$ after observing some data. When observing new (additional) data, $q_k(\boldsymbol{\theta})$ is updated by multiplying it by a likelihood function $\ell\colon \Theta \to [0,\infty)$, which relates the agent's new data to $\boldsymbol{\theta}$. The updated belief of the kth agent, $q_k^{(\ell)}(\boldsymbol{\theta})$, is thus given as

$$q_k^{(\ell)}(\boldsymbol{\theta}) = \frac{\ell(\boldsymbol{\theta})q_k(\boldsymbol{\theta})}{\int_\Theta \ell(\boldsymbol{\theta}')q_k(\boldsymbol{\theta}')\mathrm{d}\boldsymbol{\theta}'}. \tag{17}$$

To avoid degenerate cases, one usually assumes in the following axioms that all pdfs are positive on the domain Θ. Thus, we restrict the statements of the axioms to positive opinion profiles. The first axiom related to the Bayesian framework is known as *external Bayesianity* [28], [158].

Axiom 10. (External Bayesianity) *For all functions* $\ell\colon \Theta \to [0,\infty)$ *and all positive opinion profiles* $(q_1,\ldots,q_K)$ *satisfying* $0 < \int_\Theta \ell(\boldsymbol{\theta})q_k(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} < \infty$ *for all* $k \in \{1,\ldots,K\}$*, we have*

$$q^{(\ell)}(\boldsymbol{\theta}) = g[q_1^{(\ell)},\ldots,q_K^{(\ell)}](\boldsymbol{\theta}),$$

where $q_k^{(\ell)}$ *is defined in* (17) *and*

$$q^{(\ell)}(\boldsymbol{\theta}) = \frac{\ell(\boldsymbol{\theta})q(\boldsymbol{\theta})}{\int_\Theta \ell(\boldsymbol{\theta}')q(\boldsymbol{\theta}')\mathrm{d}\boldsymbol{\theta}'}, \tag{18}$$

with $q(\boldsymbol{\theta}) = g[q_1,\ldots,q_K](\boldsymbol{\theta})$.

This axiom is motivated by the following Bayesian scenario: Assume that $q_1,\ldots,q_K$ are prior pdfs of K agents. Some data are observed, and the resulting likelihood function ℓ is provided to all agents. Then, a pooling function g satisfying external Bayesianity gives the same fusion result if it first aggregates the priors q_k into a fused prior q and then q is updated according to (18), or if it aggregates the posterior pdfs $q_k^{(\ell)}$ resulting from all agents updating their priors according to (17). Thus, external Bayesianity states that pdf updating and fusion are commutative operations. Such a property is desirable in applications where the agents share identical data (i.e., a global likelihood function) but have distinct prior distributions [159].

A second axiom related to the Bayesian framework is known as *individualized Bayesianity* [80]. This axiom is motivated by the idea of combining posterior probabilities, where each agent's posterior probability is based on private data (i.e., a local likelihood function) in contrast to all agents sharing identical data.

Axiom 11. (Individualized Bayesianity) *For all* $k \in \{1,\ldots,K\}$*, all bounded, positive*[4] *functions* $\ell\colon \Theta \to [0,\infty)$*, and all positive opinion profiles* $(q_1,\ldots,q_K)$*, we have*

$$q^{(\ell)}(\boldsymbol{\theta}) = g[q_1,\ldots,q_{k-1},q_k^{(\ell)},q_{k+1},\ldots,q_K](\boldsymbol{\theta}), \tag{19}$$

where $q_k^{(\ell)}$ *and* $q^{(\ell)}$ *are defined by* (17) *and* (18)*, respectively.*

This axiom is motivated by a scenario that is partly different from the scenario motivating external Bayesianity. We again assume that $q_1,\ldots,q_K$ are prior pdfs of the agents. For some arbitrary but fixed k, the kth agent observes (private) data in terms of a likelihood function ℓ. Then, a pooling function g satisfying individualized Bayesianity gives the same fusion result if it first aggregates the priors q_k into a fused prior q and then q is updated according to (18), or if it aggregates the priors of all but the kth agent and the posterior pdf $q_k^{(\ell)}$ resulting from the kth agent updating its prior according to (17). Thus, individualized Bayesianity states that pdf updating at a *single* agent and fusion are commutative operations.

Finally, we state a novel axiom that generalizes individualized Bayesianity. We thus call it *generalized Bayesianity*.

Axiom 12. (Generalized Bayesianity) *For all bounded, positive functions* $\ell_k\colon \Theta \to [0,\infty)$*,* $k \in \{1,\ldots,K\}$*, there exists a fused likelihood function* $h[\ell_1,\ldots,\ell_K]$ *such that for all positive opinion profiles* $(q_1,\ldots,q_K)$*, we have*

$$q^{(h[\ell_1,\ldots,\ell_K])}(\boldsymbol{\theta}) = g[q_1^{(\ell_1)},\ldots,q_K^{(\ell_K)}](\boldsymbol{\theta}), \tag{20}$$

where $q_k^{(\ell_k)}$ *and* $q^{(h[\ell_1,\ldots,\ell_K])}$ *are defined by* (17) *and* (18)*, respectively.*

This axiom states that fusing $q_1^{(\ell_1)},\ldots,q_K^{(\ell_K)}$, i.e., the result of updating $q_1,\ldots,q_K$, is equivalent to updating q, i.e., the result of fusing $q_1,\ldots,q_K$, by a "fused likelihood function" $h[\ell_1,\ldots,\ell_K]$. Note that the fused likelihood function is not allowed to depend on the opinion profile $(q_1,\ldots,q_K)$.

The axioms related to the Bayesian framework presented above are not directly related to the supra-Bayesian approach presented in Sections VIII and IX below. More specifically, in the supra-Bayesian framework, we have explicit likelihood functions and thus the pooling function does not necessarily satisfy properties that relate to arbitrary likelihood functions as in the axioms above.

B. Relations between Axioms and Pooling Functions

Having presented various pooling functions in Section III-B and various axioms in Section IV-A, we next analyze which

[4]The assumption of boundedness and positivity is needed to obtain the characterization theorems involving individualized Bayesianity in Section IV-B.

pooling functions satisfy which axioms and, conversely, which axioms imply which pooling functions. Our results are summarized in Table II. In what follows, we will abbreviate the various axioms as A1, A2, etc.

Theorem 1. *The linear pooling function in* (1) *satisfies the ZPP (A2), unanimity preservation (A3), the SSFP (A4), the WSFP (A5), the likelihood principle (A6), and the weak likelihood principle (A7). In addition, it satisfies the symmetry axiom (A1) if and only if all weights are equal, i.e.,* $w_1 = w_2 = \cdots = w_K = 1/K$. *Furthermore, for a pooling function* g *the following statements are equivalent:*

(i) g *is a linear pooling function;*
(ii) g *satisfies the SSFP (A4);*
(iii) g *satisfies the WSFP (A5) and the ZPP (A2);*
(iv) g *satisfies the WSFP (A5) and unanimity preservation (A3).*

The equivalence of (i), (ii), and (iii) was first proven in [156] for pmfs and in [153] for arbitrary probability measures. However, to the best of our knowledge, a proof for pdfs has not been provided so far.[5] In [80], the equivalence of (iv) and (iii) was presented for pmfs. In Appendix A, we give a proof of Theorem 1 for pdfs.

Theorem 2. *The generalized linear pooling function in* (3) *satisfies the WSFP (A5) and the weak likelihood principle (A7). Conversely, any pooling function that satisfies the WSFP (A5) is a generalized linear pooling function. In addition, the generalized linear pooling function satisfies the symmetry axiom (A1) if and only if all weights except* w_0 *are equal, i.e.,* $w_1 = w_2 = \cdots = w_K$.

The measure-theoretic equivalence of generalized linear pooling functions with possibly negative weights and pooling functions satisfying the WSFP (A5) was proven in [153]. However, in the case of the fusion of pdfs considered here, the generalized linear pooling functions cannot have negative weights. We thus present a proof with the necessary adaptations in Appendix B.

We next turn to pooling functions that include multiplication of pdfs or of powers of pdfs. In this context, we restrict to positive opinion profiles, i.e., we assume that (6) is satisfied. Note that in this setting the ZPP (A2) is not applicable since $Q_k(\mathcal{A}) = 0$ is not possible except for sets $\mathcal{A}$ of Lebesgue measure zero; therefore, we will disregard the ZPP in the following considerations.

Theorem 3. *The log-linear pooling function in* (4) *satisfies unanimity preservation (A3), the likelihood principle (A6), the weak likelihood principle (A7), factorization preservation (A9), external Bayesianity (A10), and generalized Bayesianity (A12). In addition, it satisfies the symmetry axiom (A1) if and only if all weights are equal, i.e.,* $w_1 = w_2 = \cdots = w_K = 1/K$. *Furthermore, for a pooling function* g *the following statements are equivalent:*

(i) g *is a log-linear pooling function;*
(ii) g *satisfies the likelihood principle (A6) and external Bayesianity (A10);*
(iii) g *satisfies unanimity preservation (A3), the weak likelihood principle (A7), and external Bayesianity (A10).*

The equivalence of (i) and (ii) was proven in [28] and the equivalence of (i) and (iii) in [154]. The remaining claimed axioms follow straightforwardly from the definition of the log-linear pooling function in (4).

Theorem 4. *The generalized log-linear pooling function in* (7) *satisfies the weak likelihood principle (A7), factorization preservation (A9), external Bayesianity (A10), and generalized Bayesianity (A12). In addition, it satisfies the symmetry axiom (A1) if and only if all weights except* w_0 *are equal, i.e.,* $w_1 = w_2 = \cdots = w_K$. *Furthermore, for a pooling function* g *the following statements are equivalent:*

(i) g *is a generalized log-linear pooling function;*
(ii) g *satisfies the weak likelihood principle (A7) and external Bayesianity (A10).*

This characterization theorem was proven in [154]. We note that the assumption of fusing pdfs (rather than general measures) is essential. In particular, for pmfs axioms A7 and A10 would imply only a "modified" generalized log-linear pooling function that may contain negative weights [154]. In [154], one can also find a characterization of all pooling functions that satisfy external Bayesianity (A10). However, these pooling functions do not have a simple structure.

Theorem 5. *The Hölder pooling function in* (8) *satisfies unanimity preservation (A3), the likelihood principle (A6), and the weak likelihood principle (A7). In addition, it satisfies the symmetry axiom (A1) if and only if all weights are equal, i.e.,* $w_1 = w_2 = \cdots = w_K$.

The proof of this theorem is straightforward and thus omitted. Because the inverse-linear pooling function (9) is a special case of the Hölder pooling function, it follows that it also satisfies A3, A6, and A7.

Theorem 6. *The multiplicative pooling function in* (10) *satisfies the symmetry axiom (A1), the weak likelihood principle (A7), factorization preservation (A9), individualized Bayesianity (A11), and generalized Bayesianity (A12). Furthermore, for a pooling function* g *the following statements are equivalent:*

(i) g *is a multiplicative pooling function with calibrating pdf* q_0;
(ii) g *satisfies individualized Bayesianity (A11) and there exists a pdf* $q_0(\boldsymbol{\theta})$ *such that* $g[q_0, \ldots, q_0](\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$.

The claimed axioms follow straightforwardly from the definition of the pooling function. A result similar to the equivalence of (i) and (ii) was proven for pmfs in [80]. We provide a proof for pdfs in Appendix C.

[5] Note that the proof for arbitrary probability measures in [153] does not imply the result for pdfs. Indeed, in our pdf framework, only probability measures that are absolutely continuous with respect to a fixed reference measure (usually the Lebesgue measure) are considered. This implicates the following difference from the framework of [153]: whereas we only assume that an axiom holds for all pdfs, [153] assumes that it also holds for other probability measures such as, e.g., a Dirac measure. Therefore, if [153] states that, e.g., the assumption (ii) implies (i), then this refers to a stronger version of (ii).

	Axiom											
Pooling Function	1	2	3	4	5	6	7	8	9	10	11	12
Linear	∗	✓	✓	✓	✓	✓	✓					
Generalized Linear	∗				✓		✓					
Log-linear	∗	n.a.	✓			✓	✓		✓	✓		✓
Generalized Log-linear	∗	n.a.					✓		✓	✓		✓
Hölder	∗	n.a.	✓			✓	✓					
Inverse-linear	∗	n.a.	✓			✓	✓					
Multiplicative	✓	n.a.					✓		✓		✓	✓
Generalized Multiplicative	∗	n.a.					✓		✓			✓
Dictatorship		✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Dogmatic	✓				✓		✓					✓

Table II: Axioms satisfied by the pooling functions presented in Section III-B. (∗: satisfied if and only if all weights are equal.)

Theorem 7. *The generalized multiplicative pooling function in* (11) *satisfies the weak likelihood principle (A7), factorization preservation (A9), and generalized Bayesianity (A12). In addition, it satisfies the symmetry axiom (A1) if and only if all weights are equal, i.e.,* $w_1 = w_2 = \cdots = w_K$.

Again, the claimed axioms follow straightforwardly from the definition of the pooling function.

Theorem 8. *The dictatorship pooling function in* (12) *satisfies the ZPP (A2), unanimity preservation (A3), the SSFP (A4), the WSFP (A5), the likelihood principle (A6), the weak likelihood principle (A7), independence preservation (A8), factorization preservation (A9), external Bayesianity (A10), and generalized Bayesianity (A12). Furthermore, for a pooling function* g *the following statements are equivalent:*

(i) g *is a dictatorship pooling function;*

(ii) g *satisfies the SSFP (A4) and independence preservation (A8);*

(iii) g *satisfies the WSFP (A5) and independence preservation (A8);*

(iv) g *satisfies the SSFP (A4) and external Bayesianity (A10);*

(v) g *satisfies the WSFP (A5) and external Bayesianity (A10);*

(vi) g *satisfies the SSFP (A4) and generalized Bayesianity (A12).*

Our statements regarding the satisfied axioms follow easily from the definition of the dictatorship pooling function. The equivalence of (i) and (ii) was proven in [153, Theorem 3.1]. In Appendix D, we strengthen this result and show that the WSFP—instead of the (stronger) SSFP—in combination with independence preservation suffices to axiomatically define the dictatorship pooling function, i.e., that (iii) implies (ii). The equivalence of (i) and (iv) was proven in [160]. In fact, [160] even states the equivalence of (i) and (v) by proving that the version of external Bayesianity considered in [160] implies the ZPP. However, our formulation of external Bayesianity assumes positive opinion profiles and thus the ZPP cannot be proven. To close this gap, we further show in Appendix D that (v) implies (iv). Finally, we also show in Appendix D that (vi) implies (i).

We note that the dictatorship pooling function is a special case of both the linear and log-linear pooling functions, when one of the weights is 1 and all the others are 0. The fact that the dictatorship pooling function satisfies ten axioms shows that a pooling function that satisfies many axioms is not necessarily a useful pooling function.

Turning to the dogmatic pooling function, we first present a preliminary result that is proven in Appendix E.

Lemma 9. *Assume that a pooling function* g *satisfies the WSFP (A5) and generalized Bayesianity (A12). Then* g *is either a dogmatic pooling function or a dictatorship pooling function.*

The following characterization of the dogmatic pooling function now follows easily.

Theorem 10. *The dogmatic pooling function in* (13) *satisfies the symmetry axiom (A1), the WSFP (A5), the weak likelihood principle (A7), and generalized Bayesianity (A12). Conversely, any pooling function that satisfies the symmetry axiom (A1), the WSFP (A5), and generalized Bayesianity (A12) is a dogmatic pooling function.*

It is obvious that the dogmatic pooling function satisfies the stated axioms. The converse follows because by Lemma 9 the pooling function must be either a dogmatic pooling function or a dictatorship pooling function, but of these only the dogmatic pooling function is symmetric.

Based on the theorems above, we can establish an implication structure for the different axioms from Section IV-A, which indicates which axioms imply which other axioms. To formalize this structure, we will designate the set of all pooling functions that satisfy Axiom i as $\mathcal{F}_i$. The next theorem states the currently known implications. Venn diagrams representing the implication structure are presented in Fig. 4.

Theorem 11. *For the axioms introduced in Section IV-A, the following implications hold:*

(i) *The SSFP (A4) implies the ZPP (A2), unanimity preservation (A3), the WSFP (A5), the likelihood principle (A6), and the weak likelihood principle (A7), i.e.,* $\mathcal{F}_4 \subseteq \mathcal{F}_2 \cap \mathcal{F}_3 \cap \mathcal{F}_5 \cap \mathcal{F}_6 \cap \mathcal{F}_7$. *Furthermore,* $\mathcal{F}_4 = \mathcal{F}_2 \cap \mathcal{F}_5 = \mathcal{F}_3 \cap \mathcal{F}_5$.

(ii) *The WSFP (A5) implies the weak likelihood principle (A7), i.e.,* $\mathcal{F}_5 \subseteq \mathcal{F}_7$.

(iii) *The likelihood principle (A6) implies the weak likelihood principle (A7), i.e.,* $\mathcal{F}_6 \subseteq \mathcal{F}_7$.

(iv) *Independence preservation (A8) implies the ZPP (A2), i.e.,* $\mathcal{F}_8 \subseteq \mathcal{F}_2$.

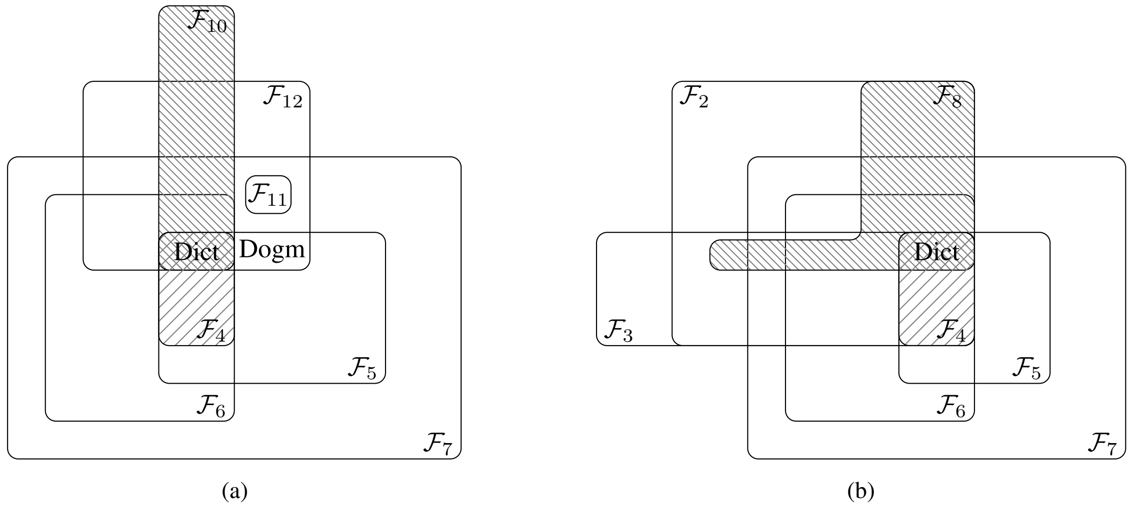


Figure 4: Venn diagrams representing the implication structure for the axioms from Section IV-A: (a) A4–A7 and A10–A12 as well as intersections resulting in dictatorship (Dict) or dogmatic (Dogm) pooling functions; (b) A2–A8 as well as intersections resulting in dictatorship (Dict) pooling functions. Note that the diagrams illustrate the currently known implications, and some regions that appear non-empty in the diagrams may actually be empty sets. For better visibility, the sets $\mathcal{F}_4$, $\mathcal{F}_8$, and $\mathcal{F}_{10}$ are highlighted by different line-patterns.

(v) *Individualized Bayesianity (A11) implies generalized Bayesianity (A12), i.e.,* $\mathcal{F}_{11} \subseteq \mathcal{F}_{12}$.

Most of these implications follow from our earlier theorems. For completeness, we provide a proof of, or references for, all implications in Appendix F.

V. The Optimization Approach

In the previous section, we identified pooling functions that satisfy certain axioms. An alternative approach to establishing pooling functions for probabilistic opinion pooling is the optimization approach. Here, a pooling function is obtained by minimizing the weighted average of some discrepancy measure between the pdfs of the K agents, $q_1(\boldsymbol{\theta}), \ldots, q_K(\boldsymbol{\theta})$, and the aggregate pdf $q(\boldsymbol{\theta})$. The underlying idea is to make the aggregate pdf as similar as possible to all the agent pdfs simultaneously. As we will see, the obtained $q(\boldsymbol{\theta})$ turns out to be some sort of average of the agent pdfs $q_1(\boldsymbol{\theta}), \ldots, q_K(\boldsymbol{\theta})$.

One class of discrepancy measures that can be considered are f-divergences. For a convex function $f\colon \mathbb{R}^+ \to \mathbb{R}$ with $f(1) = 0$, the f-divergence between two pdfs $q_k(\boldsymbol{\theta})$ and $\varphi(\boldsymbol{\theta})$ with common domain Θ is defined as [161]–[164]

$$\mathcal{D}_f(q_k \| \varphi) = \int_\Theta \varphi(\boldsymbol{\theta}) f\left(\frac{q_k(\boldsymbol{\theta})}{\varphi(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta}. \tag{21}$$

The fusion of the agent pdfs $q_1(\boldsymbol{\theta}), \ldots, q_K(\boldsymbol{\theta})$ can then be based on defining the aggregate pdf $q(\boldsymbol{\theta})$ as the pdf that minimizes a weighted average of f-divergences:[6]

$$q = \operatorname*{arg\,min}_{\varphi \in \mathcal{P}} \sum_{k=1}^K w_k \mathcal{D}_f(q_k \| \varphi), \tag{22}$$

where the weights satisfy $(w_1, \ldots, w_K) \in \mathcal{S}_K$. In what follows, we consider some specific f-divergences and derive the associated pooling functions defined by (22). These results are summarized in Table III.

A. Kullback-Leibler Divergence

For $f(x) = x \log x$, the f-divergence is the Kullback-Leibler divergence (KLD) [167]

$$\mathcal{D}_{\mathrm{KL}}(q_k \| \varphi) = \int_\Theta q_k(\boldsymbol{\theta}) \log\left(\frac{q_k(\boldsymbol{\theta})}{\varphi(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta}. \tag{23}$$

Under this choice of divergence, the pooling function that solves the optimization problem in (22) is the linear pooling function in (1):

Theorem 12. *Let* $f(x) = x \log x$ *(i.e.,* $\mathcal{D}_f(q_k \| \varphi) = \mathcal{D}_{\mathrm{KL}}(q_k \| \varphi)$*) and* $(w_1, \ldots, w_K) \in \mathcal{S}_K$*. Then, the solution to the optimization problem in* (22) *is*

$$q(\boldsymbol{\theta}) = \sum_{k=1}^K w_k q_k(\boldsymbol{\theta}).$$

[6]This minimization establishes a conceptual link to a central problem in the field of robust hypothesis testing, namely, the identification of a vector of "least favorable" pdfs within a given set of hypothesized pdfs. For two pdfs, this problem can be shown to be equivalent to the joint minimization of all f-divergences (21) for all twice differentiable convex functions f [165], [166]. The solution to this minimization can be interpreted as the pdfs that are maximally similar within the set of hypothesized pdfs, which means that a statistical test between the respective pdfs is "as hard as possible." It is interesting that an interpretation as a maximally similar pdf holds for both the optimization approach to pdf fusion and robust hypothesis testing.

Pooling Function	$f(x)$	$\mathcal{D}_f(q_k\|\varphi)$	$\chi(x)$	$\|\chi(q_k)-\chi(\varphi)\|_2^2$
Linear: $q(\boldsymbol{\theta})=\sum_{k=1}^K w_k q_k(\boldsymbol{\theta})$	$x\log x$	$\int_\Theta q_k(\boldsymbol{\theta})\log\left(\frac{q_k(\boldsymbol{\theta})}{\varphi(\boldsymbol{\theta})}\right)\mathrm{d}\boldsymbol{\theta}$	x	$\int_\Theta (q_k(\boldsymbol{\theta})-\varphi(\boldsymbol{\theta}))^2\,\mathrm{d}\boldsymbol{\theta}$
Log-linear: $q(\boldsymbol{\theta})\propto\prod_{k=1}^K (q_k(\boldsymbol{\theta}))^{w_k}$	$-\log x$	$\int_\Theta \varphi(\boldsymbol{\theta})\log\left(\frac{\varphi(\boldsymbol{\theta})}{q_k(\boldsymbol{\theta})}\right)\mathrm{d}\boldsymbol{\theta}$	$\log x$	$\int_\Theta (\log q_k(\boldsymbol{\theta})-\log\varphi(\boldsymbol{\theta}))^2\,\mathrm{d}\boldsymbol{\theta}$
Inverse-linear: $q(\boldsymbol{\theta})\propto\left(\sum_{k=1}^K \frac{w_k}{q_k(\boldsymbol{\theta})}\right)^{-1}$	$\frac{1-x}{2x}$	$\int_\Theta \frac{(q_k(\boldsymbol{\theta})-\varphi(\boldsymbol{\theta}))^2}{q_k(\boldsymbol{\theta})}\mathrm{d}\boldsymbol{\theta}$	$\frac{1}{x}$	$\int_\Theta \left(\frac{1}{q_k(\boldsymbol{\theta})}-\frac{1}{\varphi(\boldsymbol{\theta})}\right)^2\mathrm{d}\boldsymbol{\theta}$
Hölder: $q(\boldsymbol{\theta})\propto\left(\sum_{k=1}^K w_k (q_k(\boldsymbol{\theta}))^\alpha\right)^{1/\alpha}$	$\frac{x^\alpha-1}{\alpha(\alpha-1)}$	$\frac{1}{\alpha(\alpha-1)}\int_\Theta \varphi(\boldsymbol{\theta})\frac{(q_k(\boldsymbol{\theta}))^\alpha-(\varphi(\boldsymbol{\theta}))^\alpha}{(\varphi(\boldsymbol{\theta}))^\alpha}\mathrm{d}\boldsymbol{\theta}$	x^α	$\int_\Theta ((q_k(\boldsymbol{\theta}))^\alpha-(\varphi(\boldsymbol{\theta}))^\alpha)^2\,\mathrm{d}\boldsymbol{\theta}$

Table III: Optimization-based definition of pooling functions: some pooling functions along with the underlying f-divergence $\mathcal{D}_f(q_k\|\varphi)$ and squared distance function $d^2(q_k,\varphi)=\|\chi(q_k)-\chi(\varphi)\|_2^2$ used in the optimization problems in (22) and (32), respectively.

A proof of this theorem can be found in [83]. The proof is based on the fact that minimizing the weighted average of KLDs is equivalent to minimizing the cross-entropy

$$H(q_{\mathrm{mix}},\varphi)=-\int_\Theta q_{\mathrm{mix}}(\boldsymbol{\theta})\log(\varphi(\boldsymbol{\theta}))\,\mathrm{d}\boldsymbol{\theta}$$

between the mixture pdf $q_{\mathrm{mix}}(\boldsymbol{\theta})=\sum_{k=1}^K w_k q_k(\boldsymbol{\theta})$ and the pdf $\varphi\in\mathcal{P}$. That is,

$$\arg\min_{\varphi\in\mathcal{P}}\sum_{k=1}^K w_k\mathcal{D}_{\mathrm{KL}}(q_k\|\varphi)=\arg\min_{\varphi\in\mathcal{P}} H(q_{\mathrm{mix}},\varphi).$$

The cross-entropy $H(q_{\mathrm{mix}},\varphi)$ is minimized if and only if $q_{\mathrm{mix}}(\boldsymbol{\theta})$ and $\varphi(\boldsymbol{\theta})$ are equal. This follows from the fact that $H(q_{\mathrm{mix}},\varphi)$ is equal to the sum of the KLD between $q_{\mathrm{mix}}(\boldsymbol{\theta})$ and $\varphi(\boldsymbol{\theta})$ and the differential entropy of $q_{\mathrm{mix}}(\boldsymbol{\theta})$ [129, Chapter 2], i.e.,

$$H(q_{\mathrm{mix}},\varphi)=\mathcal{D}_{\mathrm{KL}}(q_{\mathrm{mix}}\|\varphi)-\int_\Theta q_{\mathrm{mix}}(\boldsymbol{\theta})\log(q_{\mathrm{mix}}(\boldsymbol{\theta}))\mathrm{d}\boldsymbol{\theta}.$$

Hence, $H(q_{\mathrm{mix}},\varphi)$ is minimized if and only if $\mathcal{D}_{\mathrm{KL}}(q_{\mathrm{mix}}\|\varphi)$ is minimized, which implies that $\varphi(\boldsymbol{\theta})=q_{\mathrm{mix}}(\boldsymbol{\theta})$.

B. Reverse Kullback-Leibler Divergence

Next, consider $f(x)=-\log x$. In this case, the f-divergence corresponds to the KLD whose arguments are reversed with respect to (23) [168], i.e.,

$$\mathcal{D}_{\mathrm{KL}}(\varphi\|q_k)=\int_\Theta\varphi(\boldsymbol{\theta})\log\left(\frac{\varphi(\boldsymbol{\theta})}{q_k(\boldsymbol{\theta})}\right)\mathrm{d}\boldsymbol{\theta}.$$

We refer to $\mathcal{D}_{\mathrm{KL}}(\varphi\|q_k)$ as the *reverse* KLD. For the reverse KLD, the solution to the optimization problem in (22) is the log-linear pooling function in (4):

Theorem 13. *Let $f(x)=-\log x$ (i.e., $\mathcal{D}_f(q_k\|\varphi)=\mathcal{D}_{\mathrm{KL}}(\varphi\|q_k)$) and $(w_1,\ldots,w_K)\in\mathcal{S}_K$. Then, the solution to the optimization problem in* (22) *is*

$$q(\boldsymbol{\theta})=c\prod_{k=1}^K(q_k(\boldsymbol{\theta}))^{w_k},$$

where $c=1/\int_\Theta\prod_{k=1}^K(q_k(\boldsymbol{\theta}))^{w_k}\,\mathrm{d}\boldsymbol{\theta}$.

A proof of this theorem can be found in [83] and [169]. The idea behind the proof is to derive a lower bound on the weighted average of reverse KLDs using Jensen's inequality and then to show that the lower bound is achieved if and only if (4) is satisfied.

C. α-Divergences

We have shown that both the linear and log-linear pooling functions can be derived using the optimization approach involving the KLD or reverse KLD, respectively. These two results can be extended to an entire family of divergences and a corresponding family of pooling functions that are both parameterized by a real parameter α. Indeed, let us consider the f-divergence $\mathcal{D}_f(q_k\|\varphi)$ induced by

$$f(x)=f_\alpha(x)\triangleq\frac{x^\alpha-1}{\alpha(\alpha-1)},$$

where $x>0$ and $\alpha\in\mathbb{R}\setminus\{0,1\}$. This yields the family of α-divergences defined as [170]–[172]

$$\mathcal{D}_\alpha(q_k\|\varphi)\triangleq\mathcal{D}_{f_\alpha}(q_k\|\varphi) \tag{24}$$

$$=\frac{1}{\alpha(\alpha-1)}\int_\Theta\varphi(\boldsymbol{\theta})\frac{(q_k(\boldsymbol{\theta}))^\alpha-(\varphi(\boldsymbol{\theta}))^\alpha}{(\varphi(\boldsymbol{\theta}))^\alpha}\mathrm{d}\boldsymbol{\theta}\,. \tag{25}$$

We remark that the α-divergence equals the so-called Hellinger divergence up to a scaling factor and is also a one-to-one transformation of the Rényi divergence [164]. Using the optimization approach for the α-divergences, we obtain the α-parameterized family of Hölder pooling functions in (8). As noted earlier, this family comprises the linear, log-linear, and inverse-linear pooling functions as special cases.

Theorem 14. *Let $f(x)=f_\alpha(x)=\frac{x^\alpha-1}{\alpha(\alpha-1)}$ (i.e., $\mathcal{D}_f(q_k\|\varphi)=\mathcal{D}_\alpha(q_k\|\varphi)$) with $\alpha\in\mathbb{R}\setminus\{0,1\}$ and $(w_1,\ldots,w_K)\in\mathcal{S}_K$. Then, the solution to the optimization problem in* (22) *is*

$$q(\boldsymbol{\theta})=c\left(\sum_{k=1}^K w_k(q_k(\boldsymbol{\theta}))^\alpha\right)^{1/\alpha}, \tag{26}$$

where $c=1/\int_\Theta\left(\sum_{k=1}^K w_k(q_k(\boldsymbol{\theta}))^\alpha\right)^{1/\alpha}\mathrm{d}\boldsymbol{\theta}$.

Although this result was mentioned in [84, Fig. 1], to the best of our knowledge, a proof does not exist in the literature. We provide a proof in Appendix G.

In the limiting case $\alpha\to 0$, the Hölder pooling function (26) becomes the log-linear pooling function (weighted geometric average) in (4), while for $\alpha=1$ it equals the linear pooling function (weighted arithmetic average) in (1). These results are consistent with the fact that $\lim_{\alpha\to 0}\mathcal{D}_\alpha(q_k\|\varphi)=\mathcal{D}_{\mathrm{KL}}(\varphi\|q_k)$ and $\lim_{\alpha\to 1}\mathcal{D}_\alpha(q_k\|\varphi)=\mathcal{D}_{\mathrm{KL}}(q_k\|\varphi)$ [171]. For $\alpha=-1$, the Hölder pooling function (26) becomes the inverse-linear pooling function (9). Furthermore, the α-divergence in the case

$\alpha = 2$ is (up to a scaling factor 2) equal to the Pearson χ^2-divergence [164], [173]

$$\chi^2(q_k, \varphi) \triangleq \int_\Theta \frac{(q_k(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta}))^2}{\varphi(\boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta} = \int_\Theta \varphi(\boldsymbol{\theta}) \frac{(q_k(\boldsymbol{\theta}))^2 - (\varphi(\boldsymbol{\theta}))^2}{(\varphi(\boldsymbol{\theta}))^2} \mathrm{d}\boldsymbol{\theta}\,.$$

The corresponding Hölder pooling function (26) is thus

$$q(\boldsymbol{\theta}) = c \left(\sum_{k=1}^{K} w_k (q_k(\boldsymbol{\theta}))^2 \right)^{1/2},$$

where $c = 1 / \int_\Theta \left(\sum_{k=1}^{K} w_k (q_k(\boldsymbol{\theta}))^2 \right)^{1/2} \mathrm{d}\boldsymbol{\theta}$.

D. Reverse α-Divergences

As for the KLD, one can exchange the order of q_k and φ in the α-divergence in (25). Again, this is equivalent to changing to a different f-divergence. More precisely, it is stated in [162, eq. (1.13)] (see also [164, Prop. 2]) that

$$\mathcal{D}_f(\varphi \| q_k) = \mathcal{D}_{f^*}(q_k \| \varphi), \tag{27}$$

where $f^*(x) = x f(1/x)$. Based on this result, we show in Appendix H that

$$\mathcal{D}_\alpha(\varphi \| q_k) = \mathcal{D}_{\alpha^*}(q_k \| \varphi),$$

where $\alpha^* = 1 - \alpha$. Thus, Theorem 14 implies the following result.

Corollary 15. *The solution to the optimization problem*

$$q = \arg\min_{\varphi \in \mathcal{P}} \sum_{k=1}^{K} w_k \mathcal{D}_\alpha(\varphi \| q_k) \tag{28}$$

is

$$q(\boldsymbol{\theta}) = c \left(\sum_{k=1}^{K} w_k (q_k(\boldsymbol{\theta}))^{\alpha^*} \right)^{1/\alpha^*}, \tag{29}$$

where $c = 1 / \int_\Theta \left(\sum_{k=1}^{K} w_k (q_k(\boldsymbol{\theta}))^{\alpha^*} \right)^{1/\alpha^*} \mathrm{d}\boldsymbol{\theta}$ *and* $\alpha^* = 1 - \alpha$.

In particular, the reverse α-divergence for $\alpha = 2$ corresponds to the Pearson χ^2-divergence in the reverse direction, i.e., $\chi^2(\varphi, q_k) = \int_\Theta \frac{(q_k(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta}))^2}{q_k(\boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta}$. In this case, $\alpha^* = 1 - 2 = -1$ and the corresponding Hölder pooling function (29) is the inverse-linear pooling function (9).

E. Symmetric Discrepancy Measures

As previously mentioned, the optimization approach defines pooling functions by minimizing a weighted average of discrepancy measures between the agent pdfs and the aggregate pdf. So far, our focus has been on minimizing a weighted average of f-divergences, where our choices of f yielded *asymmetric* discrepancy measures. Through this approach, we derived pooling functions that are the weighted arithmetic, geometric, harmonic, and Hölder averages of the agent pdfs. Interestingly, these fusion rules can also be derived using an alternative formulation, where the goal is to minimize a weighted average of *symmetric* discrepancy measures (distance functions). Let $d(q_k, \varphi)$ be a symmetric function expressing a distance between the kth agent pdf $q_k(\boldsymbol{\theta})$ and the pdf $\varphi(\boldsymbol{\theta})$, where symmetric means that $d(q_k, \varphi) = d(\varphi, q_k)$. Then, we can define the aggregate pdf to be the solution to the following optimization problem:

$$q(\boldsymbol{\theta}) = \arg\min_{\varphi \in \mathcal{P}} \sum_{k=1}^{K} w_k d^2(q_k, \varphi), \tag{30}$$

where $(w_1, \ldots, w_K) \in \mathcal{S}_K$. The resulting $q(\boldsymbol{\theta})$ has been referred to as *Fréchet mean* [44].

An important distance function is the L_2 distance function defined as

$$\|q_k - \varphi\|_2 = \sqrt{\int_\Theta (q_k(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta}))^2 \, \mathrm{d}\boldsymbol{\theta}}. \tag{31}$$

The linear pooling function can be obtained alternatively by minimizing a weighted average of squared L_2 distances:

Theorem 16. *Let* $d(q_k, \varphi) = \|q_k - \varphi\|_2$ *and* $(w_1, \ldots, w_K) \in \mathcal{S}_K$. *Then, the solution to the optimization problem in* (30) *is*

$$q(\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta}).$$

This result was mentioned without proof in [84, Fig. 1]. We provide a proof in Appendix I.

Unfortunately, for arbitrary distance functions $d(q_k, \varphi)$, an analytical solution to the optimization problem in (30) does not exist. This is due to the difficulty in satisfying the constraint $\varphi \in \mathcal{P}$, which ensures that the obtained aggregate pdf $q(\boldsymbol{\theta})$ is a valid pdf. To overcome this difficulty, following [44], we can instead solve the unconstrained version of the optimization problem in (30), i.e.,

$$\tilde{q}(\boldsymbol{\theta}) = \arg\min_{\varphi} \sum_{k=1}^{K} w_k d^2(q_k, \varphi), \tag{32}$$

and then normalize the result, i.e.,

$$q(\boldsymbol{\theta}) = \frac{\tilde{q}(\boldsymbol{\theta})}{\int_\Theta \tilde{q}(\boldsymbol{\theta}') d\boldsymbol{\theta}'}.$$

However, we emphasize that the obtained aggregate pdf $q(\boldsymbol{\theta})$ is generally different from the solution of the constrained optimization problem in (30).

Using this unconstrained approach, the minimization of the L_2 distance function (31) results again in the linear pooling function [44]. Here, the solution satisfies the constraint $q \in \mathcal{P}$ without explicitly enforcing it. Furthermore, the log-linear [44], inverse-linear, and Hölder pooling functions can be derived in an analogous manner using suitable distance functions. We can arrive at all of these results and many more in a unified manner by considering the general class of distance functions $d(q_k, \varphi)$ defined as

$$\|\chi(q_k) - \chi(\varphi)\|_2 = \sqrt{\int_\Theta (\chi(q_k(\boldsymbol{\theta})) - \chi(\varphi(\boldsymbol{\theta})))^2 \, \mathrm{d}\boldsymbol{\theta}}, \tag{33}$$

where $\chi\colon (0, \infty) \to (a, b)$ with $a \in \mathbb{R} \cup \{-\infty\}$ and $b \in \mathbb{R} \cup \{\infty\}$ is an invertible function. Solving the optimization problem (32) for the distance functions (33) leads to the rich class of pooling functions stated by the following result.

Theorem 17. *Let* $d(q_k, \varphi) = \|\chi(q_k) - \chi(\varphi)\|_2$ *and* $(w_1, \ldots, w_K) \in \mathcal{S}_K$. *Then, the solution to the optimization problem in* (32) *is*

$$\tilde{q}(\boldsymbol{\theta}) = \chi^{-1}\left(\sum_{k=1}^{K} w_k \chi(q_k(\boldsymbol{\theta}))\right). \tag{34}$$

A proof is provided in Appendix J, and the functions χ leading to the linear, log-linear, inverse-linear, and Hölder pooling functions are listed in Table III. Note that the solution $\tilde{q}(\boldsymbol{\theta})$ in (34) is always nonnegative because the domain of χ is $(0, \infty)$.

VI. Gaussian Densities

In Sections III and V, we discussed a variety of pooling functions that can be used to fuse the pdfs of several agents into a single aggregate pdf. We now consider the practically important special case where the opinions of the agents are represented by Gaussian pdfs. That is, we assume that

$$q_k(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{q_k}, \boldsymbol{\Sigma}_{q_k}), \quad k = 1, \ldots, K, \tag{35}$$

where $\mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{q_k}, \boldsymbol{\Sigma}_{q_k})$ denotes a multivariate Gaussian pdf with mean $\boldsymbol{\mu}_{q_k} = \mathbb{E}_{q_k}[\boldsymbol{\theta}]$ and covariance matrix $\boldsymbol{\Sigma}_{q_k} = \mathbb{E}_{q_k}[(\boldsymbol{\theta} - \boldsymbol{\mu}_{q_k})(\boldsymbol{\theta} - \boldsymbol{\mu}_{q_k})^{\intercal}]$. An important aspect of the Gaussian case is the fact that each agent pdf $q_k(\boldsymbol{\theta})$ is completely characterized by its first- and second-order moments $\boldsymbol{\mu}_{q_k}$ and $\boldsymbol{\Sigma}_{q_k}$.

A. Linear Pooling

The fusion of Gaussian pdfs using the linear pooling function in (1) results in an aggregate pdf that is a mixture of Gaussians, i.e.,

$$q(\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{q_k}, \boldsymbol{\Sigma}_{q_k}). \tag{36}$$

A convenient property in this context is that the expected value of a function $h(\boldsymbol{\theta})$ with respect to the pdf $q(\boldsymbol{\theta})$ in (1) is the weighted average of the expected values of $h(\boldsymbol{\theta})$ with respect to the agent pdfs $q_1(\boldsymbol{\theta}), \ldots, q_K(\boldsymbol{\theta})$, i.e., $\mathbb{E}_q[h(\boldsymbol{\theta})] = \sum_{k=1}^{K} w_k \mathbb{E}_{q_k}[h(\boldsymbol{\theta})]$. This implies that the mean of the aggregate pdf in (36), $\boldsymbol{\mu}_q = \mathbb{E}_q[\boldsymbol{\theta}]$, is simply the weighted average of the agent means, i.e.,

$$\boldsymbol{\mu}_q = \sum_{k=1}^{K} w_k \boldsymbol{\mu}_{q_k}. \tag{37}$$

Similarly, the covariance matrix of the aggregate pdf in (36), $\boldsymbol{\Sigma}_q = \mathbb{E}_q[(\boldsymbol{\theta} - \boldsymbol{\mu}_q)(\boldsymbol{\theta} - \boldsymbol{\mu}_q)^{\intercal}]$, is obtained as [174]

$$\boldsymbol{\Sigma}_q = \sum_{k=1}^{K} w_k \left(\boldsymbol{\Sigma}_{q_k} + (\boldsymbol{\mu}_{q_k} - \boldsymbol{\mu}_q)(\boldsymbol{\mu}_{q_k} - \boldsymbol{\mu}_q)^{\intercal}\right). \tag{38}$$

Thus, the mean and covariance matrix of the aggregate pdf $q(\boldsymbol{\theta})$ can be calculated easily from the agent means and covariance matrices. This is useful from a practical perspective because it provides a way for obtaining an estimate of the parameters (e.g., mean) as well as a measure of uncertainty for that estimate (e.g., covariance matrix). It is important to note, however, that since $q(\boldsymbol{\theta})$ is a mixture of Gaussians and, therefore, is non-Gaussian, it is not fully characterized by its mean and covariance matrix. Indeed, a mixture of Gaussians can have properties that a Gaussian cannot have, including heavy tails, multiple modes, and nonzero skewness [175].

In the case that the agent pdfs are Gaussian, the connection of linear opinion pooling to model averaging established in Section III-B1 extends to an estimation technique in the Kalman filtering literature called *multiple model adaptive estimation* (MMAE) [176]. MMAE uses a bank of Kalman filters to estimate an unknown state (time-varying parameter), where each Kalman filter assumes a distinct model describing the state's time evolution and its relation to the observed data. In this context, $\boldsymbol{\mu}_{q_k}$ is the local state estimate provided by the kth Kalman filter at a given time, while $\boldsymbol{\Sigma}_{q_k}$ is the covariance of that estimate. The local state estimates are then combined according to (37) to obtain a final state estimate $\boldsymbol{\mu}_q$, whose covariance $\boldsymbol{\Sigma}_q$ is determined by (38). Here, the weight w_k equals the posterior probability of the model assumed by the kth Kalman filter.

B. Log-linear Pooling

The fusion of the Gaussian pdfs $q_k(\boldsymbol{\theta})$ in (35) by the log-linear pooling function in (4) results in an aggregate pdf that is also Gaussian, i.e.,

$$q(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q),$$

with mean vector

$$\boldsymbol{\mu}_q = \left(\sum_{k=1}^{K} w_k \boldsymbol{\Sigma}_{q_k}^{-1}\right)^{-1} \sum_{j=1}^{K} w_j \boldsymbol{\Sigma}_{q_j}^{-1} \boldsymbol{\mu}_{q_j} \tag{39}$$

and covariance matrix

$$\boldsymbol{\Sigma}_q = \left(\sum_{k=1}^{K} w_k \boldsymbol{\Sigma}_{q_k}^{-1}\right)^{-1}. \tag{40}$$

Unlike the case of linear pooling, since the aggregate pdf $q(\boldsymbol{\theta})$ is Gaussian, it is unimodal and symmetric about the mean $\boldsymbol{\mu}_q$, and it is moreover fully characterized by the mean $\boldsymbol{\mu}_q$ and covariance $\boldsymbol{\Sigma}_q$.

There is a strong link between log-linear pooling of Gaussian pdfs and a second-order fusion method called *covariance intersection* [29], [32], which is often employed in distributed (decentralized) Kalman filter implementations [3], [41], [94]. In the covariance intersection context, there are K agents, each of which uses its own local observations to form a local estimate of an unknown quantity $\boldsymbol{\theta}$. The goal of covariance intersection is to fuse the local estimates in a way that does not underestimate the overall covariance of the fused estimate. Let $\boldsymbol{\mu}_{q_k}$ be the local estimate of the kth agent, whose covariance is denoted by $\boldsymbol{\Sigma}_{q_k}$. The fused state estimate $\boldsymbol{\mu}_q$ is determined according to (39), while the corresponding covariance matrix $\boldsymbol{\Sigma}_q$ is given by (40). The weights $w_1, \ldots, w_K$ used in (39) and (40) are typically chosen to minimize the determinant or the trace of $\boldsymbol{\Sigma}_q$ [29].

C. Other Pooling Functions

Finally, we consider the Hölder pooling functions. The normalization factor c in the Hölder pooling function in (8) for general $\alpha \in \mathbb{R} \setminus \{0,1\}$, involves an intractable integral and cannot be evaluated, even if the agent pdfs are Gaussian. Therefore, typically, the aggregate pdf $q(\boldsymbol{\theta})$ resulting from the Hölder pooling function is only known up to a normalization factor. Computing expected values with respect to $q(\boldsymbol{\theta})$ would require the use of numerical integration techniques such as the trapezoidal quadrature rule or Monte Carlo methods [177]. Because numerical integration techniques are plagued by the *curse of dimensionality* [178], computing expectations with respect to $q(\boldsymbol{\theta})$ under the Hölder pooling function becomes challenging when the dimension of $\boldsymbol{\theta}$ is large.

To illustrate the behavior of the linear and log-linear pooling functions, and to demonstrate the effect of different choices of α on the Hölder pooling function, we present in Fig. 5 simulation results for two different sets of $K = 2$ Gaussian agent pdfs $q_k(\theta)$ with $\theta \in \mathbb{R}$. We used the trapezoidal quadrature rule to compute the normalization factor of the aggregate pdf. Fig. 5a shows the fusion of two Gaussian pdfs with different means but the same variance. In this case, the value of α in the Hölder pooling function controls the multimodality of the aggregate pdf, in the sense that smaller (larger) values of α attenuate (enhance) the modes of the agent pdfs in the aggregate pdf. Fig. 5b shows the fusion of two Gaussian pdfs with the same mean but different variances. In this case, the value of α controls the shape of the tails of the aggregate pdf, in the sense that smaller (larger) values of α lead to less heavy (heavier) tails.

VII. Choosing the Pooling Parameters

An important consideration in opinion pooling is the choice of the parameters involved in the various pooling functions. While most of our discussion will be in regard to the weights $w_1, \dots, w_K$, we also provide some insight on the choice of the parameter α in the Hölder pooling function.

The problem of choosing the weights in probabilistic opinion pooling is well researched. The simplest approach is to assign equal weights to all agents, i.e., $w_k = 1/K$ for all k [55]. However, alternative strategies for assigning weights have been proposed for linear [179]–[181] and log-linear [159], [182], [183] pooling. These strategies are usually based on solving some optimization problem, where the definition of the objective function depends on how the weights are interpreted by the fusion center. In some instances, the optimization of the weights solely depends on the agent pdfs. In other scenarios, weight assignment takes into consideration data that are observed at the fusion center, and is based on a Bayesian interpretation involving likelihood functions or posterior distributions. These data-dependent methods have also been extended to the sequential case, where observed data are streamed and the weights are updated when new data become available [179].

In the following, we describe several options for choosing the weights in linear and log-linear pooling. We focus on methods that do not assume that the fusion center has observed any data. At this point, it is important to emphasize that in both the axiomatic and optimization approaches to probabilistic opinion pooling, the weights w_k were assumed fixed, i.e., not dependent on the agent pdfs $q_k(\boldsymbol{\theta})$. If, on the other hand, the weights are chosen adaptively according to an additional optimization procedure that involves the agent pdfs $q_k(\boldsymbol{\theta})$, then this implies a deviation from the strict mathematical framework established by both the axiomatic and optimization approaches. For example, the linear pooling function with adaptively chosen weights is no longer linear in the agent pdfs $q_k(\boldsymbol{\theta})$.

A. Linear Pooling

The problem of assigning the weights in the linear pooling function has been considered in many works; see [179] for a review. One approach is based on interpreting the weight w_k as a *veridical probability*, i.e, as the probability that the true pdf of $\boldsymbol{\theta}$ is $q_k(\boldsymbol{\theta})$ [184]. Accordingly, w_k is chosen to equal a prior or posterior estimate of that probability. This approach is connected to the model-averaging view of linear opinion pooling mentioned in Section III-B1, since in (2), $P(M_k)$ equals the probability that the model of the kth agent, M_k, is the correct one. When data are considered, the weights w_k equal the posterior probabilities of the models M_k, and this is exactly how they are assigned in the MMAE algorithm mentioned in Section VI-A [174], [176].

Alternatively, the weights can be assigned according to the predictive performance of each agent by viewing the weights as *outranking probabilities* [185]. In this view, w_k is the probability that predictions made based on $q_k(\boldsymbol{\theta})$ will outperform the predictions based on the pdfs of the other agents. This rationale for choosing the weights requires consideration of data and a mechanism for assessing the predictive performance of the agents.

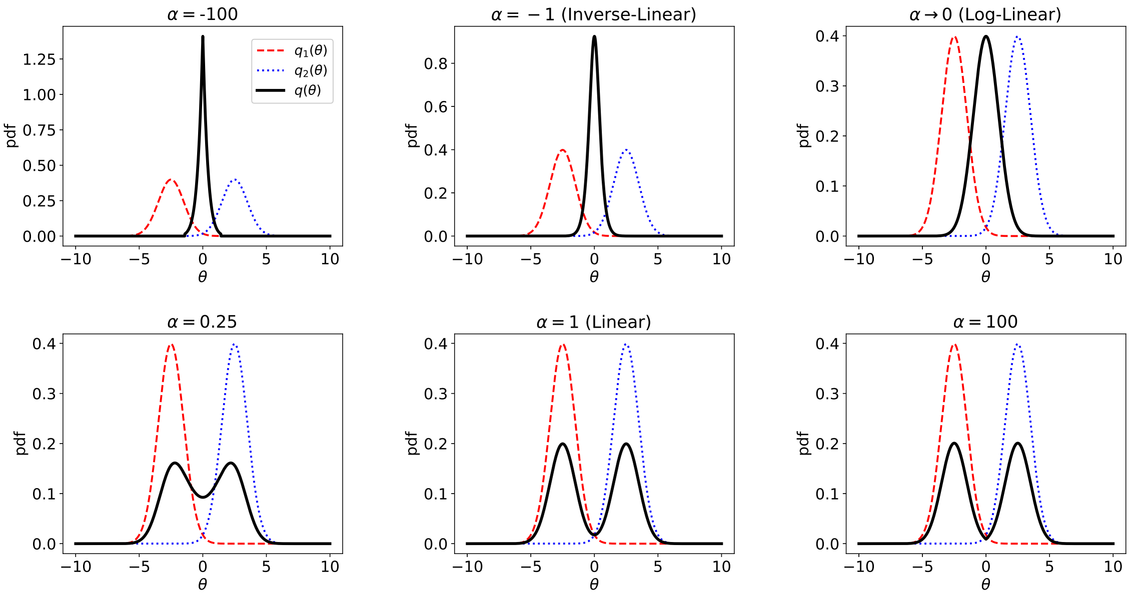
Another idea is to interpret the weights as a measure of distance [186]. Based on this interpretation, agents that have "middle of the road" opinions are assigned higher weights, while those that have more extreme (controversial) opinions are assigned lower weights. The opposite strategy would in principle also be possible, namely, giving more weight to controversial opinions. Such weight assignments can be achieved by assigning a nonnegative score γ_k to each agent pdf $q_k(\boldsymbol{\theta})$. For example, one can choose the score γ_k to be inversely related to the *maximum discrepancy* between agent k and the other agents, i.e.,

$$\gamma_k = \frac{1}{\max\limits_{j \in \{1,\dots,K\}} \mathcal{D}_{\mathrm{KL}}(q_k \| q_j)} \geq 0, \quad k = 1, \dots, K. \tag{41}$$

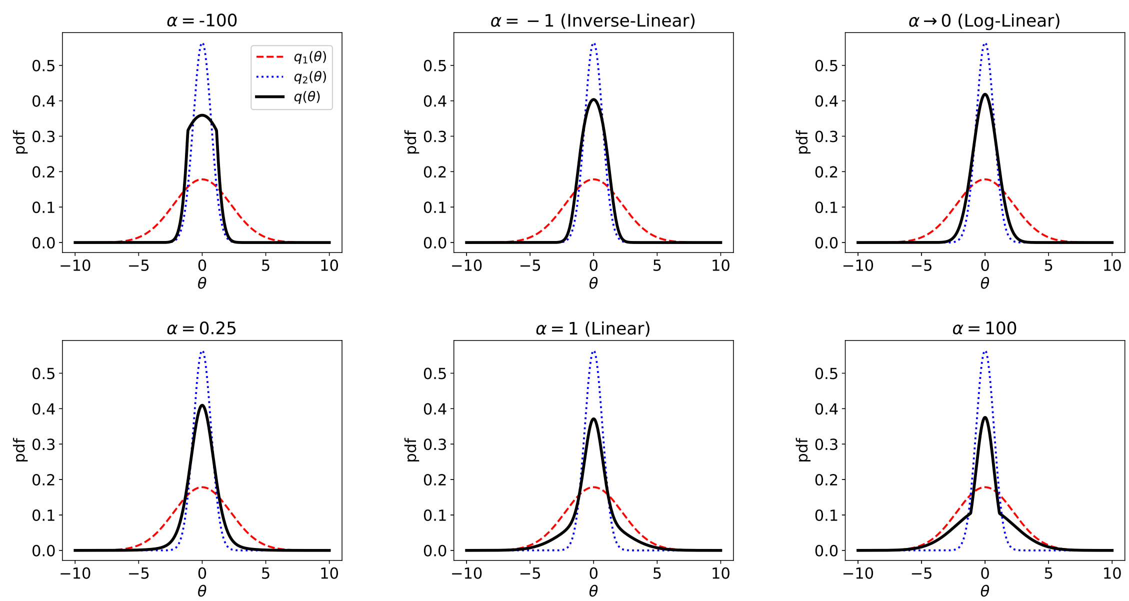
Here, the KLD is used to measure the discrepancy between agents, although other divergences can be used instead. The weight of each agent is then obtained as a normalized version of γ_k, i.e.,

$$w_k = \frac{\gamma_k}{\sum_{j=1}^{K} \gamma_j}, \quad k = 1, \dots, K.$$

Finally, there are also iterative schemes for weight assignment, where each agent considers itself to be a fusion center



(a) Different means, same variance.



(b) Same mean, different variances.

Figure 5: Results of Hölder pooling of two pdfs $q_1(\theta)$ and $q_2(\theta)$ using weights $w_1 = w_2 = 0.5$ and different values of α. The pdfs are defined as follows: (a) $q_1(\theta) = \mathcal{N}(\theta; -2.5, 1)$ and $q_2(\theta) = \mathcal{N}(\theta; 2.5, 1)$ (different means, same variance), and (b) $q_1(\boldsymbol{\theta}) = \mathcal{N}(\theta; 0, 5)$ and $q_2(\theta) = \mathcal{N}(\theta; 0, 0.5)$ (same mean, different variances). Note that $\alpha = -1$, $\alpha \to 0$, and $\alpha = 1$ correspond to the inverse-linear, log-linear, and linear pooling functions, respectively.

and assigns weights to all the other agents. The weights are iteratively updated until a consensus is reached. In [187], the weight vector of each agent is updated by multiplying it by a transition matrix, and under some conditions a consensus is reached asymptotically. The work [188] builds on this idea, but updates the weights according to how closely the agent pdfs agree, using a scoring function similar to (41).

B. Log-linear Pooling

The choice of the weights in the log-linear pooling function has been considered less intensely in the literature. Some of the aforementioned methods for linear opinion pooling can also be applied to log-linear opinion pooling; for example, the scoring rule in (41) is still reasonable. Moreover, as mentioned in Section VI-B, for Gaussian agent pdfs, log-linear pooling corresponds to the covariance intersection fusion method. Here, the weights can be chosen using schemes proposed in the covariance intersection literature, such as minimizing the trace or determinant of the covariance matrix in (40) [29].

One criterion proposed in the literature that does not require the consideration of data is the *minimum KLD* criterion [183]. If there is no basis for determining the reliability of each agent, one can choose the weights such that the aggregate pdf is maximally close to all the agent pdfs simultaneously. This is the criterion that was used in Section V to find an optimal pooling function for given weights w_k. Similarly to Section V-A, the criterion can be formulated as a minimization of the average of the KLDs between the agent pdfs $q_k(\boldsymbol{\theta})$ and the aggregate pdf $q(\boldsymbol{\theta})$. Introducing the weight vector $\boldsymbol{w} \triangleq (w_1, \ldots, w_K)$, the optimal weights are defined as

$$\boldsymbol{w}^\star = \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathcal{S}_K} L(\boldsymbol{w}),$$

with

$$\begin{aligned} L(\boldsymbol{w}) &\triangleq \frac{1}{K}\sum_{k=1}^{K} \mathcal{D}_{\mathrm{KL}}(q_k \| q) \\ &= \frac{1}{K}\sum_{k=1}^{K} \mathcal{D}_{\mathrm{KL}}\bigg(q_k \,\bigg\|\, c\prod_{\ell=1}^{K} (q_\ell(\boldsymbol{\theta}))^{w_\ell}\bigg), \end{aligned}$$

where expression (4) was inserted for $q(\boldsymbol{\theta})$. Using the KLD definition (23), one can obtain [183]

$$L(\boldsymbol{w}) = -\log c(\boldsymbol{w}) + \frac{1}{K}\sum_{k=1}^{K}\sum_{j\neq k} w_j \mathcal{D}_{\mathrm{KL}}(q_k\|q_j). \tag{42}$$

Here, $c(\boldsymbol{w})$ is the normalization factor in (5), which depends on $\boldsymbol{w}$. The objective function $L(\boldsymbol{w})$ is convex, since the first term $-\log c(\boldsymbol{w})$ is convex [159] and the second term is a linear function of $\boldsymbol{w}$. Therefore, tools from convex optimization can be used to compute the optimal weight vector $\boldsymbol{w}^\star$. We note that the minimum KLD criterion would also be a reasonable criterion for use with other pooling functions; however, the expression for $L(\boldsymbol{w})$ in (42) applies specifically to the log-linear pooling function.

Furthermore, we remark that if the average of the reverse KLDs, i.e.,

$$\begin{aligned} \tilde{L}(\boldsymbol{w}) &\triangleq \frac{1}{K}\sum_{k=1}^{K} \mathcal{D}_{\mathrm{KL}}(q \| q_k) \\ &= \frac{1}{K}\sum_{k=1}^{K} \mathcal{D}_{\mathrm{KL}}\bigg(c\prod_{\ell=1}^{K} (q_\ell(\boldsymbol{\theta}))^{w_\ell} \,\bigg\|\, q_k\bigg), \end{aligned} \tag{43}$$

was chosen as the objective function to be minimized, the optimal weights would be given by

$$\operatorname*{arg\,min}_{\boldsymbol{w}\in\mathcal{S}_K} \tilde{L}(\boldsymbol{w}) = \bigg(\frac{1}{K}, \ldots, \frac{1}{K}\bigg). \tag{44}$$

Indeed, let $q^\star(\boldsymbol{\theta})$ be defined by (4) with weights $\boldsymbol{w} = (\frac{1}{K}, \ldots, \frac{1}{K})$, i.e.,

$$q^\star(\boldsymbol{\theta}) \triangleq c\prod_{k=1}^{K} (q_k(\boldsymbol{\theta}))^{1/K}. \tag{45}$$

By Theorem 13, $q^\star(\boldsymbol{\theta})$ minimizes the objective function in (43) over all pdfs φ, i.e.,

$$q^\star = \operatorname*{arg\,min}_{\varphi\in\mathcal{P}} \frac{1}{K}\sum_{k=1}^{K} \mathcal{D}_{\mathrm{KL}}(\varphi\|q_k). \tag{46}$$

Thus, we have

$$\begin{aligned} \tilde{L}(\boldsymbol{w}) &\overset{(43)}{=} \frac{1}{K}\sum_{k=1}^{K} \mathcal{D}_{\mathrm{KL}}\bigg(c\prod_{\ell=1}^{K} (q_\ell(\boldsymbol{\theta}))^{w_\ell} \,\bigg\|\, q_k\bigg) \\ &\overset{(46)}{\geq} \frac{1}{K}\sum_{k=1}^{K} \mathcal{D}_{\mathrm{KL}}(q^\star\|q_k) \\ &\overset{(45)}{=} \frac{1}{K}\sum_{k=1}^{K} \mathcal{D}_{\mathrm{KL}}\bigg(c\prod_{\ell=1}^{K} (q_\ell(\boldsymbol{\theta}))^{1/K} \,\bigg\|\, q_k\bigg) \\ &\overset{(43)}{=} \tilde{L}\big((\tfrac{1}{K}, \ldots, \tfrac{1}{K})\big). \end{aligned}$$

Thus, for any $\boldsymbol{w}$, $\tilde{L}(\boldsymbol{w})$ is lower bounded by $\tilde{L}\big((\frac{1}{K}, \ldots, \frac{1}{K})\big)$. This proves (44).

Other approaches minimize an alternative KLD criterion [182], [189] or take a Bayesian approach by specifying a prior distribution over the weights [183]. However, these approaches require data to be available, and usually lead to closed form solutions only if the prior pdfs take the form of conjugate priors for the considered likelihood functions.

C. Hölder Pooling

In addition to the weights, the parameter α involved in the Hölder pooling function in (8) strongly impacts the resulting aggregate pdf, as was demonstrated in Fig. 5. An appropriate choice of α depends on the application at hand. For example, in risk assessment, the choice of α is relevant to a quantification of uncertainty. In a risk-averse scenario, one may opt to choose a larger value of α, or at least a positive α. Indeed, for any $\alpha > 0$, the supports of the agent pdfs are preserved by the fusion in the sense that the support of the aggregate pdf equals the union of the supports of all the agent pdfs. Furthermore, a larger α tends to yield a larger uncertainty in the aggregate pdf. This latter characteristic is related to the fact, shown in Fig. 5, that a larger α tends to promote multimodality and/or heavy-tailed properties in the aggregate pdf.

If one instead chooses a small value of α, then components of different agent pdfs that occur at different $\boldsymbol{\theta}$ locations will have substantially less influence on the aggregate pdf. This means, in particular, that an "outlier behavior" of one agent will tend to be attenuated in the fusion process. Furthermore, for $\alpha = 0$, if the pdf of any agent k is zero for some $\boldsymbol{\theta}_0$, i.e., $q_k(\boldsymbol{\theta}_0) = 0$, this implies that the aggregate pdf is also zero at $\boldsymbol{\theta}_0$ irrespectively of the values of the other agent pdfs. This "veto property" can be problematic in certain situations. Finally, for $\alpha < 0$, Hölder pooling is restricted to positive opinion profiles, which implies that all agents have to agree on the support Θ of $\boldsymbol{\theta}$.

Hölder pooling appears to be practically relevant mostly for values of α in $[0, 1]$. Here, we recall that $\alpha = 0$ and $\alpha = 1$ correspond to the log-linear pooling function and the linear pooling function, respectively; furthermore, values of α between 0 and 1 correspond to pooling functions whose characteristics—e.g., with regard to multimodality and tail decay—are intermediate between those of the linear and log-linear pooling functions, as demonstrated by Fig. 5. An application where this observation is potentially relevant was considered in Section II-A.

VIII. The Supra-Bayesian Framework

The supra-Bayesian framework is fundamentally different from the approaches discussed so far. In this section, we consider $\boldsymbol{\theta}$ to be a random variable with prior pdf $p(\boldsymbol{\theta})$ and assume that the fusion center follows a Bayesian update rule to derive a posterior pdf. Our focus will be on scenarios where observations (data) that depend on $\boldsymbol{\theta}$ are obtained by the agents but are not known to the fusion center. We will start this section with a formulation using conditionally independent observations, and extend from there to the general supra-Bayesian framework.

A. Agents Collecting Conditionally Independent Observations

Let us consider a scenario with K agents where each agent $k \in \{1, \ldots, K\}$ obtains observations $\mathbf{y}_k \in \mathbb{R}^{d_{y_k}}$. These observations are statistically related to the random vector $\boldsymbol{\theta} \in \mathbb{R}^{d_\theta}$ according to the "local" likelihood functions $p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$. We consider the observations fixed (i.e., already observed) and emphasize the dependence of $p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$ on $\boldsymbol{\theta}$ by writing the local likelihood functions as $\ell_k(\boldsymbol{\theta}) \triangleq p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$. Furthermore, each agent has access to the prior pdf $p(\boldsymbol{\theta})$ and is thus able to calculate its local posterior $\pi_k(\boldsymbol{\theta}) \triangleq p(\boldsymbol{\theta} \,|\, \mathbf{y}_k)$ according to Bayes' rule:

$$\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \,|\, \mathbf{y}_k) = \frac{\ell_k(\boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int_\Theta \ell_k(\boldsymbol{\theta}') p(\boldsymbol{\theta}') \, \mathrm{d}\boldsymbol{\theta}'} \,. \tag{47}$$

We further assume that the local observations $\mathbf{y}_k$ are conditionally independent given $\boldsymbol{\theta}$ for all $k \in \{1, \ldots, K\}$. This implies that the "global" likelihood function $\ell(\boldsymbol{\theta}) \triangleq p(\mathbf{y} \,|\, \boldsymbol{\theta})$ for $\mathbf{y} \triangleq [\mathbf{y}_1^\intercal, \ldots, \mathbf{y}_K^\intercal]^\intercal$ factors into the local likelihood functions $\ell_k(\boldsymbol{\theta}) = p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$, i.e.,

$$\ell(\boldsymbol{\theta}) = \prod_{k=1}^{K} \ell_k(\boldsymbol{\theta}) \,. \tag{48}$$

The task of the fusion center is to fuse the local posteriors $\pi_k(\boldsymbol{\theta})$ provided by the agents into an aggregate (fused) pdf $g[\pi_1, \ldots, \pi_K](\boldsymbol{\theta})$. We assume that the fusion center is aware of the statistical properties of all the observations (i.e., the conditional pdfs $p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$) and of the prior $p(\boldsymbol{\theta})$ but does not have access to the observations $\mathbf{y}_k$ directly. From a Bayesian viewpoint, the best possible fusion result is the posterior pdf of $\boldsymbol{\theta}$ using the observations from all the agents as represented by the total observation vector $\mathbf{y}$, i.e., $p(\boldsymbol{\theta} \,|\, \mathbf{y})$. We will refer to $p(\boldsymbol{\theta} \,|\, \mathbf{y})$ as *oracle posterior* because the fusion center does not know the observations $\mathbf{y}$ explicitly. Nevertheless, the following result shows that the fusion center is still able to fuse the $\pi_k(\boldsymbol{\theta})$ into the oracle posterior $p(\boldsymbol{\theta} \,|\, \mathbf{y})$.

Theorem 18. *Let $\boldsymbol{\theta}$ be a random vector with prior $p(\boldsymbol{\theta})$. Furthermore, let the local observations $\mathbf{y}_1, \ldots, \mathbf{y}_K$ given $\boldsymbol{\theta}$ be mutually independent and distributed according to $p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$. Then the global posterior $p(\boldsymbol{\theta} \,|\, \mathbf{y})$ with $\mathbf{y} = [\mathbf{y}_1^\intercal, \ldots, \mathbf{y}_K^\intercal]^\intercal$ is given by*

$$p(\boldsymbol{\theta} \,|\, \mathbf{y}) = g[\pi_1, \ldots, \pi_K](\boldsymbol{\theta}) = c \, (p(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} \pi_k(\boldsymbol{\theta}) \,, \tag{49}$$

where $c = 1/\int_\Theta (p(\boldsymbol{\theta}))^{1-K} \big(\prod_{k=1}^{K} \pi_k(\boldsymbol{\theta}) \big) \mathrm{d}\boldsymbol{\theta}$ is a normalization factor and the local posteriors $\pi_k(\boldsymbol{\theta})$ are given by (47).

Proof. We recall that $\ell(\boldsymbol{\theta}) = p(\mathbf{y} \,|\, \boldsymbol{\theta})$, $\ell_k(\boldsymbol{\theta}) = p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$, and $\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \,|\, \mathbf{y}_k)$. We have by Bayes' rule that

$$\begin{aligned} p(\boldsymbol{\theta} \,|\, \mathbf{y}) &\propto p(\boldsymbol{\theta}) \ell(\boldsymbol{\theta}) \\ &\overset{(48)}{=} p(\boldsymbol{\theta}) \prod_{k=1}^{K} \ell_k(\boldsymbol{\theta}) \\ &\propto p(\boldsymbol{\theta}) \prod_{k=1}^{K} \frac{p(\boldsymbol{\theta} \,|\, \mathbf{y}_k)}{p(\boldsymbol{\theta})} \\ &= (p(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} \pi_k(\boldsymbol{\theta}) \,. \end{aligned} \tag{50}$$

Since $p(\boldsymbol{\theta} \,|\, \mathbf{y})$ is a conditional pdf, normalizing the function in (50) gives (49). ☐

The fusion rule in (49) is recognized to be an instance of the multiplicative pooling function in (10), where the calibrating pdf $q_0(\boldsymbol{\theta})$ is given by the prior $p(\boldsymbol{\theta})$. Thus, Theorem 18 states that the multiplicative pooling function applied to the local posteriors $\pi_k(\boldsymbol{\theta})$ provides the oracle posterior $p(\boldsymbol{\theta} \,|\, \mathbf{y})$ in the case of conditionally independent local observations $\mathbf{y}_k$.

We note that the fusion center could calculate $p(\boldsymbol{\theta} \,|\, \mathbf{y})$ equally well from the local likelihood functions $\ell_k(\boldsymbol{\theta}) = p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$, rather than from the local posteriors $\pi_k(\boldsymbol{\theta})$. Indeed, the fusion rule (49) can be interpreted as first dividing each local posterior $\pi_k(\boldsymbol{\theta})$ by the prior $p(\boldsymbol{\theta})$ to obtain the local likelihood function $p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$, then fusing (multiplying) the local likelihood functions into the global likelihood function $p(\mathbf{y} \,|\, \boldsymbol{\theta})$, and finally multiplying by the prior to obtain the oracle posterior $p(\boldsymbol{\theta} \,|\, \mathbf{y})$. (This corresponds to reading the proof of Theorem 18 bottom up.) Thus, in the present scenario of conditionally independent observations $\mathbf{y}_k$, the agents may also communicate their local likelihood functions $\ell_k(\boldsymbol{\theta})$ to the fusion center, rather than their posteriors $\pi_k(\boldsymbol{\theta})$.

B. The Supra-Bayesian Framework and Local Statistics

To generalize the scenario considered in Section VIII-A, we take the perspective of the fusion center. In our Bayesian setting, the fusion center aims to calculate the posterior distribution of $\boldsymbol{\theta}$, given all the information it has access to. However, in more general settings than the case of conditionally independent observations discussed in Section VIII-A, we cannot expect that the fusion center is able to calculate the oracle posterior $p(\boldsymbol{\theta}\,|\,\mathbf{y})$. This is because the fusion center does not have direct access to the observations $\mathbf{y}_k$; rather, it observes the effect of the $\mathbf{y}_k$ only indirectly through the local posteriors $\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta}\,|\,\mathbf{y}_k)$. In addition to knowing the local posteriors $\pi_k(\boldsymbol{\theta})$, the fusion center is aware of the prior $p(\boldsymbol{\theta})$ and the conditional distribution $p(\mathbf{y}\,|\,\boldsymbol{\theta})$ (as a function of $\mathbf{y}$ and $\boldsymbol{\theta}$, not for the fixed, observed $\mathbf{y}$). Finally, the fusion center knows how the agents derive their local posteriors $\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta}\,|\,\mathbf{y}_k)$ given their local observations $\mathbf{y}_k$, i.e., it is aware that each π_k depends on $\boldsymbol{\theta}$ in a well-defined probabilistic way, namely, by the two-step process of first generating a random $\mathbf{y}_k$ given $\boldsymbol{\theta}$ according to the conditional pdf $p(\mathbf{y}_k\,|\,\boldsymbol{\theta})$ and then deriving π_k from $\mathbf{y}_k$ using (47).

This setup can be formulated generically via an abstract "observation model" $p(\pi_1,\ldots,\pi_K\,|\,\boldsymbol{\theta})$ in which the local posteriors π_K are considered as "observations." This approach is known in the literature as the supra-Bayesian model [87], [88]. In this abstract setting, we no longer have to consider the intermediate step of generating the observations $\mathbf{y}_k$ given $\boldsymbol{\theta}$, and we no longer have to assume that the local pdfs π_k are generated as posteriors. Instead, we directly define an observation model by specifying a probability distribution over the local pdfs π_k given $\boldsymbol{\theta}$. Thus, at the fusion center, the local pdfs of all agents are considered as observations, i.e., as random objects whose statistical relation to $\boldsymbol{\theta}$ is described by the "likelihood function" $p(\pi_1,\ldots,\pi_K\,|\,\boldsymbol{\theta})$. As always in Bayesian settings, we need in addition some prior $p(\boldsymbol{\theta})$. By Bayes' theorem, we can then express the posterior distribution of $\boldsymbol{\theta}$ given the local pdfs π_k as

$$p(\boldsymbol{\theta}\,|\,\pi_1,\ldots,\pi_K) = \frac{p(\pi_1,\ldots,\pi_K\,|\,\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_\Theta p(\pi_1,\ldots,\pi_K\,|\,\boldsymbol{\theta}')p(\boldsymbol{\theta}')\,\mathrm{d}\boldsymbol{\theta}'}, \tag{51}$$

which is considered to be the supra-Bayesian fusion result, also to be referred to as "supra-Bayesian posterior."

For any given $\boldsymbol{\theta}$, $p(\pi_1,\ldots,\pi_K\,|\,\boldsymbol{\theta})$ is a probability distribution over the infinite-dimensional space of functions that is given by the K-fold Cartesian product of the space of all pdfs $\mathcal{P}$. It is both mathematically and practically convenient to restrict to a finite-dimensional subset of this space. Indeed, a finite-dimensional parameterization is very often used in practical applications. In particular, if π_k depends deterministically on some finite-dimensional observation $\mathbf{y}_k$, then π_k is obviously restricted to a finite-dimensional subset. Thus, we will hereafter assume that each π_k depends deterministically and in a one-to-one manner on a finite-dimensional random vector $\mathbf{t}_k \in \mathbb{R}^{d_{t_k}}$. Then the probability distribution $p(\pi_1,\ldots,\pi_K\,|\,\boldsymbol{\theta})$ simplifies to a conventional conditional pdf $p(\mathbf{t}_1,\ldots,\mathbf{t}_K\,|\,\boldsymbol{\theta})$. This finite-dimensional setting is formalized by the following definition.

Definition 1. *A* finite-dimensional supra-Bayesian model *for a parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{d_\theta}$ consists of:*

- *a prior pdf $p(\boldsymbol{\theta})$;*
- *a conditional pdf $p(\mathbf{t}\,|\,\boldsymbol{\theta})$, where $\mathbf{t} = [\mathbf{t}_1^\intercal,\ldots,\mathbf{t}_K^\intercal]^\intercal$ with $\mathbf{t}_k \in \mathbb{R}^{d_{t_k}}$ for $k = 1,\ldots,K$;*
- *for each $k \in \{1,\ldots,K\}$, a one-to-one mapping $\psi_k\colon \mathbb{R}^{d_{t_k}} \to \mathcal{P}$.*

The vectors $\mathbf{t}_k$ are referred to as local statistics *and the functions $\pi_k(\boldsymbol{\theta}) = \psi_k[\mathbf{t}_k](\boldsymbol{\theta})$ as* local pdfs.

In a finite-dimensional supra-Bayesian model, each local pdf π_k is uniquely defined by a corresponding local statistic $\mathbf{t}_k$. As a consequence, the conditional distribution $p(\pi_1,\ldots,\pi_K\,|\,\boldsymbol{\theta})$ is implicitly given by the conditional pdf $p(\mathbf{t}\,|\,\boldsymbol{\theta})$ with $\mathbf{t} = [\mathbf{t}_1^\intercal,\ldots,\mathbf{t}_K^\intercal]^\intercal$, and we will refer to $\lambda(\boldsymbol{\theta}) \triangleq p(\mathbf{t}\,|\,\boldsymbol{\theta})$ as global likelihood function. The function ψ_k specifies which family of distributions π_k belongs to. For example, if we want to model the fact that $\pi_k(\boldsymbol{\theta})$ belongs to the family of Gaussian distributions with fixed and known covariance matrix $\boldsymbol{\Sigma}$, then we define $\psi_k[\boldsymbol{\mu}_k](\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta};\boldsymbol{\mu}_k,\boldsymbol{\Sigma})$. In this example, then, $\mathbf{t}_k = \boldsymbol{\mu}_k$.

In Definition 1, we further assumed that there is a one-to-one relation between the local pdf π_k and $\mathbf{t}_k$, i.e., two different vectors $\mathbf{t}_k$ and $\widetilde{\mathbf{t}}_k$ correspond to different pdfs π_k and $\widetilde{\pi}_k$. In addition to the fact that the pdf π_k is uniquely specified by the vector $\mathbf{t}_k$, this assumption also implies that we can uniquely determine $\mathbf{t}_k$ from π_k, i.e., $\mathbf{t}_k$ is a function of π_k and we can thus interpret it as a statistic of π_k. This justifies the designation of the vectors $\mathbf{t}_k$ as local statistics. In summary, the local statistic $\mathbf{t}_k$ represents the information provided by the pdf π_k of agent k in a more accessible, finite-dimensional way.

The following result is an immediate consequence of our definition of a finite-dimensional supra-Bayesian model (Definition 1) and Bayes' theorem: the one-to-one relationship between π_k and $\mathbf{t}_k$ for each $k \in \{1,\ldots,K\}$ implies that $p(\boldsymbol{\theta}\,|\,\pi_1,\ldots,\pi_K) = p(\boldsymbol{\theta}\,|\,\mathbf{t})$, and Bayes' theorem implies that $p(\boldsymbol{\theta}\,|\,\mathbf{t}) \propto p(\mathbf{t}\,|\,\boldsymbol{\theta})p(\boldsymbol{\theta})$.

Theorem 19. *In a finite-dimensional supra-Bayesian model, the supra-Bayesian fusion result (or supra-Bayesian posterior) is given by*

$$p(\boldsymbol{\theta}\,|\,\pi_1,\ldots,\pi_K) = p(\boldsymbol{\theta}\,|\,\mathbf{t}) = \frac{\lambda(\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_\Theta \lambda(\boldsymbol{\theta}')p(\boldsymbol{\theta}')\mathrm{d}\boldsymbol{\theta}'}\,, \tag{52}$$

where $\lambda(\boldsymbol{\theta}) = p(\mathbf{t}\,|\,\boldsymbol{\theta})$.

Since the fusion center knows $\mathbf{t}$, $p(\mathbf{t}\,|\,\boldsymbol{\theta})$, and $p(\boldsymbol{\theta})$, it is able to calculate (52). However, in general, (52) does not provide an explicit rule for fusing the pdfs $\pi_k(\boldsymbol{\theta})$ into the supra-Bayesian posterior $p(\boldsymbol{\theta}\,|\,\mathbf{t})$, i.e., it does not specify a pooling function g such that $p(\boldsymbol{\theta}\,|\,\mathbf{t}) = g[\pi_1,\ldots,\pi_K](\boldsymbol{\theta})$. Nevertheless, we can already deduce an interesting fact from the structure of (52): The supra-Bayesian posterior is proportional to the product of the prior $p(\boldsymbol{\theta})$ and the global likelihood function $\lambda(\boldsymbol{\theta})$, and thus depends on the pdfs $\pi_k(\boldsymbol{\theta})$ only indirectly via the global likelihood function $\lambda(\boldsymbol{\theta})$. Hence, the actual task in supra-Bayesian fusion is to establish a rule for obtaining the global likelihood function $\lambda(\boldsymbol{\theta}) = p(\mathbf{t}\,|\,\boldsymbol{\theta})$ from the local posteriors $\pi_k(\boldsymbol{\theta})$ or, equivalently, from the vector of local statistics $\mathbf{t} = [\mathbf{t}_1^\intercal, \ldots, \mathbf{t}_K^\intercal]^\intercal$. In what follows, we will see that this approach can result in interesting fusion rules for specific scenarios. In particular, we will consider conditionally independent agents in Section VIII-C and dependent agents in Section VIII-D. Furthermore, the special case given by the linear Gaussian model will be studied in Section IX.

C. Supra-Bayesian Fusion for Conditionally Independent Agents

Generalizing the scenario in Section VIII-A, we assume that, given $\boldsymbol{\theta}$, the information provided by each agent to the fusion center is conditionally independent of the information provided by the other agents. In our finite-dimensional supra-Bayesian model this means that the $\mathbf{t}_k$ are conditionally independent given $\boldsymbol{\theta}$, i.e., the global likelihood function $\lambda(\boldsymbol{\theta})$ factors according to

$$\lambda(\boldsymbol{\theta}) = p(\mathbf{t}\,|\,\boldsymbol{\theta}) = \prod_{k=1}^{K} p(\mathbf{t}_k\,|\,\boldsymbol{\theta}) = \prod_{k=1}^{K} \lambda_k(\boldsymbol{\theta}), \tag{53}$$

where we introduced the local likelihood functions $\lambda_k(\boldsymbol{\theta}) \triangleq p(\mathbf{t}_k\,|\,\boldsymbol{\theta})$. Because conditional independence of the $\mathbf{t}_k$ is equivalent to conditional independence of the random local pdfs π_k, we immediately obtain the following corollary by inserting (53) into (52).

Corollary 20. *In a finite-dimensional supra-Bayesian model where the local pdfs π_k are conditionally independent given $\boldsymbol{\theta}$, the supra-Bayesian fusion result (or supra-Bayesian posterior) is given by*

$$p(\boldsymbol{\theta}\,|\,\pi_1, \ldots, \pi_K) = p(\boldsymbol{\theta}\,|\,\mathbf{t}) = \frac{\big(\prod_{k=1}^{K} \lambda_k(\boldsymbol{\theta})\big) p(\boldsymbol{\theta})}{\int_\Theta \big(\prod_{k=1}^{K} \lambda_k(\boldsymbol{\theta}')\big) p(\boldsymbol{\theta}')\, \mathrm{d}\boldsymbol{\theta}'}, \tag{54}$$

where $\lambda_k(\boldsymbol{\theta}) = p(\mathbf{t}_k\,|\,\boldsymbol{\theta})$.

To establish a link to the scenario of Section VIII-A, let us consider the local statistics $\mathbf{t}_k$ and the global likelihood function $\lambda(\boldsymbol{\theta}) = p(\mathbf{t}\,|\,\boldsymbol{\theta})$ in that scenario. Recall that in Section VIII-A, we assumed that each agent has observations $\mathbf{y}_k \in \mathbb{R}^{d_{y_k}}$ related to $\boldsymbol{\theta}$ according to the local observation likelihood function $\ell_k(\boldsymbol{\theta}) = p(\mathbf{y}_k\,|\,\boldsymbol{\theta})$, and these observations are conditionally independent given $\boldsymbol{\theta}$. The local pdfs $\pi_k(\boldsymbol{\theta})$—which, in this scenario, are the local posteriors $p(\boldsymbol{\theta}\,|\,\mathbf{y}_k)$—are given by (47), and they are thus parametrized by the local observations $\mathbf{y}_k$. However, in common observation models, the observations $\mathbf{y}_k$ cannot be uniquely reconstructed from the posterior pdf $\pi_k(\boldsymbol{\theta})$. Indeed, local statistics $\mathbf{t}_k$ that parametrize the local posteriors $\pi_k(\boldsymbol{\theta})$ in a one-to-one manner are usually obtained as some function $T_k(\mathbf{y}_k)$ of the observations, where $T_k \colon \mathbb{R}^{d_{y_k}} \to \mathbb{R}^{d_{t_k}}$ with $d_{t_k} \leq d_{y_k}$ is in general not invertible. The random variable $\mathbf{t}_k = T_k(\mathbf{y}_k)$ is then a sufficient statistic [190, Sec. 6.2] of $\mathbf{y}_k$ for $\boldsymbol{\theta}$, i.e.,

$$p(\boldsymbol{\theta}\,|\,\mathbf{y}_k) = p(\boldsymbol{\theta}\,|\,\mathbf{t}_k)\,. \tag{55}$$

Thus, our local statistic $\mathbf{t}_k$ uniquely parametrizing the local posterior $\pi_k(\boldsymbol{\theta})$ is given by $\mathbf{t}_k = T_k(\mathbf{y}_k)$, with a noninvertible, possibly dimension-reducing function T_k. The local statistics $\mathbf{t}_k$ given $\boldsymbol{\theta}$ are conditionally independent for $k = 1, \ldots, K$ because they are deterministic functions of the conditionally independent observations $\mathbf{y}_k$. Hence, the factorization (53) holds, and indeed we have a finite-dimensional supra-Bayesian model with a prior $p(\boldsymbol{\theta})$, a likelihood function $p(\mathbf{t}\,|\,\boldsymbol{\theta})$, and local pdfs π_k that are given by $\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta}\,|\,\mathbf{t}_k)$, i.e., $\psi_k[\mathbf{t}_k](\boldsymbol{\theta}) = p(\boldsymbol{\theta}\,|\,\mathbf{t}_k)$. Thus, the supra-Bayesian fusion result $p(\boldsymbol{\theta}\,|\,\mathbf{t})$ is given by the expression in (54). We will now demonstrate that $p(\boldsymbol{\theta}\,|\,\mathbf{t})$ coincides with the fusion result given in (49). Recalling that $\lambda_k(\boldsymbol{\theta}) = p(\mathbf{t}_k\,|\,\boldsymbol{\theta})$, the supra-Bayesian fusion result (54) becomes

$$\begin{aligned} p(\boldsymbol{\theta}\,|\,\mathbf{t}) &\propto \bigg(\prod_{k=1}^{K} p(\mathbf{t}_k\,|\,\boldsymbol{\theta})\bigg) p(\boldsymbol{\theta}) \\ &\propto \bigg(\prod_{k=1}^{K} \frac{p(\boldsymbol{\theta}\,|\,\mathbf{t}_k)}{p(\boldsymbol{\theta})}\bigg) p(\boldsymbol{\theta}) \\ &= (p(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} p(\boldsymbol{\theta}\,|\,\mathbf{t}_k), \end{aligned}$$

where we used Bayes' theorem. By (55), we further have

$$\begin{aligned} p(\boldsymbol{\theta}\,|\,\mathbf{t}) &\propto (p(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} p(\boldsymbol{\theta}\,|\,\mathbf{y}_k) \\ &= (p(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} \pi_k(\boldsymbol{\theta})\,, \end{aligned} \tag{56}$$

which indeed equals the fusion rule (49). In particular, a comparison with (49) shows that for conditionally independent $\mathbf{y}_k$, the supra-Bayesian posterior $p(\boldsymbol{\theta}\,|\,\mathbf{t})$ coincides with the oracle posterior $p(\boldsymbol{\theta}\,|\,\mathbf{y})$. Thus, in this case, $\mathbf{t}$ is a sufficient statistic of $\mathbf{y}$ for $\boldsymbol{\theta}$.

Example 1 (Exponential Families). A convenient and versatile class of likelihood functions is given by exponential families [191]. We thus specialize the results discussed above to these models. A local observation likelihood function of the exponential family type can be written as

$$p(\mathbf{y}_k\,|\,\boldsymbol{\theta}) = h_k(\mathbf{y}_k) \exp\big(\boldsymbol{\eta}(\boldsymbol{\theta})^\intercal T_k(\mathbf{y}_k) - A_k(\boldsymbol{\theta})\big), \tag{57}$$

with some functions $h_k(\mathbf{y}_k) \geq 0$, $\boldsymbol{\eta}(\boldsymbol{\theta}) \in \mathbb{R}^{d_\theta}$, and $T_k(\mathbf{y}_k) \in \mathbb{R}^{d_\theta}$. The function $A_k(\boldsymbol{\theta})$ is determined by the other functions via the fact that $p(\mathbf{y}_k\,|\,\boldsymbol{\theta})$ is normalized. We assume that the observations $\mathbf{y}_k$ are conditionally independent given $\boldsymbol{\theta}$. Furthermore, the fusion center is supposed to know the conditional pdfs $p(\mathbf{y}_k\,|\,\boldsymbol{\theta})$ in terms of the functions $\boldsymbol{\eta}$, h_k, T_k, and A_k for all k (but, as always, it does not know the $\mathbf{y}_k$), and to be also aware of the prior $p(\boldsymbol{\theta})$.

It is known that the local statistic $\mathbf{t}_k = T_k(\mathbf{y}_k)$ is a sufficient statistic of $\mathbf{y}_k$ for $\boldsymbol{\theta}$ [191, Prop. 1.5]. To verify that there is a one-to-one relation between the local posterior π_k and $\mathbf{t}_k$, we have to show that $\mathbf{t}_k$ can be recovered from π_k. We have

$$\pi_k(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}) p(\mathbf{y}_k\,|\,\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}) \exp(\boldsymbol{\eta}(\boldsymbol{\theta})^\intercal \mathbf{t}_k - A_k(\boldsymbol{\theta}))\,. \tag{58}$$

Then

$$\log\bigg(\frac{\pi_k(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} \exp(A_k(\boldsymbol{\theta}))\bigg) = \boldsymbol{\eta}(\boldsymbol{\theta})^\intercal \mathbf{t}_k + C\,, \tag{59}$$

where C is a constant that does not depend on $\boldsymbol{\theta}$. To be able to solve (59) for $\mathbf{t}_k$ and C, we make the technical assumption that there exist $d_\theta + 1$ different $\boldsymbol{\theta}_j$ such that the matrix

$$\mathbf{B} \triangleq \begin{pmatrix} \boldsymbol{\eta}(\boldsymbol{\theta}_1)^\intercal & 1 \\ \vdots & \vdots \\ \boldsymbol{\eta}(\boldsymbol{\theta}_{d_\theta+1})^\intercal & 1 \end{pmatrix} \in \mathbb{R}^{(d_\theta+1)\times(d_\theta+1)} \tag{60}$$

is nonsingular. Then, evaluating (59) at $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{d_\theta+1}$ gives a system of $d_\theta + 1$ equations that can be written as

$$\mathbf{B}\begin{pmatrix} \mathbf{t}_k \\ C \end{pmatrix} = \begin{pmatrix} \log\left(\frac{\pi_k(\boldsymbol{\theta}_1)}{p(\boldsymbol{\theta}_1)} \exp(A_k(\boldsymbol{\theta}_1))\right) \\ \vdots \\ \log\left(\frac{\pi_k(\boldsymbol{\theta}_{d_\theta+1})}{p(\boldsymbol{\theta}_{d_\theta+1})} \exp(A_k(\boldsymbol{\theta}_{d_\theta+1}))\right) \end{pmatrix}.$$

Because $\mathbf{B}$ is nonsingular, this equation can be solved for $\mathbf{t}_k$ and C. Thus, we are able to recover $\mathbf{t}_k$ from π_k. We conclude that our exponential family model is a finite-dimensional supra-Bayesian model.

Using (58) in (56), the supra-Bayesian fusion result is obtained as

$$\begin{aligned} p(\boldsymbol{\theta} \,|\, \mathbf{t}) &\propto p(\boldsymbol{\theta})^{1-K} \prod_{k=1}^{K} p(\boldsymbol{\theta}) \exp\big(\boldsymbol{\eta}(\boldsymbol{\theta})^\intercal \mathbf{t}_k - A_k(\boldsymbol{\theta})\big) \\ &= p(\boldsymbol{\theta}) \exp\big(\boldsymbol{\eta}(\boldsymbol{\theta})^\intercal \bar{\mathbf{t}} - \bar{A}(\boldsymbol{\theta})\big), \end{aligned} \tag{61}$$

with

$$\bar{\mathbf{t}} = \sum_{k=1}^{K} \mathbf{t}_k, \quad \bar{A}(\boldsymbol{\theta}) = \sum_{k=1}^{K} A_k(\boldsymbol{\theta}).$$

We see that, for conditionally independent observations $\mathbf{y}_k$, $p(\boldsymbol{\theta} \,|\, \mathbf{t})$ depends on the observations $\mathbf{y}_k$ only via the local statistics $\mathbf{t}_k = T_k(\mathbf{y}_k)$, and furthermore, supra-Bayesian fusion essentially amounts to the summation of the local statistics $\mathbf{t}_k$ and of the normalization functions $A_k(\boldsymbol{\theta})$.

This simple summation rule is augmented when the prior $p(\boldsymbol{\theta})$ is chosen as

$$p(\boldsymbol{\theta}) \propto \exp(\boldsymbol{\eta}(\boldsymbol{\theta})^\intercal \mathbf{t}_0 - A_0(\boldsymbol{\theta})), \tag{62}$$

for some vector $\mathbf{t}_0$ and function $A_0(\boldsymbol{\theta})$. Inserting (62) into (61), we obtain

$$p(\boldsymbol{\theta} \,|\, \mathbf{t}) \propto \exp\big(\boldsymbol{\eta}(\boldsymbol{\theta})^\intercal \mathbf{t}_{\text{post}} - A_{\text{post}}(\boldsymbol{\theta})\big), \tag{63}$$

with

$$\mathbf{t}_{\text{post}} = \bar{\mathbf{t}} + \mathbf{t}_0 = \sum_{k=0}^{K} \mathbf{t}_k \tag{64}$$

and

$$A_{\text{post}}(\boldsymbol{\theta}) = \bar{A}(\boldsymbol{\theta}) + A_0(\boldsymbol{\theta}) = \sum_{k=0}^{K} A_k(\boldsymbol{\theta}). \tag{65}$$

In particular, when all $A_k(\boldsymbol{\theta})$ for $k = 1, \ldots, K$ are equal to the same $A(\boldsymbol{\theta})$ and $A_0(\boldsymbol{\theta}) = a_0 A(\boldsymbol{\theta})$, then the prior becomes the conjugate prior [191, Def. 4.18]

$$p(\boldsymbol{\theta}) \propto \exp(\boldsymbol{\eta}(\boldsymbol{\theta})^\intercal \mathbf{t}_0 - a_0 A(\boldsymbol{\theta})),$$

with the two hyperparameters $\mathbf{t}_0$ and $a_0 > 0$. Here, the supra-Bayesian fusion result simplifies to

$$p(\boldsymbol{\theta} \,|\, \mathbf{t}) \propto \exp\big(\boldsymbol{\eta}(\boldsymbol{\theta})^\intercal \mathbf{t}_{\text{post}} - (K + a_0) A(\boldsymbol{\theta})\big).$$

We see that $p(\boldsymbol{\theta} \,|\, \mathbf{t})$ has the same form as the prior $p(\boldsymbol{\theta})$, while the hyperparameters $\mathbf{t}_0$ and a_0 are replaced by $\mathbf{t}_{\text{post}} = \mathbf{t}_0 + \bar{\mathbf{t}}$ and $a_0 + K$, respectively.

An important special case of the exponential family setting is given by linear Gaussian observations. This case will be considered in Section IX, both for conditionally dependent and independent observations (see in particular Example 2 in Section IX-A).

D. Supra-Bayesian Fusion for Agents Collecting Dependent Observations

Similar to the setting of independent agents studied above, we consider K agents that obtain observations $\mathbf{y}_k \in \mathbb{R}^{d_{y_k}}$ distributed according to the local observation likelihood functions $p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$, with $k \in \{1, \ldots, K\}$. Again, each agent has access also to the prior pdf $p(\boldsymbol{\theta})$, and the local posterior pdfs $\pi_k(\boldsymbol{\theta})$ are still given by (47). However, in contrast to the previous subsection, we do not assume that the observations are conditionally independent. We assume that the fusion center is aware of the conditional pdf[7] $p(\mathbf{y} \,|\, \boldsymbol{\theta})$ of all observations $\mathbf{y} = [\mathbf{y}_1^\intercal, \ldots, \mathbf{y}_K^\intercal]^\intercal$ given $\boldsymbol{\theta}$, the prior pdf $p(\boldsymbol{\theta})$, and the local posterior pdfs $\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \,|\, \mathbf{y}_k)$. We emphasize that although the fusion center has access to $p(\mathbf{y} \,|\, \boldsymbol{\theta})$ as a function of $\mathbf{y}$ and $\boldsymbol{\theta}$, it does not know the global observation $\mathbf{y}$ and thus cannot use $p(\mathbf{y} \,|\, \boldsymbol{\theta})$ as a global likelihood function.

To establish a supra-Bayesian fusion scheme for this scenario, we again consider a finite-dimensional supra-Bayesian model, i.e., for each agent k there exists a local statistic $\mathbf{t}_k$ such that $\pi_k(\boldsymbol{\theta}) = \psi_k[\mathbf{t}_k](\boldsymbol{\theta})$, and there is a one-to-one relation between $\mathbf{t}_k$ and the local posterior π_k. Because $\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \,|\, \mathbf{y}_k)$, the local pdf π_k is also uniquely determined by $\mathbf{y}_k$, and thus the one-to-one relation between π_k and $\mathbf{t}_k$ implies that there exists a function $T_k \colon \mathbb{R}^{d_{y_k}} \to \mathbb{R}^{d_{t_k}}$ such that $\mathbf{t}_k = T_k(\mathbf{y}_k)$. As before, the function T_k is not one-to-one in general, i.e., it is not possible to recover $\mathbf{y}_k$ from $\mathbf{t}_k$. However, $\mathbf{t}_k$ is again a sufficient statistic of $\mathbf{y}_k$ for $\boldsymbol{\theta}$, i.e., $p(\boldsymbol{\theta} \,|\, \mathbf{y}_k) = p(\boldsymbol{\theta} \,|\, \mathbf{t}_k)$.

Because the local observations $\mathbf{y}_k$ are subvectors of the global observation $\mathbf{y} = [\mathbf{y}_1^\intercal, \ldots, \mathbf{y}_K^\intercal]^\intercal \in \mathbb{R}^{d_y}$, we can introduce $T \colon \mathbb{R}^{d_y} \to \mathbb{R}^{\sum_{k=1}^{K} d_{t_k}}$ as

$$T(\mathbf{y}) = [T_1(\mathbf{y}_1)^\intercal, \ldots, T_K(\mathbf{y}_K)^\intercal]^\intercal,$$

and thus we have

$$\mathbf{t} = [\mathbf{t}_1^\intercal, \ldots, \mathbf{t}_K^\intercal]^\intercal = T(\mathbf{y}).$$

The random vector $\mathbf{t}$ summarizes all the information that the agents communicate to the fusion center, and it is thus known to the fusion center (whereas $\mathbf{y}$ is not). Note that although each $\mathbf{t}_k$ is a sufficient statistic of $\mathbf{y}_k$ for $\boldsymbol{\theta}$, the global statistic $\mathbf{t}$ is, in general, not a sufficient statistic of $\mathbf{y}$. This is due to

[7]Note that the conditional pdfs $p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$ are marginals of the conditional pdf $p(\mathbf{y} \,|\, \boldsymbol{\theta})$.

the fact that $\mathbf{t}$ generally does not capture all the dependencies between the individual $\mathbf{y}_k$.

Because $\mathbf{t} = T(\mathbf{y})$, we can use the general change-of-variables formula [192, Sec. 3.4.3] to calculate the conditional pdf $p(\mathbf{t} \,|\, \boldsymbol{\theta})$ from the conditional pdf $p(\mathbf{y} \,|\, \boldsymbol{\theta})$, provided the function T is differentiable. Since $\mathbf{t}$ summarizes the information communicated by the agents to the fusion center, $\lambda(\boldsymbol{\theta}) = p(\mathbf{t} \,|\, \boldsymbol{\theta})$ is the global likelihood function that the fusion center has to use in the calculation of the supra-Bayesian posterior $p(\boldsymbol{\theta} \,|\, \mathbf{t})$ according to (52). Therefore, to obtain the supra-Bayesian fusion rule $g[\pi_1, \ldots, \pi_K]$, based on (52), we have to perform the following three steps:

1) Identify the local statistics $\mathbf{t}_k$ that uniquely represent the local posterior pdfs π_k within the given statistical model;
2) apply the general change-of-variables formula to transform the (known) conditional pdf $p(\mathbf{y} \,|\, \boldsymbol{\theta})$ into the global likelihood function $\lambda(\boldsymbol{\theta}) = p(\mathbf{t} \,|\, \boldsymbol{\theta})$;
3) calculate the supra-Bayesian posterior $p(\boldsymbol{\theta} \,|\, \mathbf{t})$ according to (52).

While this three-step process can in principle be performed in any setting satisfying our assumptions, an explicit characterization of the resulting supra-Bayesian fusion rule (pooling function) $g[\pi_1, \ldots, \pi_K]$ can only be derived for special cases. The important case of a linear Gaussian model will be explored in the following.

IX. Supra-Bayesian Fusion for the Linear Gaussian Model

We consider supra-Bayesian pdf fusion for the linear observation model

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n}, \tag{66}$$

where $\mathbf{H} \in \mathbb{R}^{d_y \times d_\theta}$ is a known observation matrix and $\mathbf{n} \in \mathbb{R}^{d_y}$ is additive zero-mean Gaussian noise with a known covariance matrix $\boldsymbol{\Sigma}$, i.e., $p(\mathbf{n}) = \mathcal{N}(\mathbf{n}; \mathbf{0}, \boldsymbol{\Sigma})$. Thus, $\mathbf{y}$ given $\boldsymbol{\theta}$ is Gaussian distributed with mean $\mathbf{H}\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e.,

$$p(\mathbf{y} \,|\, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}; \mathbf{H}\boldsymbol{\theta}, \boldsymbol{\Sigma})\,. \tag{67}$$

The local observation at agent k is given as $\mathbf{y}_k = \mathbf{H}_k\boldsymbol{\theta} + \mathbf{n}_k \in \mathbb{R}^{d_{y_k}}$, where

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_K \end{pmatrix}, \tag{68}$$

with $\mathbf{H}_k \in \mathbb{R}^{d_{y_k} \times d_\theta}$, and $\mathbf{n} = [\mathbf{n}_1^\intercal, \ldots, \mathbf{n}_K^\intercal]^\intercal$. Thus, each local observation $\mathbf{y}_k$ given $\boldsymbol{\theta}$ is again Gaussian with mean $\mathbf{H}_k\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}_{kk} \in \mathbb{R}^{d_{y_k} \times d_{y_k}}$. We note that the overall covariance matrix $\boldsymbol{\Sigma}$ is block-structured according to

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1K} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{K1} & \cdots & \boldsymbol{\Sigma}_{KK} \end{pmatrix}, \tag{69}$$

where the off-diagonal cross-covariance matrices $\boldsymbol{\Sigma}_{kk'}$ for $k \neq k'$ describe the conditional dependency between the observations of different agents. The case of conditionally independent observations $\mathbf{y}_k$ is obtained for $\boldsymbol{\Sigma}_{kk'} = \mathbf{0}$ for all $k \neq k'$. For simplicity, we further assume that for all $k = 1, \ldots, K$, $d_{y_k} \geq d_\theta$, $\mathbf{H}_k$ has full rank, and $\boldsymbol{\Sigma}_{kk}$ is positive definite. The local observation likelihood functions are here given by

$$\begin{aligned} \ell_k(\boldsymbol{\theta}) &= p(\mathbf{y}_k \,|\, \boldsymbol{\theta}) \\ &= \mathcal{N}(\mathbf{y}_k; \mathbf{H}_k\boldsymbol{\theta}, \boldsymbol{\Sigma}_{kk}) \\ &\propto \exp\left(-\frac{(\mathbf{y}_k - \mathbf{H}_k\boldsymbol{\theta})^\intercal \boldsymbol{\Sigma}_{kk}^{-1}(\mathbf{y}_k - \mathbf{H}_k\boldsymbol{\theta})}{2}\right). \end{aligned} \tag{70}$$

A. Local Statistics

We can rewrite (70) as

$$\begin{aligned} \ell_k(\boldsymbol{\theta}) &\propto \exp\left(-\frac{(\boldsymbol{\theta} - \mathbf{V}_k\mathbf{y}_k)^\intercal \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{H}_k(\boldsymbol{\theta} - \mathbf{V}_k\mathbf{y}_k)}{2}\right) \\ &= \exp\left(-\frac{(\boldsymbol{\theta} - \mathbf{t}_k)^\intercal \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{H}_k(\boldsymbol{\theta} - \mathbf{t}_k)}{2}\right), \end{aligned} \tag{71}$$

where

$$\mathbf{V}_k = (\mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{H}_k)^{-1} \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \tag{72}$$

and

$$\mathbf{t}_k = \mathbf{V}_k\mathbf{y}_k = (\mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{H}_k)^{-1} \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{y}_k. \tag{73}$$

The proportionality in (71) is as a function of $\boldsymbol{\theta}$, i.e., the proportionality constant will depend on $\mathbf{y}_k$.

We claim that $\mathbf{t}_k$ in (73) qualifies as a local statistic in a finite-dimensional supra-Bayesian model. For a proof, we note that the local posteriors are again given as $\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \,|\, \mathbf{y}_k) \propto \ell_k(\boldsymbol{\theta}) p(\boldsymbol{\theta})$. To see that there is a one-to-one relation between the local posterior π_k and the finite-dimensional parameter $\mathbf{t}_k \in \mathbb{R}^{d_\theta}$, recall that the fusion center is aware of the prior $p(\boldsymbol{\theta})$ and the matrices $\mathbf{H}$ and $\boldsymbol{\Sigma}$. In particular, the fusion center is aware of $\mathbf{H}_k$ and $\boldsymbol{\Sigma}_{kk}$, and thus it is able to recover from $\mathbf{t}_k$ the local observation likelihood function $\ell_k(\boldsymbol{\theta})$ in (71) and, in turn, the local posterior $\pi_k(\boldsymbol{\theta}) \propto \ell_k(\boldsymbol{\theta}) p(\boldsymbol{\theta})$. Conversely, the fusion center is able to obtain $\mathbf{t}_k$ from the local posterior $\pi_k(\boldsymbol{\theta})$ by first dividing by the prior $p(\boldsymbol{\theta})$ and normalizing as a function of $\boldsymbol{\theta}$ (to obtain a function proportional to $\ell_k(\boldsymbol{\theta})$), and finally calculating the mean of the resulting pdf in $\boldsymbol{\theta}$ (which is $\mathbf{t}_k$ according to (71)). Thus, $\mathbf{t}_k$ is related to π_k in a one-to-one manner, and hence it is a local statistic.

Example 2 (Conditionally Independent Agents). In the case of conditionally independent agents, i.e., the observations $\mathbf{y}_k$ are conditionally independent given $\boldsymbol{\theta}$, we can easily calculate the supra-Bayesian posterior. Indeed, the structure of the local likelihood function in (71) shows that we are in the exponential family setting of Example 1. More specifically, we can rewrite (71) as

$$\ell_k(\boldsymbol{\theta}) \propto \exp\left(\boldsymbol{\theta}^\intercal \tilde{\mathbf{t}}_k - \frac{\boldsymbol{\theta}^\intercal \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{H}_k \boldsymbol{\theta}}{2}\right), \tag{74}$$

where

$$\tilde{\mathbf{t}}_k = \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{H}_k \mathbf{t}_k = \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{y}_k$$

is a bijective transformation of $\mathbf{t}_k$ and thus also a valid choice for a local statistic. Considering a Gaussian prior $p(\boldsymbol{\theta})$ with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$, we can rewrite $p(\boldsymbol{\theta})$ as

$$p(\boldsymbol{\theta}) \propto \exp\left(\boldsymbol{\theta}^\intercal \tilde{\mathbf{t}}_0 - \frac{\boldsymbol{\theta}^\intercal \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}}{2}\right), \tag{75}$$

where $\tilde{\mathbf{t}}_0 = \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0$. Comparing (74) with (57) and (75) with (62), we see that $\ell_k(\boldsymbol{\theta}) = p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$ belongs to the exponential family (57) with $\mathbf{t}_k$ formally replaced by $\tilde{\mathbf{t}}_k$ and $A_k(\boldsymbol{\theta}) = \frac{\boldsymbol{\theta}^\intercal \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1}\mathbf{H}_k\boldsymbol{\theta}}{2}$. Furthermore, $p(\boldsymbol{\theta})$ conforms to (62) with $A_0(\boldsymbol{\theta}) = \frac{\boldsymbol{\theta}^\intercal \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}}{2}$. With our assumption of conditionally independent agents, we can use the result (63)–(65) and obtain for the supra-Bayesian fusion result

$$p(\boldsymbol{\theta} \,|\, \mathbf{t}) \propto \exp\left(\boldsymbol{\theta}^\intercal \left(\sum_{k=0}^{K} \tilde{\mathbf{t}}_k\right) - \frac{\boldsymbol{\theta}^\intercal \big(\boldsymbol{\Sigma}_0^{-1} + \sum_{k=1}^{K} \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1}\mathbf{H}_k\big)\boldsymbol{\theta}}{2}\right). \tag{76}$$

This is again a Gaussian pdf, with mean

$$\boldsymbol{\mu}_1 = \boldsymbol{\Sigma}_1 \sum_{k=0}^{K} \tilde{\mathbf{t}}_k = \boldsymbol{\Sigma}_1 \left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \sum_{k=1}^{K} \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1}\mathbf{y}_k\right)$$

and covariance matrix

$$\boldsymbol{\Sigma}_1 = \left(\boldsymbol{\Sigma}_0^{-1} + \sum_{k=1}^{K} \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1}\mathbf{H}_k\right)^{-1}.$$

It is straightforward to verify that (76) is equal to the oracle posterior $p(\boldsymbol{\theta} \,|\, \mathbf{y})$. Thus, we see once again (cf. Section VIII-C) that although the supra-Bayesian fusion result depends on the observations $\mathbf{y}_k$ only via the local statistics $\tilde{\mathbf{t}}_k$, it still equals the oracle posterior $p(\boldsymbol{\theta} \,|\, \mathbf{y})$, as if the fusion center had access to all observations $\mathbf{y}_k$ directly. As we will see below, this crucially depends on our assumption of conditionally independent agents and is no longer true if we assume conditional dependencies between the observations.

B. Global Likelihood Function

In the previous subsection, for the general linear Gaussian model with conditionally dependent $\mathbf{y}_k$, we identified local statistics $\mathbf{t}_k = T_k(\mathbf{y}_k) = \mathbf{V}_k\mathbf{y}_k$ that are related in a one-to-one manner to the local posteriors π_k. The next step according to our three-step program from Section VIII-D is to calculate the global likelihood function $\lambda(\boldsymbol{\theta}) = p(\mathbf{t} \,|\, \boldsymbol{\theta})$ by transforming the conditional pdf $p(\mathbf{y} \,|\, \boldsymbol{\theta})$ into the conditional pdf $p(\mathbf{t} \,|\, \boldsymbol{\theta})$. According to (67), the conditional pdf of $\mathbf{y}$ given $\boldsymbol{\theta}$ is[8]

$$p(\mathbf{y} \,|\, \boldsymbol{\theta}) \propto \exp\left(-\frac{(\mathbf{y} - \mathbf{H}\boldsymbol{\theta})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\theta})}{2}\right). \tag{77}$$

We further have that

$$\mathbf{t} = [\mathbf{t}_1^\intercal, \ldots, \mathbf{t}_K^\intercal]^\intercal = \mathbf{V}\mathbf{y}\,, \tag{78}$$

where $\mathbf{V} = \mathrm{diag}(\mathbf{V}_1, \ldots, \mathbf{V}_K)$ denotes the block-diagonal matrix with block entries $\mathbf{V}_k$ on the diagonal. Thus, $\mathbf{t}$ is a linear function of $\mathbf{y}$ and hence $\mathbf{t}$ given $\boldsymbol{\theta}$ is Gaussian and has mean $\mathbf{V}\mathbf{H}\boldsymbol{\theta}$ and covariance matrix

$$\widetilde{\boldsymbol{\Sigma}} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^\intercal. \tag{79}$$

We assume that $\widetilde{\boldsymbol{\Sigma}}$ is nonsingular. The mean can be simplified to

$$\mathbf{V}\mathbf{H}\boldsymbol{\theta} = \begin{pmatrix} \mathbf{V}_1\mathbf{H}_1 \\ \mathbf{V}_2\mathbf{H}_2 \\ \vdots \\ \mathbf{V}_K\mathbf{H}_K \end{pmatrix} \boldsymbol{\theta} = \begin{pmatrix} \mathbf{I}_{d_\theta} \\ \mathbf{I}_{d_\theta} \\ \vdots \\ \mathbf{I}_{d_\theta} \end{pmatrix} \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\theta} \\ \vdots \\ \boldsymbol{\theta} \end{pmatrix} = \mathbf{1}_K \otimes \boldsymbol{\theta}\,,$$

where we used (68) and the fact that, by (72),

$$\mathbf{V}_k\mathbf{H}_k = \mathbf{I}_{d_\theta}\,. \tag{80}$$

The global likelihood function $\lambda(\boldsymbol{\theta})$ is thus obtained as

$$\begin{aligned} \lambda(\boldsymbol{\theta}) &= p(\mathbf{t} \,|\, \boldsymbol{\theta}) \\ &= \mathcal{N}(\mathbf{t}; \mathbf{1}_K \otimes \boldsymbol{\theta}, \widetilde{\boldsymbol{\Sigma}}) \\ &\propto \exp\left(-\frac{(\mathbf{t} - \mathbf{1}_K \otimes \boldsymbol{\theta})^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{t} - \mathbf{1}_K \otimes \boldsymbol{\theta})}{2}\right). \end{aligned} \tag{81}$$

To summarize, for the linear Gaussian model, local statistics $\mathbf{t}_k$ characterizing the local posteriors π_k are given by (73), and the corresponding global likelihood function $\lambda(\boldsymbol{\theta}) = p(\mathbf{t} \,|\, \boldsymbol{\theta})$ is given by (81).

C. Supra-Bayesian Fusion Rule for a Scalar θ

After identifying local statistics $\mathbf{t}_k$ and calculating the global likelihood function $\lambda(\boldsymbol{\theta}) = p(\mathbf{t} \,|\, \boldsymbol{\theta})$, the final step in the derivation of the supra-Bayesian fusion rule is to calculate the supra-Bayesian posterior $p(\boldsymbol{\theta} \,|\, \mathbf{t})$ according to (52). We first develop the supra-Bayesian fusion rule for the case that $d_\theta = 1$, i.e., for a scalar random variable $\theta \in \mathbb{R}$. Here, the observation matrix $\mathbf{H}$ reduces to a vector $\mathbf{h} \in \mathbb{R}^{d_y}$ and the observation model (66) is given by

$$\mathbf{y} = \mathbf{h}\,\theta + \mathbf{n}\,.$$

Similarly, the local observation at agent k is given as $\mathbf{y}_k = \mathbf{h}_k\theta + \mathbf{n}_k$ with $\mathbf{h}_k \in \mathbb{R}^{d_{y_k}}$, and the local statistic at agent k follows from (73) as

$$t_k = \mathbf{v}_k^\intercal \mathbf{y}_k \in \mathbb{R}\,, \tag{82}$$

where $\mathbf{V}_k$ reduces to the (row) vector

$$\mathbf{v}_k^\intercal = \frac{1}{\mathbf{h}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1}\mathbf{h}_k}\mathbf{h}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1}\,. \tag{83}$$

Note that $\mathbf{V} = \mathrm{diag}(\mathbf{v}_1^\intercal, \ldots, \mathbf{v}_K^\intercal)$ is still a matrix. In this case, we can give the following explicit fusion rule, which is derived in Appendix K-A.

Theorem 21. *For $d_\theta = 1$, let $\ell_k(\theta) = p(\mathbf{y}_k \,|\, \theta)$ denote the local observation likelihood functions given by* (70) *for $k = 1, \ldots, K$ and let $\lambda(\theta) = p(\mathbf{t} \,|\, \theta)$ be the global likelihood function given by* (81)*. Then*

$$\lambda(\theta) \propto \prod_{k=1}^{K} (\ell_k(\theta))^{w_k}, \tag{84}$$

[8] This conditional pdf only exists if the covariance matrix $\boldsymbol{\Sigma}$ is positive definite. However, the derivations that follow do not require the existence of a pdf and are also valid if $\boldsymbol{\Sigma}$ is positive semidefinite.

where

$$w_k = \frac{\mathbf{1}_K^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{e}_k}{\mathbf{h}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1}\mathbf{h}_k}, \tag{85}$$

with $\widetilde{\boldsymbol{\Sigma}} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^\intercal$ *and* $\mathbf{e}_k$ *denoting the kth unit vector in* $\mathbb{R}^K$*. Furthermore, for a given prior* $p(\theta)$ *and local posteriors* $\pi_k(\theta) = p(\theta \,|\, \mathbf{y}_k) \propto p(\theta)\ell_k(\theta)$*, the supra-Bayesian fusion result* $g[\pi_1, \dots, \pi_K](\theta) = p(\theta \,|\, \mathbf{t}) \propto p(\theta)\lambda(\theta)$ *is given by*

$$g[\pi_1, \dots, \pi_K](\theta) \propto (p(\theta))^{1-\sum_{k=1}^K w_k} \prod_{k=1}^K (\pi_k(\theta))^{w_k}. \tag{86}$$

We emphasize that in this theorem we do not assume that the observations $\mathbf{y}_k$ are conditionally independent given θ. Furthermore, it should be noted that the weights w_k in (85) do not generally sum to one, and they may be negative. Thus, the fusion rule (86) is an instance of the generalized multiplicative pooling function in (11).

Finally, if the prior $p(\theta)$ is Gaussian, we can show that the supra-Bayesian fusion result $p(\theta \,|\, \mathbf{t})$ is again Gaussian and reduce the fusion rule (86) to a second-order rule involving only the mean and variance:

Corollary 22. *Under the assumptions of Theorem 21, let the prior* $p(\theta)$ *be Gaussian with mean* μ_0 *and variance* σ_0^2*, i.e.,* $p(\theta) = \mathcal{N}(\theta; \mu_0, \sigma_0^2)$*. Then the supra-Bayesian fusion result* $p(\theta \,|\, \mathbf{t})$ *is again Gaussian, i.e.,* $p(\theta \,|\, \mathbf{t}) = \mathcal{N}(\theta; \mu_1, \sigma_1^2)$*, with mean*

$$\mu_1 = \frac{\widehat{\sigma}^2\sigma_0^2}{\widehat{\sigma}^2 + \sigma_0^2}\mathbf{1}_K^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{t} + \frac{\widehat{\sigma}^2}{\widehat{\sigma}^2 + \sigma_0^2}\mu_0 \tag{87}$$

and variance

$$\sigma_1^2 = \frac{\widehat{\sigma}^2\sigma_0^2}{\widehat{\sigma}^2 + \sigma_0^2},$$

where

$$\widehat{\sigma}^2 = \frac{1}{\mathbf{1}_K^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_K} \tag{88}$$

and $\mathbf{t} = [t_1^\intercal, \dots, t_K^\intercal]^\intercal$ *is given by* (82) *and* (83)*.*

As mentioned before, the supra-Bayesian fusion result $p(\theta \,|\, \mathbf{t})$ is in general different from the oracle posterior $p(\theta \,|\, \mathbf{y})$. Indeed, the oracle posterior is proportional to the product of the prior $p(\theta)$ and the global observation likelihood function $p(\mathbf{y} \,|\, \theta)$ in (77). It can then easily be seen that the oracle posterior $p(\theta \,|\, \mathbf{y})$ is also Gaussian but with mean

$$\mu_2 = \frac{\widehat{\sigma}_2^2\sigma_0^2}{\widehat{\sigma}^2 + \sigma_0^2}\mathbf{h}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{y} + \frac{\widehat{\sigma}_2^2}{\widehat{\sigma}_2^2 + \sigma_0^2}\mu_0 \tag{89}$$

and variance

$$\sigma_2^2 = \frac{\widehat{\sigma}_2^2\sigma_0^2}{\widehat{\sigma}_2^2 + \sigma_0^2},$$

where

$$\widehat{\sigma}_2^2 = \frac{1}{\mathbf{h}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{h}}. \tag{90}$$

To better understand the difference, we note that in (87)

$$\mathbf{1}_K^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{t} = \mathbf{h}^\intercal \mathbf{V}^\intercal (\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^\intercal)^{-1}\mathbf{V}\mathbf{y}$$

and in (88)

$$\mathbf{1}_K^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_K = \mathbf{h}^\intercal \mathbf{V}^\intercal (\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^\intercal)^{-1}\mathbf{V}\mathbf{h},$$

where we used (78)–(80). Comparing with $\mathbf{h}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{y}$ and $\mathbf{h}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{h}$ arising in (89) and (90), respectively, we conclude that the difference between the oracle posterior and the supra-Bayesian posterior is that the matrix $\boldsymbol{\Sigma}^{-1}$ is replaced by $\mathbf{V}^\intercal(\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^\intercal)^{-1}\mathbf{V}$.

A simplified version of Theorem 21 has been shown in [91] and is the setting of the early supra-Bayesian approaches. More specifically, it is assumed in [91] that a fusion center obtains from K agents estimates μ_k of a scalar random variable θ. These estimates can be interpreted as our local statistics t_k. Furthermore, the fusion center has a Gaussian prior for θ and knows that the vector of the estimation errors of all agents, $\mathbf{u} = [u_1, \dots, u_K]^\intercal$ with $u_k = \mu_k - \theta$, also follows a Gaussian distribution with zero mean and some covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ (in general, the errors may be correlated). Equivalently, conditionally on θ, the estimates $\boldsymbol{\mu} = [\mu_1, \dots, \mu_K]^\intercal$ follow a Gaussian distribution with mean $\mathbf{1}_K\theta$ and the same covariance matrix $\widetilde{\boldsymbol{\Sigma}}$. Thus, the setting in [91] directly assumes the conditional distribution of $\mathbf{t}$ given θ without starting from any detailed observation model.

To get a better intuition about the role of the weights w_k and the meaning of negative weights in the setting of Theorem 21, we will consider a specific example.

Example 3 (Private and Shared Observations). We assume that agent k has r_k private observations, i.e., observations that no other agent observes, and r_0 shared observations, i.e., observations that all agents know jointly. The resulting total number of observations is thus $d_y = \sum_{k=1}^K (r_0 + r_k)$. However, there are only $r_0 + \sum_{k=1}^K r_k$ different observations. We assume that these different observations given θ are independent and have variance one and mean θ. To embed this scenario into our linear model, we choose $\mathbf{h}_k = \mathbf{1}_{r_0+r_k}$ and the submatrices of the covariance matrix $\boldsymbol{\Sigma}$ in (69) as

$$\boldsymbol{\Sigma}_{kk'} = \begin{pmatrix} \mathbf{I}_{r_0} & \mathbf{0}_{r_0 \times r_{k'}} \\ \mathbf{0}_{r_k \times r_0} & \mathbf{0}_{r_k \times r_{k'}} \end{pmatrix} \in \mathbb{R}^{(r_0+r_k)\times(r_0+r_{k'})} \tag{91}$$

for $k \neq k'$ and

$$\boldsymbol{\Sigma}_{kk} = \mathbf{I}_{r_0+r_k}.$$

Thus, we have that

$$\mathbf{y}_k = \mathbf{1}_{r_0+r_k}\theta + \mathbf{n}_k,$$

where $\mathbf{n}_k$ is a vector of independent and identically distributed standard Gaussian random variables, i.e., $p(\mathbf{n}_k) = \mathcal{N}(\mathbf{n}_k; \mathbf{0}_{(r_0+r_k)\times 1}, \mathbf{I}_{r_0+r_k})$. The covariance structure (91) between the $\mathbf{n}_k$, for $k \in \{1, \dots, K\}$, implies that for $i \in \{1, \dots, r_0\}$ the ith entry of $\mathbf{n}_k$ and the ith entry of $\mathbf{n}_{k'}$ with $k' \neq k$ coincide with probability one:

$$\mathbb{E}[(n_{k,i} - n_{k',i})^2] = \underbrace{\mathbb{E}[n_{k,i}^2]}_{=1} + \underbrace{\mathbb{E}[n_{k',i}^2]}_{=1} - 2\underbrace{\mathbb{E}[n_{k,i}n_{k',i}]}_{=1} = 0.$$

Thus, the first r_0 observations are the same for all agents.

With these choices and assuming that $r_k > 0$ and $r_0 > 0$, a tedious but straightforward calculation (for details see Appendix K-B) shows that the weights w_k in (85) simplify to

$$w_k = 1 - \frac{K-1}{r_k}\left(\sum_{k'=0}^K \frac{1}{r_{k'}}\right)^{-1}. \tag{92}$$

In particular, we see that all weights are upper-bounded by 1 and are emphasized according to their amount of independent information as given by r_k. More surprising is the possibility of negative weights for agents with few private observations (e.g., the setting $K = 3$, $r_1 = 1$, and $r_0 = r_2 = r_3 = 4$ gives $w_1 = -1/7$). An explanation for this result is that negatively weighting agents with few private observations can counteract the multiple-counting of the shared observations that are part of all agents' posteriors. More generally, it follows from (92) that $w_k \geq 0$ if and only if

$$r_k \geq (K-1)\left(\sum_{k'=0}^{K} \frac{1}{r_{k'}}\right)^{-1}$$

or, equivalently,

$$\sum_{k'=0}^{K} \frac{r_k}{r_{k'}} \geq K-1.$$

The sum of all weights is given by

$$\sum_{k=1}^{K} w_k = K - (K-1)\left(\sum_{k'=0}^{K} \frac{1}{r_{k'}}\right)^{-1} \sum_{k=1}^{K} \frac{1}{r_k}. \tag{93}$$

From this expression, we readily conclude that

$$1 \leq \sum_{k=1}^{K} w_k \leq K. \tag{94}$$

Indeed, this follows from the fact that the second term on the right-hand side of (93), $(K-1)\big(\sum_{k'=0}^{K} \frac{1}{r_{k'}}\big)^{-1} \sum_{k=1}^{K} \frac{1}{r_k}$, is nonnegative and upper-bounded by $K-1$ since $\big(\sum_{k'=0}^{K} \frac{1}{r_{k'}}\big)^{-1} \sum_{k=1}^{K} \frac{1}{r_k} \leq 1$. The double bound (94) shows that although some weights may be negative, the sum of all weights is always between the sum of all weights in the log-linear pooling function in (4) (there, the sum was 1) and the sum of all weights in the multiplicative pooling function in (10) (there, all weights were 1, and hence the sum was K).

Another conclusion we can draw is that varying the number of shared observations r_0—while keeping the number of private observations r_k fixed—corresponds to an "interpolation" between the multiplicative pooling function and the log-linear pooling function. Consider first the case that the agents have the same number of private observations, i.e., $r_1 = \cdots = r_K$. When $r_0 = 0$, a derivation similar to that in Appendix K-B gives $w_k = 1$. This implies that when the agents do not share any observations, the pooling function in (86) corresponds exactly to the standard multiplicative pooling function in (10). On the other hand, as the number of shared observations r_0 increases, the pooling function behaves closer to a symmetric log-linear pooling function (i.e., using $w_k = 1/K$). Indeed, it follows from (92) that

$$\lim_{r_0 \to \infty} w_k = \frac{1}{K}.$$

If we remove the restriction that $r_1 = \cdots = r_K$, the connection to multiplicative pooling still holds; however, the connection to log-linear pooling only holds under the condition of nonnegative weights, i.e., $w_k \geq 0$ for all k, which may be violated if some agents hold only few private observations as compared to the total number of observations.

D. Supra-Bayesian Fusion Rule for a Vector $\boldsymbol{\theta}$

We can generalize Theorem 21 to a vector $\boldsymbol{\theta} \in \mathbb{R}^{d_\theta}$ with $d_\theta > 1$. However, formally, the weights w_k in (85) become matrices $\mathbf{W}_k$ and thus cannot be used as powers in a fusion rule. Hence, the following fusion result is more complicated and the relation to the one-dimensional case is not obvious. A proof is provided in Appendix L-A.

Theorem 23. *Let $\ell_k(\boldsymbol{\theta}) = p(\mathbf{y}_k \,|\, \boldsymbol{\theta})$ denote the local observation likelihood functions given by* (70) *for $k = 1, \ldots, K$ and let $\lambda(\boldsymbol{\theta}) = p(\mathbf{t} \,|\, \boldsymbol{\theta})$ be the global likelihood function given by* (81)*. Then*

$$\lambda(\boldsymbol{\theta}) \propto \xi_0(\boldsymbol{\theta}) \prod_{k=1}^{K} \ell_k(\mathbf{W}_k \boldsymbol{\theta}) \tag{95}$$

where

$$\mathbf{W}_k = (\mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{H}_k)^{-1} (\mathbf{e}_k \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1} (\mathbf{1}_K \otimes \mathbf{I}_{d_\theta}), \tag{96}$$

with $\mathbf{e}_k$ denoting the kth unit vector in $\mathbb{R}^K$ and $\widetilde{\boldsymbol{\Sigma}} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^\intercal$, and

$$\xi_0(\boldsymbol{\theta}) = \exp\left(-\frac{\boldsymbol{\theta}^\intercal \mathbf{G} \boldsymbol{\theta}}{2}\right). \tag{97}$$

Here,

$$\mathbf{G} = \widehat{\boldsymbol{\Sigma}}^{-1} - \sum_{k=1}^{K} \mathbf{W}_k^\intercal \mathbf{H}_k^\intercal \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{H}_k \mathbf{W}_k \tag{98}$$

with

$$\widehat{\boldsymbol{\Sigma}}^{-1} = (\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1} (\mathbf{1}_K \otimes \mathbf{I}_{d_\theta}). \tag{99}$$

Furthermore, for a given prior $p(\boldsymbol{\theta})$ and local posteriors $\pi_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \,|\, \mathbf{y}_k) \propto p(\boldsymbol{\theta})\ell_k(\boldsymbol{\theta})$, the supra-Bayesian fusion result $g[\pi_1, \ldots, \pi_K](\boldsymbol{\theta}) = p(\boldsymbol{\theta} \,|\, \mathbf{t}) \propto p(\boldsymbol{\theta})\lambda(\boldsymbol{\theta})$ is given by

$$g[\pi_1, \ldots, \pi_K](\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}) \xi_0(\boldsymbol{\theta}) \prod_{k=1}^{K} \frac{\pi_k(\mathbf{W}_k \boldsymbol{\theta})}{p(\mathbf{W}_k \boldsymbol{\theta})}. \tag{100}$$

Finally, if the prior $p(\boldsymbol{\theta})$ is Gaussian, then the supra-Bayesian fusion result $p(\boldsymbol{\theta} \,|\, \mathbf{t})$ is again Gaussian and the fusion rule (100) can be reduced to a second-order rule involving only the mean and covariance matrix:

Corollary 24. *Under the assumptions of Theorem 23, let the prior $p(\boldsymbol{\theta})$ be Gaussian with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$, i.e., $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. Then the supra-Bayesian fusion result $p(\boldsymbol{\theta} \,|\, \mathbf{t})$ is again Gaussian, i.e., $p(\boldsymbol{\theta} \,|\, \mathbf{t}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, with mean*

$$\boldsymbol{\mu}_1 = \big(\widehat{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}_0^{-1}\big)^{-1} \big((\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1} \mathbf{t} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0\big) \tag{101}$$

and covariance matrix

$$\boldsymbol{\Sigma}_1 = \big(\widehat{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}_0^{-1}\big)^{-1}. \tag{102}$$

Here, we recall that $\mathbf{t} = [\mathbf{t}_1^\intercal, \ldots, \mathbf{t}_K^\intercal]^\intercal$ with $\mathbf{t}_k$ given by (73). A proof of Corollary 24 is provided in Appendix L-B.

The supra-Bayesian fusion result in (100) has an intriguing structure in that the agent pdfs are first preprocessed by a multiplication in the argument and then combined via a generalized multiplicative pooling function. The relevance of this fusion rule beyond the linear Gaussian setting, especially for approximately linear Gaussian observation models, is an open issue.

As in the scalar case, the supra-Bayesian fusion result $p(\boldsymbol{\theta}\,|\,\mathbf{t})$ is in general different from the oracle posterior $p(\boldsymbol{\theta}\,|\,\mathbf{y})$. Again, the oracle posterior is proportional to the product of the prior $p(\boldsymbol{\theta})$ and the global observation likelihood function $p(\mathbf{y}\,|\,\boldsymbol{\theta})$ in (77); it is easily seen that $p(\boldsymbol{\theta}\,|\,\mathbf{y})$ is also Gaussian but with mean

$$\boldsymbol{\mu}_2 = \big(\widehat{\boldsymbol{\Sigma}}_2^{-1} + \boldsymbol{\Sigma}_0^{-1}\big)^{-1}\big(\mathbf{H}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\big) \tag{103}$$

and covariance matrix

$$\boldsymbol{\Sigma}_2 = \big(\widehat{\boldsymbol{\Sigma}}_2^{-1} + \boldsymbol{\Sigma}_0^{-1}\big)^{-1},$$

where

$$\widehat{\boldsymbol{\Sigma}}_2^{-1} = \mathbf{H}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{H}\,. \tag{104}$$

The difference can be better understood by noting that in (101)

$$(\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^{\mathsf{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{t} = \mathbf{H}^{\mathsf{T}}\mathbf{V}^{\mathsf{T}}(\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}})^{-1}\mathbf{V}\mathbf{y}$$

and in (99)

$$(\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^{\mathsf{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{1}_K \otimes \mathbf{I}_{d_\theta}) = \mathbf{H}^{\mathsf{T}}\mathbf{V}^{\mathsf{T}}(\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}})^{-1}\mathbf{V}\mathbf{H}\,,$$

where we used (78)–(80). Comparing with $\mathbf{H}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}$ and $\mathbf{H}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{H}$ in (103) and (104), respectively, we conclude that, as in the scalar case considered earlier, the difference between the oracle posterior and the supra-Bayesian posterior is that $\boldsymbol{\Sigma}^{-1}$ is replaced by $\mathbf{V}^{\mathsf{T}}(\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}})^{-1}\mathbf{V}$.

X. Outlook

The fusion of pdfs presents numerous interesting aspects beyond those considered in our treatment. Moreover, certain extensions can be envisioned. In what follows, we suggest some related directions of future research.

- Our discussion of pdf fusion emphasized theoretical considerations. In practical implementations, a finite-dimensional representation or parametrization of the agent pdfs $q_k(\boldsymbol{\theta})$ is required. Popular examples are Gaussian, Gaussian mixture, and particle representations [31], [68], [69]. Since these representations are usually approximations of the true pdfs, a relevant issue is the tradeoff between low representation complexity (small number of parameters) and high accuracy of approximation. Furthermore, algorithms implementing a given pooling function for a given type of parametric representation are required. Examples of finite-dimensional parametric fusion rules were considered in Sections VI and IX.
- In the case of a centralized agent network where each agent pdf $q_k(\boldsymbol{\theta})$ is transmitted to the fusion center via a channel, communication cost is another practical issue. Although a low-dimensional parametric representation of the agent pdfs may be used to achieve a low communication cost, the reduction of communication cost is ultimately a source coding (rate-distortion) problem.
- In many cases, the aggregate pdf $q(\boldsymbol{\theta}) = g[q_1, \ldots, q_K](\boldsymbol{\theta})$ is not used as the final result but arises as part of a method performing a statistical inference task such as estimation, detection, classification, or clustering. In this setting, the pooling function (or certain parameters within a given family of pooling functions) should be chosen or optimized such that the performance of the statistical inference method is maximized. Note that this is different from the optimization approach considered in Section V.
- Our discussion assumed the existence of a fusion center that has access to all pdfs $q_k(\boldsymbol{\theta})$. In a decentralized agent network, there is no fusion center and each agent is able to communicate only with certain neighboring agents. Besides the basic necessity of using a distributed communication-and-fusion protocol, challenging aspects in the decentralized setting include communication cost, efficient representation of pdfs, and double counting of information along cycles in the network graph.
- In many scenarios, the agent pdfs $q_k(\boldsymbol{\theta})$ are time-varying and a temporal sequence $q_k^{(n)}(\boldsymbol{\theta})$, where $n = 1, 2, \ldots$ is a discrete time index, is available at the kth agent. This serial setting suggests a sequential variant of pdf fusion in which at each time n the fused pdf is not calculated from scratch but the previous fusion result is updated using the new set of $q_k^{(n)}(\boldsymbol{\theta})$. Practical implementations of sequential updating can be based on both parametric and nonparametric representations of the pdfs.
- The fusion of multiobject pdfs or probability hypothesis densities of finite point processes (random finite sets), especially in the context of multitarget tracking, is a topic of active research [2], [8], [44], [110], [111]. While the current focus is on the finite point process counterparts of the linear and log-linear pooling functions, it would also be interesting to investigate the applicability of the other pooling functions considered in Sections III and V. In particular, the fact that the family of Hölder pooling functions offers fusion characteristics that are intermediate between those of the linear and log-linear pooling functions may be relevant to multitarget tracking. Furthermore, it may be rewarding to reformulate and develop our results on supra-Bayesian pdf fusion in the context of finite point processes.
- Big data problems allow a natural application of pdf fusion. When the data to be processed are so large in size that they exceed the capacity of a single computer, it is logical to partition them and process the different parts separately. Furthermore, data related to some quantity of interest may be available in heterogeneous form, so that all of the data cannot be processed within a single framework and hence different parts have to be processed separately. In either case, the individual processing results can be represented as summaries, which then need to be fused into one overall summary. The concepts and techniques presented in this article provide suggestions regarding the construction and fusion of the summaries. This is of particular interest in the context of modern machine learning methods [65]–[67], [136], [193].
- Ensemble learning [128], [130], i.e., the combination of the results of multiple learning algorithms, is currently one of the most successful learning paradigms. At the

same time, there is a growing demand for probabilistic machine learning methods that provide along with a point estimate also a measure of reliability. Until now, only few works have considered ensembles of probabilistic machine learning methods. We conjecture that the success of the ensemble learning paradigm will soon lead to its increased use also in probabilistic machine learning. At that point, it is likely that probabilistic opinion pooling will outperform the simple linear voting rules that are currently used to combine point estimates.

- With a collaborative machine learning methodology known as federated learning, a learning algorithm is trained across multiple decentralized edge devices or servers that hold local data, which are not exchanged [134]. In other words, model parameters are learned collectively by many interconnected devices without sharing or disclosing local training data. The devices send summaries instead of raw data to a server for fusion. Here, again, fusion plays a central role. The fusion process can be challenging in the case of a large number of heterogeneous devices with different constraints. Using pdfs to represent the local summaries enables the use of different pdf representations at the individual devices, from simple parametric models to complex kernel density estimates, which can still be combined in a meaningful way. Moreover, different levels of quality of the local data can be taken into account by using appropriate weights in the pooling function used for pdf fusion.
- A potential theoretical basis of pdf fusion that has not been explored in this work is information geometry, which studies probability theory and statistics using tools from differential geometry [42]. The focus of information geometry is on statistical manifolds whose points correspond to probability distributions. This theoretical framework can be exploited for fusion by assuming that local estimates are posterior pdfs that correspond to a parametric family with the structure of a Riemannian manifold [42]. One can then formulate pdf fusion, e.g., by considering the fused pdf to be an informative barycenter of the manifold [194].
- Within the finite-dimensional supra-Bayesian setting, an explicit fusion rule was obtained only for linear Gaussian observation models (see Section IX). This fusion rule can formally be used also for nonlinear/non-Gaussian models with known first and second moments. However, it is here unclear how close the obtained fusion result will be to the true supra-Bayesian fusion result. A characterization of the error for approximately linear Gaussian observation models is an interesting topic for future research. Another interesting topic is the derivation of explicit supra-Bayesian fusion rules for simple nonlinear/non-Gaussian observation models.
- Our supra-Bayesian framework is currently limited to a finite-dimensional setting. Although this is the setting most frequently encountered in practical applications, it would be interesting to find a definition of a likelihood function for random pdfs that do not admit a finite-dimensional parameterization. For this, nonparametric Bayesian models [195] appear to be a feasible starting point. The challenge is to model a useful and nontrivial dependence on the parameter $\boldsymbol{\theta}$ that accounts for the constraint that random pdfs must be nonnegative and integrate to one with probability one.

XI. Concluding Remarks

The problem of fusing multiple pdfs $q_k(\boldsymbol{\theta})$, $k = 1, \ldots, K$ of a continuous random vector $\boldsymbol{\theta}$ into an aggregate pdf $q(\boldsymbol{\theta}) = g[q_1, \ldots, q_K](\boldsymbol{\theta})$ has many possible solutions and, indeed, several different approaches to this fusion problem have been developed in the past decades. We have attempted to survey and study these approaches and the related solutions in a structured and coherent manner. Our discussion has emphasized a first basic distinction between the axiomatic approach, the optimization approach, and the conceptually more complex supra-Bayesian framework.

Regarding the axiomatic approach, we formulated a set of axioms and determined the axioms satisfied by each considered pooling function. This analysis demonstrated the prominent role of the linear, log-linear, and multiplicative pooling functions within the axiomatic framework. However, it also revealed that several desirable axioms are effectively incompatible and postulating those simultaneously implies a dictatorship pooling function.

Regarding the optimization approach, besides other results, we proved that the minimization of the weighted sum of α-divergences yields the family of Hölder mean pooling functions. This family contains the two most popular pooling functions—the linear and log-linear pooling functions—as special cases. Moreover, it offers an infinite number of further interesting pooling functions with different multimodality and tail decay characteristics depending on the choice of a single parameter.

The supra-Bayesian framework is different from the classical probabilistic opinion pooling framework in that the pdfs $q_k(\boldsymbol{\theta})$ are modeled as random observations, and additional information regarding the statistical structure of $\boldsymbol{\theta}$ is available to the fusion center. In this framework, the optimal aggregate pdf $q(\boldsymbol{\theta})$ is the global posterior pdf of $\boldsymbol{\theta}$ given the pdfs $q_k(\boldsymbol{\theta})$. Since random functions are difficult to work with, we introduced the finite-dimensional supra-Bayesian model based on random "local statistics." Using this framework, we formulated a general procedure for obtaining the supra-Bayesian posterior pdf conditioned on all the local statistics, and we derived explicit fusion rules for special cases.

While the theory of pdf fusion appears mature, interesting directions of future work are related to implementation and application aspects. We provided some suggestions including implementations using parametric representations, integration into probabilistic methods for multisensor signal processing and machine learning, and extensions to decentralized scenarios and point processes.

Acknowledgments

We would like to thank the anonymous reviewers for insightful comments and constructive criticism, which have resulted in a significant improvement of this article. In particular, we are thankful for a comment on a conceptual relationship between pdf fusion and robust hypothesis testing as mentioned in Section V. We are also grateful to Mr. Thomas Kropfreiter for sharing his expertise in target tracking.

Appendix A
Proof of Theorem 1

A. Axioms Satisfied by the Linear Pooling Function

We first show that all the mentioned axioms are satisfied by the linear pooling function. Let $g[q_1,\dots,q_K](\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta})$ with $(w_1,\dots,w_K) \in \mathcal{S}_K$. We first show the ZPP (A2). Assume that for some event $\mathcal{A}$, we have $Q_k(\mathcal{A}) = 0$ for all $k = 1,\dots,K$. Because $Q_k(\mathcal{A}) = \int_{\mathcal{A}} q_k(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$ and $q_k(\boldsymbol{\theta})$ is nonnegative, this implies $q_k(\boldsymbol{\theta}) = 0$ for almost all $\boldsymbol{\theta} \in \mathcal{A}$ and all $k = 1,\dots,K$. Thus,

$$q(\boldsymbol{\theta}) = g[q_1,\dots,q_K](\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta}) = 0\,,$$

for almost all $\boldsymbol{\theta} \in \mathcal{A}$. Hence, $Q(\mathcal{A}) = \int_{\mathcal{A}} q(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = 0$, which concludes the proof of the ZPP.

We next show unanimity preservation (A3). To this end, assume that $q_k(\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$ for all $k = 1,\dots,K$. Then

$$\begin{aligned} q(\boldsymbol{\theta}) &= g[q_1,\dots,q_K](\boldsymbol{\theta}) \\ &= \sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta}) \\ &= q_0(\boldsymbol{\theta}) \sum_{k=1}^{K} w_k \\ &= q_0(\boldsymbol{\theta})\,, \end{aligned}$$

which shows unanimity preservation.

To show the SSFP (A4), we define $h\colon [0,1]^K \to [0,1]$ as

$$h(p_1,\dots,p_K) \triangleq \sum_{k=1}^{K} w_k p_k\,. \tag{105}$$

For an arbitrary set $\mathcal{A} \subseteq \Theta$ and any opinion profile $(q_1,\dots,q_K)$, we have that

$$\begin{aligned} Q(\mathcal{A}) &= \int_{\mathcal{A}} q(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \\ &= \int_{\mathcal{A}} \sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \\ &= \sum_{k=1}^{K} w_k \int_{\mathcal{A}} q_k(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \\ &= \sum_{k=1}^{K} w_k Q_k(\mathcal{A}) \\ &= h(Q_1(\mathcal{A}),\dots,Q_K(\mathcal{A}))\,, \end{aligned}$$

i.e., h satisfies the condition stated in A4.

The WSFP (A5) follows by setting $h_{\mathcal{A}} = h$ with h given in (105). The likelihood principles (A6 and A7) are obviously satisfied with

$$h(t_1,\dots,t_K) \triangleq \sum_{k=1}^{K} w_k t_k$$

and $h_{\boldsymbol{\theta}} = h$, respectively, as is the symmetry statement.

B. Equivalence Statement

We now prove the other direction, namely, that each of the assumptions (ii)–(iv) stated in Theorem 1 implies that g is a linear pooling function. More specifically, we will show the chain of implications (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). Because we already showed (i) $\Rightarrow$ (iv), this implies that (i)–(iv) are equivalent, and thus concludes the proof.

1) (iv) implies (iii): We assume that g satisfies (iv), i.e., the WSFP (A5) and unanimity preservation (A3). We will show that this implies that g satisfies the ZPP (A2), i.e., (iv) implies (iii). Let $h_{\mathcal{A}}\colon [0,1]^K \to [0,1]$ denote the function satisfying (16) for all opinion profiles. For any set $\mathcal{A}$ that satisfies $|\mathcal{A}^c| > 0$, let us choose $q_k(\boldsymbol{\theta}) = \mathbb{1}_{\mathcal{A}^c}(\boldsymbol{\theta})/|\mathcal{A}^c|$ for all $k = 1,\dots,K$. Then $Q_k(\mathcal{A}) = \int_{\mathcal{A}} q_k(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = 0$ for all k. By unanimity preservation, this implies $Q(\mathcal{A}) = 0$. On the other hand, we have

$$Q(\mathcal{A}) \overset{(16)}{=} h_{\mathcal{A}}(Q_1(\mathcal{A}),\dots,Q_K(\mathcal{A})) = h_{\mathcal{A}}(0,\dots,0)$$

and hence

$$h_{\mathcal{A}}(0,\dots,0) = 0 \tag{106}$$

for any set $\mathcal{A}$ such that $|\mathcal{A}^c| > 0$.

To show the ZPP, assume that for a given opinion profile $(q_1,\dots,q_K)$, we have $Q_k(\mathcal{A}) = 0$ for all $k = 1,\dots,K$. Note that this is only possible if $|\mathcal{A}^c| > 0$ as otherwise $Q_k(\Theta) = Q_k(\mathcal{A}) + Q_k(\mathcal{A}^c) = 0$. Thus, we can calculate $Q(\mathcal{A})$ as

$$Q(\mathcal{A}) \overset{(16)}{=} h_{\mathcal{A}}(Q_1(\mathcal{A}),\dots,Q_K(\mathcal{A})) = h_{\mathcal{A}}(0,\dots,0) \overset{(106)}{=} 0\,,$$

which shows that the ZPP (A2) is satisfied.

2) (iii) implies (ii): Next we show that (iii), i.e., the WSFP (A5) and the ZPP (A2), implies (ii), i.e., the SSFP (A4). Let again $h_{\mathcal{A}}\colon [0,1]^K \to [0,1]$ denote the function satisfying (16) for all opinion profiles. Our proof consists of three steps:

1) Show that for two nontrivial events $\mathcal{A}$ and $\mathcal{B}$ (i.e., $|\mathcal{A}|, |\mathcal{A}^c|, |\mathcal{B}|, |\mathcal{B}^c| > 0$) that have a nontrivial intersection and a nontrivial union, we have $h_{\mathcal{A}} = h_{\mathcal{B}}$.
2) Show that for any nontrivial events $\mathcal{A}$ and $\mathcal{B}$, there exists a nontrivial event $\mathcal{C}$ such that $\mathcal{A} \cap \mathcal{C}$, $\mathcal{A} \cup \mathcal{C}$, $\mathcal{B} \cap \mathcal{C}$, and $\mathcal{B} \cup \mathcal{C}$ are all nontrivial. This implies by step 1 that $h_{\mathcal{A}} = h_{\mathcal{C}}$ and $h_{\mathcal{B}} = h_{\mathcal{C}}$, and thus $h_{\mathcal{A}} = h_{\mathcal{B}}$. Thus, setting $h \triangleq h_{\mathcal{A}}$, we have $h_{\mathcal{A}'} = h$ for all nontrivial events $\mathcal{A}'$, and hence the same function h satisfies (14) for all nontrivial events.
3) Show that the function h satisfies (14) also for trivial events.

To show step 1, we consider two nontrivial events $\mathcal{A}$ and $\mathcal{B}$ that have a nontrivial intersection, in particular, $|\mathcal{A} \cap \mathcal{B}| > 0$, and a nontrivial union, in particular, $|(\mathcal{A} \cup \mathcal{B})^c| > 0$. We fix arbitrary $(p_1,\dots,p_K) \in [0,1]^K$ and will show that

$h_{\mathcal{A}}(p_1, \ldots, p_K) = h_{\mathcal{B}}(p_1, \ldots, p_K)$. Because $|\mathcal{A} \cap \mathcal{B}| > 0$ and $|(\mathcal{A} \cup \mathcal{B})^c| > 0$, there exists an opinion profile $(q_1, \ldots, q_K)$ such that

$$Q_k(\mathcal{A} \cap \mathcal{B}) = p_k \tag{107}$$

and

$$Q_k\big((\mathcal{A} \cup \mathcal{B})^c\big) = 1 - p_k\,, \tag{108}$$

for all $k = 1, \ldots, K$. Because $\Theta = (\mathcal{A} \cup \mathcal{B}) \cup (\mathcal{A} \cup \mathcal{B})^c$ is a disjoint union and $Q_k(\Theta) = 1$, (108) implies $Q_k(\mathcal{A} \cup \mathcal{B}) = p_k$. Hence, as also $Q_k(\mathcal{A} \cap \mathcal{B}) = p_k$ by (107), the difference set $(\mathcal{A} \cup \mathcal{B}) \setminus (\mathcal{A} \cap \mathcal{B})$ satisfies

$$Q_k\big((\mathcal{A} \cup \mathcal{B}) \setminus (\mathcal{A} \cap \mathcal{B})\big) = 0\,. \tag{109}$$

Because $(\mathcal{A} \cup \mathcal{B}) \setminus (\mathcal{A} \cap \mathcal{B}) = (\mathcal{A} \setminus \mathcal{B}) \cup (\mathcal{B} \setminus \mathcal{A})$, this implies $Q_k(\mathcal{A} \setminus \mathcal{B}) = 0$ and $Q_k(\mathcal{B} \setminus \mathcal{A}) = 0$. Thus,

$$\begin{aligned} Q_k(\mathcal{A}) &= Q_k\big((\mathcal{A} \setminus \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B})\big) \\ &= Q_k(\mathcal{A} \setminus \mathcal{B}) + Q_k(\mathcal{A} \cap \mathcal{B}) \\ &= p_k \end{aligned} \tag{110}$$

and, similarly,

$$Q_k(\mathcal{B}) = p_k\,. \tag{111}$$

Furthermore,

$$Q_k\big((\mathcal{A} \setminus \mathcal{B}) \cup (\mathcal{B} \setminus \mathcal{A})\big) = Q_k\big((\mathcal{A} \cup \mathcal{B}) \setminus (\mathcal{A} \cap \mathcal{B})\big) \stackrel{(109)}{=} 0\,, \tag{112}$$

for all $k = 1, \ldots, K$. By the ZPP, (112) implies $Q\big((\mathcal{A} \setminus \mathcal{B}) \cup (\mathcal{B} \setminus \mathcal{A})\big) = 0$ and, in turn, $Q\big((\mathcal{A} \setminus \mathcal{B})\big) = Q\big((\mathcal{B} \setminus \mathcal{A})\big) = 0$. Thus,

$$\begin{aligned} Q(\mathcal{A}) &= Q(\mathcal{A} \setminus \mathcal{B}) + Q(\mathcal{A} \cap \mathcal{B}) \\ &= Q(\mathcal{A} \cap \mathcal{B}) \\ &= Q(\mathcal{B} \setminus \mathcal{A}) + Q(\mathcal{A} \cap \mathcal{B}) \\ &= Q(\mathcal{B})\,. \end{aligned} \tag{113}$$

For the functions $h_{\mathcal{A}}$ and $h_{\mathcal{B}}$, these properties imply

$$\begin{aligned} h_{\mathcal{A}}(p_1, \ldots, p_K) &\stackrel{(110)}{=} h_{\mathcal{A}}(Q_1(\mathcal{A}), \ldots, Q_K(\mathcal{A})) \\ &\stackrel{(16)}{=} Q(\mathcal{A}) \\ &\stackrel{(113)}{=} Q(\mathcal{B}) \\ &\stackrel{(16)}{=} h_{\mathcal{B}}(Q_1(\mathcal{B}), \ldots, Q_K(\mathcal{B})) \\ &\stackrel{(111)}{=} h_{\mathcal{B}}(p_1, \ldots, p_K)\,, \end{aligned}$$

i.e.,

$$h_{\mathcal{A}}(p_1, \ldots, p_K) = h_{\mathcal{B}}(p_1, \ldots, p_K) \tag{114}$$

for any nontrivial events $\mathcal{A}, \mathcal{B} \subseteq \Theta$ that have a nontrivial intersection and a nontrivial union.

To show step 2, we first construct a set $\mathcal{C} \subseteq \mathcal{A} \cup \mathcal{B}$ such that $\mathcal{A} \cap \mathcal{C}$, $\mathcal{B} \cap \mathcal{C}$, $\mathcal{A} \cup \mathcal{C}$, and $\mathcal{B} \cup \mathcal{C}$ are nontrivial. If $|\mathcal{A} \cap \mathcal{B}| > 0$, then $\mathcal{C} = \mathcal{A} \cap \mathcal{B}$ can easily be seen to satisfy these assumptions. If $|\mathcal{A} \cap \mathcal{B}| = 0$, we choose $\mathcal{C} = \mathcal{C}_{\mathcal{A}} \cup \mathcal{C}_{\mathcal{B}}$ where $\mathcal{C}_{\mathcal{A}} \subseteq \mathcal{A}$ with $|\mathcal{C}_{\mathcal{A}}|, |\mathcal{A} \setminus \mathcal{C}_{\mathcal{A}}| > 0$ and $\mathcal{C}_{\mathcal{B}} \subseteq \mathcal{B}$ with $|\mathcal{C}_{\mathcal{B}}|, |\mathcal{B} \setminus \mathcal{C}_{\mathcal{B}}| > 0$. The separations $\mathcal{A} = \mathcal{C}_{\mathcal{A}} \cup (\mathcal{A} \setminus \mathcal{C}_{\mathcal{A}})$ and $\mathcal{B} = \mathcal{C}_{\mathcal{B}} \cup (\mathcal{B} \setminus \mathcal{C}_{\mathcal{B}})$ are possible because the Lebesgue measure is nonatomic, i.e., any set of positive Lebesgue measure can be separated into two disjoint sets of positive Lebesgue measure.

We now choose

$$h(p_1, \ldots, p_K) = h_{\mathcal{A}}(p_1, \ldots, p_K) \tag{115}$$

for any nontrivial set $\mathcal{A}$. Then, for any nontrivial set $\mathcal{B} \subseteq \Theta$, we construct $\mathcal{C}$ as above and obtain

$$\begin{aligned} h(Q_1(\mathcal{B}), \ldots, Q_K(\mathcal{B})) &\stackrel{(115)}{=} h_{\mathcal{A}}(Q_1(\mathcal{B}), \ldots, Q_K(\mathcal{B})) \\ &\stackrel{(114)}{=} h_{\mathcal{C}}(Q_1(\mathcal{B}), \ldots, Q_K(\mathcal{B})) \\ &\stackrel{(114)}{=} h_{\mathcal{B}}(Q_1(\mathcal{B}), \ldots, Q_K(\mathcal{B})) \\ &\stackrel{(16)}{=} Q(\mathcal{B})\,, \end{aligned} \tag{116}$$

i.e., (14) is satisfied for any nontrivial set $\mathcal{B}$.

It remains to show step 3, i.e., that with this choice of h, (14) is also satisfied by trivial sets. For trivial sets $\mathcal{A}$, i.e., such that $|\mathcal{A}| = 0$ or $|\mathcal{A}^c| = 0$, we have $Q_k(\mathcal{A}) = 0$ or $Q_k(\mathcal{A}) = 1$ for all $k = 1, \ldots, K$, respectively. Also the fused result must satisfy $Q(\mathcal{A}) = 0$ or $Q(\mathcal{A}) = 1$, respectively. Thus, we have to show $h(0, \ldots, 0) = 0$ and $h(1, \ldots, 1) = 1$ for our choice of h in (115). To this end, let $\mathcal{B} \subseteq \Theta$ be any nontrivial set and choose an opinion profile $(q_1, \ldots, q_K)$ such that $Q_k(\mathcal{B}) = 0$ for all $k = 1, \ldots, K$. Then the ZPP implies $Q(\mathcal{B}) = 0$. On the other hand, since $\mathcal{B}$ is a nontrivial set and thus (116) is satisfied, we have

$$h(0, \ldots, 0) = h(Q_1(\mathcal{B}), \ldots, Q_K(\mathcal{B})) \stackrel{(116)}{=} Q(\mathcal{B})\,.$$

Thus, $h(0, \ldots, 0) = 0$. Furthermore, $Q_k(\mathcal{B}) = 0$ and $Q(\mathcal{B}) = 0$ imply $Q_k(\mathcal{B}^c) = 1$ and $Q(\mathcal{B}^c) = 1$, respectively. Hence,

$$h(1, \ldots, 1) = h(Q_1(\mathcal{B}^c), \ldots, Q_K(\mathcal{B}^c)) \stackrel{(116)}{=} Q(\mathcal{B}^c)\,.$$

Thus, $h(1, \ldots, 1) = 1$. Hence, we identified a function h such that (14) holds for all sets $\mathcal{A} \subseteq \Theta$. This concludes the proof that (iii) implies (ii).

3) (ii) implies (i): Finally, we show that (ii), i.e., the SSFP (A4) implies (i), i.e., that g is a linear pooling function. Let h denote the function satisfying (14). Furthermore, let $\mathcal{A}, \mathcal{B}, \mathcal{C} \subseteq \Theta$ be disjoint events of positive Lebesgue measure. For arbitrary $p_1, \tilde{p}_1, \ldots, p_K, \tilde{p}_K \in [0, 1]$ satisfying $p_k + \tilde{p}_k \leq 1$ for all $k = 1, \ldots, K$, we define an opinion profile $(q_1, \ldots, q_K)$ such that $Q_k(\mathcal{A}) = p_k$, $Q_k(\mathcal{B}) = \tilde{p}_k$, and $Q_k(\mathcal{C}) = 1 - p_k - \tilde{p}_k$. Because $\mathcal{A}$ and $\mathcal{B}$ are disjoint, $Q_k(\mathcal{A} \cup \mathcal{B}) = p_k + \tilde{p}_k$ and $Q(\mathcal{A} \cup \mathcal{B}) = Q(\mathcal{A}) + Q(\mathcal{B})$. Thus,

$$\begin{aligned} &h(p_1 + \tilde{p}_1, \ldots, p_K + \tilde{p}_K) \\ &\quad = h(Q_1(\mathcal{A} \cup \mathcal{B}), \ldots, Q_K(\mathcal{A} \cup \mathcal{B})) \\ &\quad \stackrel{(14)}{=} Q(\mathcal{A} \cup \mathcal{B}) \\ &\quad = Q(\mathcal{A}) + Q(\mathcal{B}) \\ &\quad \stackrel{(14)}{=} h(Q_1(\mathcal{A}), \ldots, Q_K(\mathcal{A})) + h(Q_1(\mathcal{B}), \ldots, Q_K(\mathcal{B})) \\ &\quad = h(p_1, \ldots, p_K) + h(\tilde{p}_1, \ldots, \tilde{p}_K)\,, \end{aligned}$$

i.e., h is an additive function on its domain $[0, 1]^K$. It can moreover be extended to an additive function on $\mathbb{R}^K$. Because h is also bounded by 1 on $[0, 1]^K$, it must be linear according to [196, Th. 1, p. 215], i.e.,

$$h(p_1, \ldots, p_K) = \sum_{k=1}^{K} w_k p_k\,. \tag{117}$$

Here, the weights w_k must be in $[0,1]$ because $h(p_1,\ldots,p_K) \in [0,1]$ for all $(p_1,\ldots,p_K) \in [0,1]^K$. Furthermore, because $1 = Q(\Theta) = h(Q_1(\Theta),\ldots,Q_K(\Theta)) = h(1,\ldots,1) = \sum_{k=1}^K w_k$, the weights must sum to one. We thus have for any event $\mathcal{A} \subseteq \Theta$

$$\begin{aligned}\int_{\mathcal{A}} q(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} &= Q(\mathcal{A})\\ &\overset{(14)}{=} h(Q_1(\mathcal{A}),\ldots,Q_K(\mathcal{A}))\\ &\overset{(117)}{=} \sum_{k=1}^K w_k Q_k(\mathcal{A})\\ &= \int_{\mathcal{A}} \sum_{k=1}^K w_k q_k(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\,,\end{aligned}$$

which implies $q(\boldsymbol{\theta}) = \sum_{k=1}^K w_k q_k(\boldsymbol{\theta})$.

Appendix B
Proof of Theorem 2

A. Axioms Satisfied by the Generalized Linear Pooling Function

We first show that all the mentioned axioms are satisfied by the generalized linear pooling function. Let $g[q_1,\ldots,q_K](\boldsymbol{\theta}) = \sum_{k=0}^K w_k q_k(\boldsymbol{\theta})$ with $(w_0,\ldots,w_K) \in \mathcal{S}_{K+1}$. To show the WSFP (A5), we define for an event $\mathcal{A} \subseteq \Theta$

$$h_{\mathcal{A}}(p_1,\ldots,p_K) = w_0 \int_{\mathcal{A}} q_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} + \sum_{k=1}^K w_k p_k\,, \qquad (118)$$

for all $(p_1,\ldots,p_K) \in [0,1]^K$. For any opinion profile $(q_1,\ldots,q_K)$, we then have that

$$\begin{aligned}Q(\mathcal{A}) &= \int_{\mathcal{A}} q(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\\ &= \int_{\mathcal{A}} \sum_{k=0}^K w_k q_k(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\\ &= \sum_{k=0}^K w_k \int_{\mathcal{A}} q_k(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\\ &= w_0 \int_{\mathcal{A}} q_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} + \sum_{k=1}^K w_k Q_k(\mathcal{A})\\ &\overset{(118)}{=} h_{\mathcal{A}}(Q_1(\mathcal{A}),\ldots,Q_K(\mathcal{A}))\,,\end{aligned}$$

i.e., $h_{\mathcal{A}}$ satisfies the condition stated in A5. The weak likelihood principle (A7) is obviously satisfied with

$$h_{\boldsymbol{\theta}}(t_1,\ldots,t_K) = w_0 q_0(\boldsymbol{\theta}) + \sum_{k=1}^K w_k t_k\,,$$

as is the symmetry statement in Theorem 2.

B. Converse Statement

We now prove the converse statement in Theorem 2, i.e., that any pooling function g that satisfies the WSFP (A5) is a generalized linear pooling function. For each event $\mathcal{A} \subseteq \Theta$, let $h_{\mathcal{A}}\colon [0,1]^K \to [0,1]$ denote the function satisfying (16) for all opinion profiles. Our proof consists of three steps: First, we construct the pdf q_0 and the corresponding weight $w_0 \le 1$. In the second step, we show that by adapting each function $h_{\mathcal{A}}$ to $\tilde{h}_{\mathcal{A}} = \big(h_{\mathcal{A}} - w_0 \int_{\mathcal{A}} q_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\big)/(1-w_0)$, we obtain a linear pooling function. Finally, we show that this implies that g is a generalized linear pooling function.

Step 1: Construct q_0 and w_0: We define

$$Q_0(\mathcal{A}) = h_{\mathcal{A}}(0,\ldots,0) \qquad (119)$$

for all nontrivial (i.e., $|\mathcal{A}|, |\mathcal{A}^c| > 0$) events $\mathcal{A} \subseteq \Theta$. The pdf q_0 will be a weighted version of a density associated with Q_0. Thus, we first show that Q_0 can be expressed as an integral $Q_0(\mathcal{A}) = \int_{\mathcal{A}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$.

Let $\mathcal{A}_0$ be a fixed nontrivial event. Because $|\mathcal{A}_0^c| > 0$, there exists an opinion profile $(q_1^{(\mathcal{A}_0)},\ldots,q_K^{(\mathcal{A}_0)})$ such that $Q_k^{(\mathcal{A}_0)}(\mathcal{A}_0) = \int_{\mathcal{A}_0} q_k^{(\mathcal{A}_0)}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = 0$ for all $k = 1,\ldots,K$. We denote the fused pdf of this particular profile as $q^{(\mathcal{A}_0)}(\boldsymbol{\theta})$ and the resulting probability measure as

$$Q^{(\mathcal{A}_0)}(\mathcal{A}) = \int_{\mathcal{A}_0} q^{(\mathcal{A}_0)}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\,. \qquad (120)$$

Then,

$$Q^{(\mathcal{A}_0)}(\mathcal{A}) \overset{(16)}{=} h_{\mathcal{A}}(Q_1^{(\mathcal{A}_0)}(\mathcal{A}),\ldots,Q_K^{(\mathcal{A}_0)}(\mathcal{A}))\,.$$

In particular, for any event $\mathcal{A} \subseteq \mathcal{A}_0$, we have $Q_k^{(\mathcal{A}_0)}(\mathcal{A}) = 0$ for all $k = 1,\ldots,K$ (because $Q_k^{(\mathcal{A}_0)}(\mathcal{A}) \le Q_k^{(\mathcal{A}_0)}(\mathcal{A}_0) = 0$), and thus we obtain further

$$Q^{(\mathcal{A}_0)}(\mathcal{A}) = h_{\mathcal{A}}(0,\ldots,0) \overset{(119)}{=} Q_0(\mathcal{A})\,.$$

Recalling (120), we conclude that the fused pdf $q^{(\mathcal{A}_0)}(\boldsymbol{\theta})$ satisfies

$$Q_0(\mathcal{A}) = \int_{\mathcal{A}} q^{(\mathcal{A}_0)}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \quad \text{for any event } \mathcal{A} \subseteq \mathcal{A}_0. \qquad (121)$$

Following the same steps with $\mathcal{A}_0$ replaced by $\mathcal{A}_0^c$, we obtain a pdf $q^{(\mathcal{A}_0^c)}(\boldsymbol{\theta})$ such that we have

$$Q_0(\mathcal{A}) = \int_{\mathcal{A}} q^{(\mathcal{A}_0^c)}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \quad \text{for any event } \mathcal{A} \subseteq \mathcal{A}_0^c. \qquad (122)$$

Now for an arbitrary nontrivial event $\mathcal{B} \subseteq \Theta$, there exists an opinion profile $(q_1^{(\mathcal{B})},\ldots,q_K^{(\mathcal{B})})$ such that $Q_k^{(\mathcal{B})}(\mathcal{B}) = \int_{\mathcal{B}} q_k^{(\mathcal{B})}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = 0$ for all $k = 1,\ldots,K$. Again we denote the fused probability measure as $Q^{(\mathcal{B})}$. We thus obtain for $Q_0(\mathcal{B})$ as defined by (119)

$$\begin{aligned}Q_0(\mathcal{B}) &= h_{\mathcal{B}}(0,\ldots,0)\\ &= h_{\mathcal{B}}(Q_1^{(\mathcal{B})}(\mathcal{B}),\ldots,Q_K^{(\mathcal{B})}(\mathcal{B}))\\ &\overset{(16)}{=} Q^{(\mathcal{B})}(\mathcal{B})\,.\end{aligned} \qquad (123)$$

Because $\mathcal{B}$ can be decomposed into disjoint subsets according to $\mathcal{B} = (\mathcal{B} \cap \mathcal{A}_0) \cup (\mathcal{B} \cap \mathcal{A}_0^c)$, we further obtain from (123)

$$\begin{aligned}Q_0(\mathcal{B}) &= Q^{(\mathcal{B})}(\mathcal{B} \cap \mathcal{A}_0) + Q^{(\mathcal{B})}(\mathcal{B} \cap \mathcal{A}_0^c)\\ &\overset{(16)}{=} h_{\mathcal{B}\cap\mathcal{A}_0}(Q_1^{(\mathcal{B})}(\mathcal{B} \cap \mathcal{A}_0),\ldots,Q_K^{(\mathcal{B})}(\mathcal{B} \cap \mathcal{A}_0))\\ &\quad + h_{\mathcal{B}\cap\mathcal{A}_0^c}(Q_1^{(\mathcal{B})}(\mathcal{B} \cap \mathcal{A}_0^c),\ldots,Q_K^{(\mathcal{B})}(\mathcal{B} \cap \mathcal{A}_0^c))\\ &\overset{(a)}{=} h_{\mathcal{B}\cap\mathcal{A}_0}(0,\ldots,0) + h_{\mathcal{B}\cap\mathcal{A}_0^c}(0,\ldots,0)\end{aligned}$$

$$\overset{(119)}{=} Q_0(\mathcal{B}\cap\mathcal{A}_0) + Q_0(\mathcal{B}\cap\mathcal{A}_0^c)\,,$$

where we used $Q_k^{(\mathcal{B})}(\mathcal{B}) = 0$ in (a). Using (121) with $\mathcal{A} = \mathcal{B}\cap\mathcal{A}_0$ and (122) with $\mathcal{A} = \mathcal{B}\cap\mathcal{A}_0^c$, this implies

$$\begin{aligned} Q_0(\mathcal{B}) &= \int_{\mathcal{B}\cap\mathcal{A}_0} q^{(\mathcal{A}_0)}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} + \int_{\mathcal{B}\cap\mathcal{A}_0^c} q^{(\mathcal{A}_0^c)}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \\ &= \int_{\mathcal{B}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\,, \end{aligned}$$

where we defined

$$\tilde{q}_0(\boldsymbol{\theta}) \triangleq \begin{cases} q^{(\mathcal{A}_0)}(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{A}_0 \\ q^{(\mathcal{A}_0^c)}(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{A}_0^c. \end{cases}$$

We thus found an integral representation for Q_0 and can define

$$Q_0(\mathcal{B}) = \int_{\mathcal{B}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \tag{124}$$

also for trivial events $\mathcal{B}$. The nonnegativity of $q^{(\mathcal{A}_0)}(\boldsymbol{\theta})$ and $q^{(\mathcal{A}_0^c)}(\boldsymbol{\theta})$ implies that $\tilde{q}_0(\boldsymbol{\theta})$ is nonnegative and, in turn, that Q_0 is a measure. However, $\tilde{q}_0(\boldsymbol{\theta})$ is not a pdf in general.

We define

$$w_0 \triangleq Q_0(\Theta) \tag{125}$$

(note that this implies $w_0 \geq 0$) and

$$q_0(\boldsymbol{\theta}) \triangleq \frac{\tilde{q}_0(\boldsymbol{\theta})}{w_0}, \tag{126}$$

provided $w_0 \neq 0$. If $w_0 = 0$, we choose $q_0(\boldsymbol{\theta})$ as an arbitrary pdf. We claim that $w_0 \leq 1$. To prove this claim, let $\mathcal{B}_n$ be a sequence of nontrivial events such that $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$ and $\lim_{n\to\infty}\mathcal{B}_n = \Theta$. For each $\mathcal{B}_n$, there exists an opinion profile $(q_1^{(\mathcal{B}_n)},\ldots,q_K^{(\mathcal{B}_n)})$ such that $Q_k^{(\mathcal{B}_n)}(\mathcal{B}_n) = \int_{\mathcal{B}_n} q_k^{(\mathcal{B}_n)}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = 0$ for all $k = 1,\ldots,K$. Again we denote the sequence of fused probability measures as $Q^{(\mathcal{B}_n)}$. Following the steps in (123), we have that

$$Q_0(\mathcal{B}_n) = Q^{(\mathcal{B}_n)}(\mathcal{B}_n) \leq 1\,,$$

because $Q^{(\mathcal{B}_n)}$ is a probability measure. The continuity from below of measures [197, Lem. 3.4] implies $w_0 = Q_0(\Theta) = \lim_{n\to\infty} Q_0(\mathcal{B}_n) \leq 1$.

A similar argument can be employed to show (for later use) that for any nontrivial event $\mathcal{A}\subseteq\Theta$ and arbitrary probabilities p_k

$$h_{\mathcal{A}}(p_1,\ldots,p_K) \geq Q_0(\mathcal{A})\,. \tag{127}$$

Indeed, for any nontrivial event $\mathcal{A}\subseteq\Theta$, let $\mathcal{B}_n\subseteq\mathcal{A}$ be a sequence satisfying $|\mathcal{A}\setminus\mathcal{B}_n| > 0$, $\mathcal{B}_n\subseteq\mathcal{B}_{n+1}$ for all $n\in\mathbb{N}$, and $\lim_{n\to\infty}\mathcal{B}_n = \mathcal{A}$. Then for each $n\in\mathbb{N}$ there exists an opinion profile $(q_1^{(\mathcal{B}_n)},\ldots,q_K^{(\mathcal{B}_n)})$ satisfying $Q_k^{(\mathcal{B}_n)}(\mathcal{B}_n) = 0$, $Q_k^{(\mathcal{B}_n)}(\mathcal{A}\setminus\mathcal{B}_n) = p_k$, and, in turn, $Q_k^{(\mathcal{B}_n)}(\mathcal{A}) = p_k$. Again we denote the sequence of fused probability measures as $Q^{(\mathcal{B}_n)}$. Following the steps in (123), $Q_0(\mathcal{B}_n) = Q^{(\mathcal{B}_n)}(\mathcal{B}_n)$. Thus, we have

$$\begin{aligned} h_{\mathcal{A}}(p_1,\ldots,p_K) &= h_{\mathcal{A}}(Q_1^{(\mathcal{B}_n)}(\mathcal{A}),\ldots,Q_K^{(\mathcal{B}_n)}(\mathcal{A})) \\ &\overset{(16)}{=} Q^{(\mathcal{B}_n)}(\mathcal{A}) \\ &\geq Q^{(\mathcal{B}_n)}(\mathcal{B}_n) \\ &= Q_0(\mathcal{B}_n)\,. \end{aligned} \tag{128}$$

Here, $h_{\mathcal{A}}(p_1,\ldots,p_K)$ does not depend on n. Hence, we can take the limit on the right-hand side of (128) and obtain

$$h_{\mathcal{A}}(p_1,\ldots,p_K) \geq \lim_{n\to\infty} Q_0(\mathcal{B}_n) = Q_0(\mathcal{A})\,,$$

using again the continuity from below of Q_0.

Step 2: Define $\tilde{h}_{\mathcal{A}}$ and prove that it defines a linear pooling function: We define

$$\tilde{h}_{\mathcal{A}}(p_1,\ldots,p_K) \triangleq \frac{h_{\mathcal{A}}(p_1,\ldots,p_K) - Q_0(\mathcal{A})}{1-w_0}\,. \tag{129}$$

Here, we have to assume that $w_0 < 1$. Thus, we first show that g is a generalized linear pooling function in the case $w_0 = 1$. In this case, for any nontrivial event $\mathcal{A}\subseteq\Theta$ and arbitrary probabilities p_k, we choose an opinion profile that satisfies $Q_k(\mathcal{A}) = p_k$ and hence $Q_k(\mathcal{A}^c) = 1-p_k$ for all $k = 1,\ldots,K$. We then have

$$\begin{aligned} 1 &= Q(\mathcal{A}) + Q(\mathcal{A}^c) \\ &\overset{(16)}{=} h_{\mathcal{A}}(p_1,\ldots,p_K) + h_{\mathcal{A}^c}(1-p_1,\ldots,1-p_K) \\ &\overset{(127)}{\geq} Q_0(\mathcal{A}) + Q_0(\mathcal{A}^c) \\ &\overset{(125)}{=} w_0 \\ &= 1\,. \end{aligned}$$

Thus, the inequality in the third line is actually an equality, which is only possible if $h_{\mathcal{A}}(p_1,\ldots,p_K) = Q_0(\mathcal{A})$. Because $\mathcal{A}$ and the p_k were chosen arbitrarily, we have $h_{\mathcal{A}}(p_1,\ldots,p_K) = Q_0(\mathcal{A})$ independently of the probabilities p_k. By (16), this further implies for any opinion profile $(q_1,\ldots,q_K)$ that the aggregate pdf q satisfies

$$\begin{aligned} \int_{\mathcal{A}} q(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} &= Q(\mathcal{A}) \\ &\overset{(16)}{=} h_{\mathcal{A}}(Q_1(\mathcal{A}),\ldots,Q_K(\mathcal{A})) \\ &= Q_0(\mathcal{A}) \\ &\overset{(124)}{=} \int_{\mathcal{A}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \end{aligned}$$

for all events $\mathcal{A}$. Hence, $q(\boldsymbol{\theta}) = \tilde{q}_0(\boldsymbol{\theta})$, which implies that g is a dogmatic pooling function (which is a special case of a generalized linear pooling function with weights $w_0 = 1$, $w_k = 0$ for $k = 1,\ldots,K$). This concludes the proof for the special case $w_0 = 1$, and thus we can assume $w_0 < 1$ in what follows.

We define a new fusion rule $\tilde{g}$ by

$$\tilde{g}[q_1,\ldots,q_K](\boldsymbol{\theta}) \triangleq \frac{g[q_1,\ldots,q_K](\boldsymbol{\theta}) - w_0 q_0(\boldsymbol{\theta})}{1-w_0} \tag{130}$$

and claim that it satisfies the WSFP with the functions $\tilde{h}_{\mathcal{A}}$ defined by (129). Indeed, we have for any opinion profile $(q_1,\ldots,q_K)$ and any event $\mathcal{A}\subseteq\Theta$ that

$$\begin{aligned} \tilde{Q}(\mathcal{A}) &= \int_{\mathcal{A}} \tilde{g}[q_1,\ldots,q_K](\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \\ &= \int_{\mathcal{A}} \frac{g[q_1,\ldots,q_K](\boldsymbol{\theta}) - w_0 q_0(\boldsymbol{\theta})}{1-w_0}\,\mathrm{d}\boldsymbol{\theta} \end{aligned}$$

$$\stackrel{(a)}{=} \frac{h_{\mathcal{A}}(Q_1(\mathcal{A}), \dots, Q_K(\mathcal{A})) - \int_{\mathcal{A}} \tilde{q}_0(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}}{1 - w_0}$$
$$\stackrel{(124)}{=} \frac{h_{\mathcal{A}}(Q_1(\mathcal{A}), \dots, Q_K(\mathcal{A})) - Q_0(\mathcal{A})}{1 - w_0}$$
$$\stackrel{(129)}{=} \tilde{h}_{\mathcal{A}}(Q_1(\mathcal{A}), \dots, Q_K(\mathcal{A}))\,, \tag{131}$$

where we used in (a) that, by (16), $h_{\mathcal{A}}(Q_1(\mathcal{A}), \dots, Q_K(\mathcal{A})) = Q(\mathcal{A}) = \int_{\mathcal{A}} g[q_1, \dots, q_K](\boldsymbol{\theta})$ and, by (126), $w_0 q_0(\boldsymbol{\theta}) = \tilde{q}_0(\boldsymbol{\theta})$. Furthermore, we claim that $\tilde{g}$ satisfies the ZPP. To prove this, let $(q_1, \dots, q_K)$ be an opinion profile and $\mathcal{A}$ a nontrivial event such that $Q_k(\mathcal{A}) = 0$ for all $k = 1, \dots, K$. Because $h_{\mathcal{A}}(0, \dots, 0) = Q_0(\mathcal{A})$,

$$\begin{aligned} \tilde{Q}(\mathcal{A}) &\stackrel{(131)}{=} \tilde{h}_{\mathcal{A}}(Q_1(\mathcal{A}), \dots, Q_K(\mathcal{A})) \\ &= \tilde{h}_{\mathcal{A}}(0, \dots, 0) \\ &\stackrel{(129)}{=} \frac{h_{\mathcal{A}}(0, \dots, 0) - Q_0(\mathcal{A})}{1 - w_0} \\ &= \frac{Q_0(\mathcal{A}) - Q_0(\mathcal{A})}{1 - w_0} \\ &= 0\,, \end{aligned}$$

proving the ZPP.

Finally, to see that $\tilde{g}$ is a valid pooling function, we first show that for any $(p_1, \dots, p_K) \in [0,1]^K$, the function $\tilde{h}_{\mathcal{A}}(p_1, \dots, p_K)$ is nonnegative. This follows from (129), (127), and our assumption $w_0 < 1$. Hence, the measure $\tilde{Q}$ is nonnegative and thus also the associated density $\tilde{g}[q_1, \dots, q_K](\boldsymbol{\theta})$ must be nonnegative. The fact that $\tilde{g}[q_1, \dots, q_K](\boldsymbol{\theta})$ integrates to one follows directly from the definition (130) and the fact that $g[q_1, \dots, q_K](\boldsymbol{\theta})$ and q_0 are pdfs.

Because $\tilde{g}$ is a pooling function that satisfies the WSFP and the ZPP, Theorem 1 implies that it is a linear pooling function, i.e.,

$$\tilde{g}[q_1, \dots, q_K](\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta})\,, \tag{132}$$

with $(w_1, \dots, w_K) \in \mathcal{S}_K$.

Step 3: Conclude that $\tilde{g}$ is a generalized linear pooling function: Combining (130) and (132), we obtain

$$\sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta}) = \frac{g[q_1, \dots, q_K](\boldsymbol{\theta}) - w_0 q_0(\boldsymbol{\theta})}{1 - w_0}$$

or, equivalently,

$$g[q_1, \dots, q_K](\boldsymbol{\theta}) = w_0 q_0(\boldsymbol{\theta}) + \sum_{k=1}^{K} (1 - w_0) w_k q_k(\boldsymbol{\theta})\,.$$

From $\sum_{k=1}^{K} w_k = 1$, it follows that $w_0 + \sum_{k=1}^{K} (1 - w_0) w_k$ is one. Thus, g is a generalized linear pooling function.

Appendix C
Proof of the Equivalence Statement in Theorem 6

We only show that (ii), i.e., g satisfies individualized Bayesianity (A11) and $g[q_0, \dots, q_0](\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$ for some pdf q_0, implies (i), i.e., g is a multiplicative pooling function. The other direction is obvious.

Thus, let us assume that $g[q_0, \dots, q_0](\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$ for some pdf q_0. We have to show that, for any opinion profile $(q_1, \dots, q_K)$ such that q_k / q_0 is bounded for all $k = 1, \dots, K$ (recall that we only consider those opinion profiles in the multiplicative pooling function), q is of the form (10), i.e.,

$$g[q_1, \dots, q_K](\boldsymbol{\theta}) \propto (q_0(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} q_k(\boldsymbol{\theta})\,.$$

To this end, we first note that $q_k = q_0^{(\ell_k)}$ (see (17)) with $\ell_k = q_k / q_0$ for all $k = 1, \dots, K$. Thus,

$$g[q_1, \dots, q_K](\boldsymbol{\theta}) = g\Big[q_0^{(q_1/q_0)}, \dots, q_0^{(q_K/q_0)}\Big](\boldsymbol{\theta})\,.$$

By iteratively using individualized Bayesianity (19) with $\ell = q_k / q_0$ for each $k = 1, \dots, K$, we obtain further

$$\begin{aligned} g[q_1, \dots, q_K](\boldsymbol{\theta}) &\propto g\Big[q_0, q_0^{(q_2/q_0)}, \dots, q_0^{(q_K/q_0)}\Big](\boldsymbol{\theta}) \frac{q_1(\boldsymbol{\theta})}{q_0(\boldsymbol{\theta})} \\ &\propto g[q_0, \dots, q_0](\boldsymbol{\theta}) \prod_{k=1}^{K} \frac{q_k(\boldsymbol{\theta})}{q_0(\boldsymbol{\theta})} \\ &= (q_0(\boldsymbol{\theta}))^{1-K} \prod_{k=1}^{K} q_k(\boldsymbol{\theta})\,, \end{aligned}$$

which is (10) and thus concludes the proof.

Appendix D
Partial Proof of Theorem 8

1) (iii) implies (ii): We first show that (iii), i.e., the WSFP (A5) and independence preservation (A8), implies (ii), i.e., the SSFP (A4) and independence preservation (A8). To this end, we show that independence preservation implies the ZPP (A2). The ZPP and the assumed WSFP in turn imply the SSFP by Theorem 1.

To show that independence preservation implies the ZPP, assume that for some event $\mathcal{A}$, we have $Q_k(\mathcal{A}) = 0$ for all $k = 1, \dots, K$. This implies that

$$Q_k(\mathcal{A} \cap \mathcal{A}) = Q_k(\mathcal{A}) = 0 = Q_k(\mathcal{A}) Q_k(\mathcal{A})\,.$$

Independence preservation now implies that also Q must satisfy $Q(\mathcal{A} \cap \mathcal{A}) = Q(\mathcal{A}) Q(\mathcal{A})$, and thus that either $Q(\mathcal{A}) = 0$ or $Q(\mathcal{A}) = 1$. In the first case, the proof of the ZPP is finished. In the second case, i.e., $Q(\mathcal{A}) = 1$, there must exist a subset $\mathcal{B} \subseteq \mathcal{A}$ such that $Q(\mathcal{B}) = 1/2$. However, because $\mathcal{B} \subseteq \mathcal{A}$ and $Q_k(\mathcal{A}) = 0$, we have that also $Q_k(\mathcal{B}) = 0$, and thus we again have that $Q_k(\mathcal{B} \cap \mathcal{B}) = 0 = Q_k(\mathcal{B}) Q_k(\mathcal{B})$. This implies that $Q(\mathcal{B})$ is either 0 or 1, which is a contradiction to $Q(\mathcal{B}) = 1/2$. Thus, $Q(\mathcal{A}) = 0$ is the only valid conclusion, which proves that the ZPP is satisfied.

2) (v) implies (iv): We next show that (v), i.e., the WSFP (A5) and external Bayesianity (A10), implies (iv), i.e., the SSFP (A4) and external Bayesianity (A10). Thus, we have to show that the WSFP and external Bayesianity imply the SSFP.

By Theorem 2, the WSFP implies the weak likelihood principle (A7). Furthermore, by Theorem 4, the weak likelihood principle and external Bayesianity imply that g is a generalized log-linear pooling function, i.e.,

$$g[q_1,\ldots,q_K](\boldsymbol{\theta}) = c\,\xi_0(\boldsymbol{\theta}) \prod_{k=1}^{K} (q_k(\boldsymbol{\theta}))^{w_k} \tag{133}$$

for all positive opinion profiles. Finally, by Theorem 2, the WSFP implies that g is also a generalized linear pooling function, i.e.,

$$g[q_1,\ldots,q_K](\boldsymbol{\theta}) = \sum_{k=0}^{K} w'_k q_k(\boldsymbol{\theta}) \tag{134}$$

for all opinion profiles. Thus, combining (133) and (134), we have

$$c\,\xi_0(\boldsymbol{\theta}) \prod_{k=1}^{K} (q_k(\boldsymbol{\theta}))^{w_k} = \sum_{k=0}^{K} w'_k q_k(\boldsymbol{\theta}) \tag{135}$$

for all positive opinion profiles. Note that $q_0(\boldsymbol{\theta})$ is not necessarily positive, i.e., it may be zero for certain values of $\boldsymbol{\theta}$.

We choose an arbitrary positive pdf $\tilde{q}_0(\boldsymbol{\theta})$ and $\varepsilon \in (0,1)$ and consider the opinion profile $(\varepsilon\tilde{q}_0 + (1-\varepsilon)q_0, \ldots, \varepsilon\tilde{q}_0 + (1-\varepsilon)q_0)$. Since $\tilde{q}_0(\boldsymbol{\theta})$ is positive, this is a positive opinion profile for any $\varepsilon \in (0,1)$. Using it in (135) gives

$$\begin{aligned} &c\,\xi_0(\boldsymbol{\theta})\big(\varepsilon\tilde{q}_0(\boldsymbol{\theta}) + (1-\varepsilon)q_0(\boldsymbol{\theta})\big) \\ &\quad = w'_0 q_0(\boldsymbol{\theta}) + (1-w'_0)\big(\varepsilon\tilde{q}_0(\boldsymbol{\theta}) + (1-\varepsilon)q_0(\boldsymbol{\theta})\big), \end{aligned}$$

where $\sum_{k=1}^{K} w_k = 1$ and $\sum_{k=0}^{K} w'_k = 1$ were used, or, equivalently,

$$c\,\xi_0(\boldsymbol{\theta}) = w'_0 \frac{q_0(\boldsymbol{\theta})}{\varepsilon\tilde{q}_0(\boldsymbol{\theta}) + (1-\varepsilon)q_0(\boldsymbol{\theta})} + 1 - w'_0\,. \tag{136}$$

Taking the limit $\varepsilon \to 0$ in (136), we obtain

$$c\,\xi_0(\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } q_0(\boldsymbol{\theta}) > 0 \\ 1 - w'_0 & \text{if } q_0(\boldsymbol{\theta}) = 0\,. \end{cases} \tag{137}$$

Inserting into (133) and evaluating (133) for the opinion profile $(\tilde{q}_0, \ldots, \tilde{q}_0)$ yields

$$g[\tilde{q}_0,\ldots,\tilde{q}_0](\boldsymbol{\theta}) = \begin{cases} \tilde{q}_0(\boldsymbol{\theta}) & \text{if } q_0(\boldsymbol{\theta}) > 0 \\ (1-w'_0)\tilde{q}_0(\boldsymbol{\theta}) & \text{if } q_0(\boldsymbol{\theta}) = 0\,. \end{cases}$$

Because $g[\tilde{q}_0,\ldots,\tilde{q}_0](\boldsymbol{\theta})$ is a pdf, this implies

$$\begin{aligned} 1 &= \int_{\Theta} g[\tilde{q}_0,\ldots,\tilde{q}_0](\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \\ &= \int_{\{\boldsymbol{\theta}\in\Theta: q_0(\boldsymbol{\theta})>0\}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \\ &\quad + (1-w'_0)\int_{\{\boldsymbol{\theta}\in\Theta: q_0(\boldsymbol{\theta})=0\}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\,. \end{aligned} \tag{138}$$

On the other hand, because $\tilde{q}_0(\boldsymbol{\theta})$ is a pdf, we have

$$\begin{aligned} 1 &= \int_{\Theta} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \\ &= \int_{\{\boldsymbol{\theta}\in\Theta: q_0(\boldsymbol{\theta})>0\}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} + \int_{\{\boldsymbol{\theta}\in\Theta: q_0(\boldsymbol{\theta})=0\}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\,. \end{aligned} \tag{139}$$

Combining (138) and (139), we obtain

$$(1-w'_0)\int_{\{\boldsymbol{\theta}\in\Theta: q_0(\boldsymbol{\theta})=0\}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = \int_{\{\boldsymbol{\theta}\in\Theta: q_0(\boldsymbol{\theta})=0\}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$$

or equivalently

$$w'_0 \int_{\{\boldsymbol{\theta}\in\Theta: q_0(\boldsymbol{\theta})=0\}} \tilde{q}_0(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = 0\,.$$

Since $\tilde{q}_0(\boldsymbol{\theta})$ is a positive pdf on Θ, this can only hold if either $w'_0 = 0$ or $|\{\boldsymbol{\theta}\in\Theta : q_0(\boldsymbol{\theta}) = 0\}| = 0$. In the first case, (134) implies that g is actually a linear pooling function and thus, by Theorem 1, g satisfies the SSFP. In the second case, $q_0(\boldsymbol{\theta}) > 0$ almost everywhere and thus (137) states that $c\,\xi_0(\boldsymbol{\theta}) = 1$. Using the opinion profile $(q_1,\ldots,q_K) = (\tilde{q}_0,\ldots,\tilde{q}_0)$ in (135) now gives

$$\tilde{q}_0(\boldsymbol{\theta}) = w'_0 q_0(\boldsymbol{\theta}) + (1-w'_0)\tilde{q}_0(\boldsymbol{\theta})\,. \tag{140}$$

In particular, let us partition Θ into disjoint sets $\mathcal{A}_1$, $\mathcal{A}_2$ satisfying $\int_{\mathcal{A}_1} q_0(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = \int_{\mathcal{A}_2} q_0(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = 1/2$, and let us choose

$$\tilde{q}_0(\boldsymbol{\theta}) = \begin{cases} \frac{3}{2}q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{A}_1 \\ \frac{1}{2}q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{A}_2\,. \end{cases}$$

Then (140) yields for all $\boldsymbol{\theta}\in\mathcal{A}_1$

$$\frac{3}{2}q_0(\boldsymbol{\theta}) = w'_0 q_0(\boldsymbol{\theta}) + (1-w'_0)\frac{3}{2}q_0(\boldsymbol{\theta}) = \left(\frac{3}{2} - \frac{1}{2}w'_0\right)q_0(\boldsymbol{\theta})\,.$$

This implies $w'_0 = 0$, and hence we again conclude from (134) that g is actually a linear pooling function, and thus, by Theorem 1, that g satisfies the SSFP.

3) (vi) implies (i): Finally, we prove that (vi), i.e., the SSFP (A4) and generalized Bayesianity (A12), implies (i), i.e., that g is a dictatorship pooling function. By Theorem 1, the SSFP implies that g is a linear pooling function, i.e.,

$$g[q_1,\ldots,q_K](\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta}) \tag{141}$$

with $(w_1,\ldots,w_K) \in \mathcal{S}_K$. We will show that for an arbitrary k the weight w_k is either 0 or 1, which is equivalent to g being a dictatorship pooling function.

We first choose a positive function f and two disjoint sets $\mathcal{A}$ and $\mathcal{B}$ such that $\Theta = \mathcal{A}\cup\mathcal{B}$ and $\int_{\mathcal{A}} f(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = \int_{\mathcal{B}} f(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = 1$. We fix an arbitrary k and define an opinion profile $(q_1,\ldots,q_K)$ by setting

$$q_k(\boldsymbol{\theta}) = \begin{cases} \frac{1}{3}f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{A} \\ \frac{2}{3}f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{B} \end{cases}$$

and $q_{k'} = q_0$ for all $k' \neq k$, where

$$q_0(\boldsymbol{\theta}) = \frac{1}{2}f(\boldsymbol{\theta})\,. \tag{142}$$

Inserting this opinion profile into the fusion rule (141) and using $\sum_{k'\neq k} w_{k'} = 1 - w_k$ gives

$$\begin{aligned} g[q_1,\ldots,q_K](\boldsymbol{\theta}) &= \begin{cases} \big(\frac{1}{3}w_k + \frac{1}{2}(1-w_k)\big)f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{A} \\ \big(\frac{2}{3}w_k + \frac{1}{2}(1-w_k)\big)f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{B} \end{cases} \\ &= \begin{cases} \big(\frac{1}{2} - \frac{1}{6}w_k\big)f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{A} \\ \big(\frac{1}{2} + \frac{1}{6}w_k\big)f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\mathcal{B}\,. \end{cases} \end{aligned} \tag{143}$$

Next, we use generalized Bayesianity with $\ell_{k'} = \ell$ for all $k' = 1, \dots, K$, where

$$\ell(\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ 2 & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases}$$

We easily obtain (see (17))

$$q_k^{(\ell)}(\boldsymbol{\theta}) = \begin{cases} \frac{1}{5} f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \frac{4}{5} f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,, \end{cases}$$

and

$$q_0^{(\ell)}(\boldsymbol{\theta}) = \begin{cases} \frac{1}{3} f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \frac{2}{3} f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases} \tag{144}$$

Now, (141) gives

$$\begin{aligned} g[q_1^{(\ell)}, \dots, q_K^{(\ell)}](\boldsymbol{\theta}) &= \begin{cases} \big(\frac{1}{5} w_k + \frac{1}{3}(1-w_k)\big) f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(\frac{4}{5} w_k + \frac{2}{3}(1-w_k)\big) f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B} \end{cases} \\ &= \begin{cases} \big(\frac{1}{3} - \frac{2}{15} w_k\big) f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(\frac{2}{3} + \frac{2}{15} w_k\big) f(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases} \end{aligned} \tag{145}$$

On the other hand, because g satisfies generalized Bayesianity, there exists a function $h[\ell, \dots, \ell]$ such that

$$g\big[q_1^{(\ell)}, \dots, q_K^{(\ell)}\big](\boldsymbol{\theta}) = \frac{g[q_1, \dots, q_K](\boldsymbol{\theta}) h[\ell, \dots, \ell](\boldsymbol{\theta})}{c_\ell}, \tag{146}$$

where $c_\ell = \int_\Theta g[q_1, \dots, q_K](\boldsymbol{\theta}) h[\ell, \dots, \ell](\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}$. Inserting (143) and (145) into (146) gives

$$\frac{1}{3} - \frac{2}{15} w_k = \frac{\big(\frac{1}{2} - \frac{1}{6} w_k\big) h[\ell, \dots, \ell](\boldsymbol{\theta})}{c_\ell} \tag{147}$$

for all $\boldsymbol{\theta} \in \mathcal{A}$ and

$$\frac{2}{3} + \frac{2}{15} w_k = \frac{\big(\frac{1}{2} + \frac{1}{6} w_k\big) h[\ell, \dots, \ell](\boldsymbol{\theta})}{c_\ell} \tag{148}$$

for all $\boldsymbol{\theta} \in \mathcal{B}$.

Using again the generalized Bayesianity of g, we also have

$$g\big[q_0^{(\ell)}, \dots, q_0^{(\ell)}\big](\boldsymbol{\theta}) = \frac{g[q_0, \dots, q_0](\boldsymbol{\theta}) h[\ell, \dots, \ell](\boldsymbol{\theta})}{c_{0,\ell}}, \tag{149}$$

where $c_{0,\ell} = \int_\Theta g[q_0, \dots, q_0](\boldsymbol{\theta}) h[\ell, \dots, \ell](\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}$. Because linear pooling functions are unanimity preserving (see Theorem 1), we have $g\big[q_0^{(\ell)}, \dots, q_0^{(\ell)}\big](\boldsymbol{\theta}) = q_0^{(\ell)}(\boldsymbol{\theta})$ and $g[q_0, \dots, q_0](\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$, and thus (149) is equivalent to

$$q_0^{(\ell)}(\boldsymbol{\theta}) = \frac{q_0(\boldsymbol{\theta}) h[\ell, \dots, \ell](\boldsymbol{\theta})}{c_{0,\ell}},$$

or, inserting (144) and (142),

$$\frac{1}{3} = \frac{\frac{1}{2} h[\ell, \dots, \ell](\boldsymbol{\theta})}{c_{0,\ell}}$$

for all $\boldsymbol{\theta} \in \mathcal{A}$ and

$$\frac{2}{3} = \frac{\frac{1}{2} h[\ell, \dots, \ell](\boldsymbol{\theta})}{c_{0,\ell}}$$

for all $\boldsymbol{\theta} \in \mathcal{B}$. We thus obtain

$$h[\ell, \dots, \ell](\boldsymbol{\theta}) = \begin{cases} \frac{2}{3} c_{0,\ell} & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \frac{4}{3} c_{0,\ell} & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases}$$

Inserting this into (147) and (148) yields

$$\frac{1}{3} - \frac{2}{15} w_k = \frac{\big(\frac{1}{2} - \frac{1}{6} w_k\big) \frac{2}{3} c_{0,\ell}}{c_\ell}$$

and

$$\frac{2}{3} + \frac{2}{15} w_k = \frac{\big(\frac{1}{2} + \frac{1}{6} w_k\big) \frac{4}{3} c_{0,\ell}}{c_\ell}$$

or, equivalently,

$$\frac{\frac{1}{3} - \frac{2}{15} w_k}{\frac{1}{3} - \frac{1}{9} w_k} = \frac{c_{0,\ell}}{c_\ell} = \frac{\frac{2}{3} + \frac{2}{15} w_k}{\frac{2}{3} + \frac{2}{9} w_k}\,.$$

This amounts to the quadratic equation $w_k^2 - w_k = 0$, which has the solutions $w_k = 0$ and $w_k = 1$. Since k was arbitrary, this concludes the proof.

Appendix E
Proof of Lemma 9

By Theorem 2, the WSFP implies that g is a generalized linear pooling function, i.e.,

$$g[q_1, \dots, q_K](\boldsymbol{\theta}) = \sum_{k=0}^{K} w_k q_k(\boldsymbol{\theta}) \tag{150}$$

with $(w_0, \dots, w_K) \in \mathcal{S}_{K+1}$. We will show that w_0 is either 0 or 1, which is equivalent to g being either a linear pooling function or a dogmatic pooling function.

We first choose two disjoint sets $\mathcal{A}$ and $\mathcal{B}$ such that $\Theta = \mathcal{A} \cup \mathcal{B}$ and $\int_\mathcal{A} q_0(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} = \int_\mathcal{B} q_0(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} = 1/2$. Furthermore, we choose an opinion profile $(q_1, \dots, q_K)$ as

$$q_k(\boldsymbol{\theta}) = \begin{cases} \frac{2}{3} q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \frac{4}{3} q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B} \end{cases}$$

for all $k = 1, \dots, K$. Inserting this opinion profile into (150) and using $\sum_{k=1}^K w_k = 1 - w_0$ gives

$$\begin{aligned} g[q_1, \dots, q_K](\boldsymbol{\theta}) &= \begin{cases} \big(w_0 + \frac{2}{3}(1-w_0)\big) q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(w_0 + \frac{4}{3}(1-w_0)\big) q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B} \end{cases} \\ &= \begin{cases} \big(\frac{2}{3} + \frac{1}{3} w_0\big) q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(\frac{4}{3} - \frac{1}{3} w_0\big) q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases} \end{aligned} \tag{151}$$

Next, we use generalized Bayesianity with $\ell_k = \ell$ for all $k = 1, \dots, K$, where

$$\ell(\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ 2 & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases} \tag{152}$$

We easily obtain (see (17))

$$q_k^{(\ell)}(\boldsymbol{\theta}) = \begin{cases} \frac{2}{5} q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \frac{8}{5} q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B} \end{cases}$$

and then (150) gives

$$\begin{aligned} g[q_1^{(\ell)}, \dots, q_K^{(\ell)}](\boldsymbol{\theta}) &= \begin{cases} \big(w_0 + \frac{2}{5}(1-w_0)\big) q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(w_0 + \frac{8}{5}(1-w_0)\big) q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B} \end{cases} \\ &= \begin{cases} \big(\frac{2}{5} + \frac{3}{5} w_0\big) q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(\frac{8}{5} - \frac{3}{5} w_0\big) q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases} \end{aligned} \tag{153}$$

Because g satisfies generalized Bayesianity, we have that there exists a function $h[\ell,\dots,\ell]$ such that

$$g\big[q_1^{(\ell)},\dots,q_K^{(\ell)}\big](\boldsymbol{\theta}) = \frac{g[q_1,\dots,q_K](\boldsymbol{\theta})h[\ell,\dots,\ell](\boldsymbol{\theta})}{c_\ell}, \tag{154}$$

where $c_\ell = \int_\Theta g[q_1,\dots,q_K](\boldsymbol{\theta})h[\ell,\dots,\ell](\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$. Inserting (151) and (153) into (154) gives

$$\frac{2}{5} + \frac{3}{5}w_0 = \frac{\big(\frac{2}{3}+\frac{1}{3}w_0\big)h[\ell,\dots,\ell](\boldsymbol{\theta})}{c_\ell} \tag{155}$$

for all $\boldsymbol{\theta} \in \mathcal{A}$ and

$$\frac{8}{5} - \frac{3}{5}w_0 = \frac{\big(\frac{4}{3}-\frac{1}{3}w_0\big)h[\ell,\dots,\ell](\boldsymbol{\theta})}{c_\ell} \tag{156}$$

for all $\boldsymbol{\theta} \in \mathcal{B}$.

Using again the generalized Bayesianity of g, we also have

$$g\big[q_0^{(\ell)},\dots,q_0^{(\ell)}\big](\boldsymbol{\theta}) = \frac{g[q_0,\dots,q_0](\boldsymbol{\theta})h[\ell,\dots,\ell](\boldsymbol{\theta})}{c_{0,\ell}}, \tag{157}$$

where $c_{0,\ell} = \int_\Theta g[q_0,\dots,q_0](\boldsymbol{\theta})h[\ell,\dots,\ell](\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$. Using (152) and (17), we obtain

$$q_0^{(\ell)}(\boldsymbol{\theta}) = \begin{cases} \frac{2}{3}q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \frac{4}{3}q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases}$$

Inserting into (150) yields

$$\begin{aligned} g\big[q_0^{(\ell)},\dots,q_0^{(\ell)}\big](\boldsymbol{\theta}) &= \begin{cases} \big(w_0 + \frac{2}{3}(1-w_0)\big)q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(w_0 + \frac{4}{3}(1-w_0)\big)q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B} \end{cases} \\ &= \begin{cases} \big(\frac{2}{3} + \frac{1}{3}w_0\big)q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(\frac{4}{3} - \frac{1}{3}w_0\big)q_0(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases} \end{aligned} \tag{158}$$

Furthermore, again by (150), $g[q_0,\dots,q_0](\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$. Inserting this and (158) into (157), we obtain

$$\frac{2}{3} + \frac{1}{3}w_0 = \frac{h[\ell,\dots,\ell](\boldsymbol{\theta})}{c_{0,\ell}}$$

for all $\boldsymbol{\theta} \in \mathcal{A}$ and

$$\frac{4}{3} - \frac{1}{3}w_0 = \frac{h[\ell,\dots,\ell](\boldsymbol{\theta})}{c_{0,\ell}}$$

for all $\boldsymbol{\theta} \in \mathcal{B}$. Thus,

$$h[\ell,\dots,\ell](\boldsymbol{\theta}) = \begin{cases} \big(\frac{2}{3} + \frac{1}{3}w_0\big)c_{0,\ell} & \text{if } \boldsymbol{\theta} \in \mathcal{A} \\ \big(\frac{4}{3} - \frac{1}{3}w_0\big)c_{0,\ell} & \text{if } \boldsymbol{\theta} \in \mathcal{B}\,. \end{cases}$$

Inserting this into (155) and (156) gives

$$\frac{2}{5} + \frac{3}{5}w_0 = \frac{\big(\frac{2}{3}+\frac{1}{3}w_0\big)^2 c_{0,\ell}}{c_\ell}$$

and

$$\frac{8}{5} - \frac{3}{5}w_0 = \frac{\big(\frac{4}{3}-\frac{1}{3}w_0\big)^2 c_{0,\ell}}{c_\ell}$$

or, equivalently,

$$\frac{\frac{2}{5}+\frac{3}{5}w_0}{\big(\frac{2}{3}+\frac{1}{3}w_0\big)^2} = \frac{c_{0,\ell}}{c_\ell} = \frac{\frac{8}{5}-\frac{3}{5}w_0}{\big(\frac{4}{3}-\frac{1}{3}w_0\big)^2}\,.$$

This amounts to the cubic equation $w_0^3 - 3w_0^2 + 2w_0 = 0$, which has the solutions $w_0 = 0$, $w_0 = 1$, and $w_0 = 2$. Since w_0 cannot be larger than one, only the solutions $w_0 = 0$ and $w_0 = 1$ remain. In the first case, g is a linear pooling function, which satisfies the SSFP by Theorem 1. Hence, since g satisfies both the SSFP and generalized Bayesiantity, it reduces to a dictatorship pooling function by Theorem 8. In the second case, g is a dogmatic pooling function.

Appendix F
Proof of Theorem 11

The implications in (i) follow from Theorem 1 because the SSFP implies that g is a linear pooling function and in turn satisfies the ZPP (A2), unanimity preservation (A3), the WSFP (A5), the likelihood principle (A6), and the weak likelihood principle (A7). Similarly, the implications in (ii) follow from Theorem 2. Implication (iii) follows directly from the concerned axioms. Implication (iv) is shown in the first part of the proof of Theorem 8 in Appendix D. It remains to show implication (v), i.e., that individualized Bayesianity implies generalized Bayesianity. This can easily be seen by defining

$$h[\ell_1,\dots,\ell_K](\boldsymbol{\theta}) \triangleq \prod_{k=1}^K \ell_k(\boldsymbol{\theta})\,. \tag{159}$$

Indeed, because g satisfies individualized Bayesianity, iterative application of (19) implies

$$\begin{aligned} g[q_1^{(\ell_1)},\dots,q_K^{(\ell_K)}](\boldsymbol{\theta}) &\propto g\Big[q_1,q_2^{(\ell_2)},\dots,q_K^{(\ell_K)}\Big](\boldsymbol{\theta})\ell_1(\boldsymbol{\theta}) \\ &\propto g[q_1,\dots,q_K](\boldsymbol{\theta})\prod_{k=1}^K \ell_k(\boldsymbol{\theta}) \\ &\propto g[q_1,\dots,q_K]^{\left(\prod_{k=1}^K \ell_k\right)}(\boldsymbol{\theta}). \end{aligned}$$

Thus, (20) is satisfied by h defined in (159).

Appendix G
Proof of Theorem 14 (Constrained Minimization of the Weighted Average of α-Divergences)

Let $f_\alpha(x) = \frac{x^\alpha - 1}{\alpha(\alpha-1)}$. The inverse function is given by

$$f_\alpha^{-1}(x) = (x\alpha(\alpha-1)+1)^{1/\alpha}\,. \tag{160}$$

Furthermore, we have that for two functions $p_1(\boldsymbol{\theta})$ and $p_2(\boldsymbol{\theta})$

$$\begin{aligned} f_\alpha\left(\frac{p_1(\boldsymbol{\theta})}{p_2(\boldsymbol{\theta})}\right) &= \frac{\left(\frac{p_1(\boldsymbol{\theta})}{p_2(\boldsymbol{\theta})}\right)^\alpha - 1}{\alpha(\alpha-1)} \\ &= \frac{(p_1(\boldsymbol{\theta}))^\alpha - (p_2(\boldsymbol{\theta}))^\alpha}{(p_2(\boldsymbol{\theta}))^\alpha\alpha(\alpha-1)} \\ &= \frac{(p_1(\boldsymbol{\theta}))^\alpha - 1}{(p_2(\boldsymbol{\theta}))^\alpha\alpha(\alpha-1)} - \frac{(p_2(\boldsymbol{\theta}))^\alpha - 1}{(p_2(\boldsymbol{\theta}))^\alpha\alpha(\alpha-1)} \\ &= \frac{f_\alpha(p_1(\boldsymbol{\theta})) - f_\alpha(p_2(\boldsymbol{\theta}))}{(p_2(\boldsymbol{\theta}))^\alpha}. \end{aligned} \tag{161}$$

Therefore, the objective function in (22) for $f(x) = f_\alpha(x)$ can be written as

$$\sum_{k=1}^K w_k \mathcal{D}_\alpha(q_k \| \varphi) = \sum_{k=1}^K w_k \int_\Theta \varphi(\boldsymbol{\theta}) f_\alpha\left(\frac{q_k(\boldsymbol{\theta})}{\varphi(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta} \stackrel{(161)}{=} \sum_{k=1}^K w_k \int_\Theta \varphi(\boldsymbol{\theta}) \frac{f_\alpha(q_k(\boldsymbol{\theta})) - f_\alpha(\varphi(\boldsymbol{\theta}))}{(\varphi(\boldsymbol{\theta}))^\alpha} \mathrm{d}\boldsymbol{\theta}.$$

Interchanging the summation and the integral gives

$$\begin{aligned}\sum_{k=1}^K w_k \mathcal{D}_\alpha(q_k \| \varphi) &= \int_\Theta \varphi(\boldsymbol{\theta}) \sum_{k=1}^K w_k \frac{f_\alpha(q_k(\boldsymbol{\theta})) - f_\alpha(\varphi(\boldsymbol{\theta}))}{(\varphi(\boldsymbol{\theta}))^\alpha} \mathrm{d}\boldsymbol{\theta} \\ &\stackrel{(a)}{=} \int_\Theta \varphi(\boldsymbol{\theta}) \frac{\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right) - f_\alpha(\varphi(\boldsymbol{\theta}))}{(\varphi(\boldsymbol{\theta}))^\alpha} \mathrm{d}\boldsymbol{\theta} \\ &= \int_\Theta \varphi(\boldsymbol{\theta}) \frac{f_\alpha\left(f_\alpha^{-1}\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right)\right) - f_\alpha(\varphi(\boldsymbol{\theta}))}{(\varphi(\boldsymbol{\theta}))^\alpha} \mathrm{d}\boldsymbol{\theta} \\ &\stackrel{(161)}{=} \int_\Theta \varphi(\boldsymbol{\theta}) f_\alpha\left(\frac{f_\alpha^{-1}\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right)}{\varphi(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta},\end{aligned}$$

where we used in (a) that $\sum_{k=1}^K w_k = 1$. Since φ is a pdf and $f_\alpha(x) = \frac{x^\alpha - 1}{\alpha(\alpha-1)}$ is a convex function for $\alpha \in \mathbb{R} \setminus \{0, 1\}$, we can apply Jensen's inequality[9] to obtain the following lower bound on the objective function:

$$\begin{aligned}\int_\Theta \varphi(\boldsymbol{\theta}) f_\alpha\left(\frac{f_\alpha^{-1}\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right)}{\varphi(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta} &\geq f_\alpha\left(\int_\Theta \varphi(\boldsymbol{\theta}) \frac{f_\alpha^{-1}\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right)}{\varphi(\boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta}\right) \\ &= f_\alpha\left(\int_\Theta f_\alpha^{-1}\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right) \mathrm{d}\boldsymbol{\theta}\right), \end{aligned} \quad (162)$$

with equality if and only if the function

$$\zeta(\boldsymbol{\theta}) \triangleq \frac{f_\alpha^{-1}\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right)}{\varphi(\boldsymbol{\theta})}$$

is constant almost everywhere. Note that this is equivalent to $\varphi(\boldsymbol{\theta}) \propto f_\alpha^{-1}(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta})))$. Since the right-hand side of (162) is independent of φ, it is a lower bound for any choice of φ, and hence the function $\varphi(\boldsymbol{\theta})$ minimizing the objective function (which is the desired solution $q(\boldsymbol{\theta})$ in (22)) is the one for which this lower bound is achieved with equality, i.e.,

$$\begin{aligned} q(\boldsymbol{\theta}) &\propto f_\alpha^{-1}\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right) \\ &\stackrel{(160)}{=} \left(\left(\sum_{k=1}^K w_k f_\alpha(q_k(\boldsymbol{\theta}))\right) \alpha(\alpha-1) + 1\right)^{1/\alpha} \\ &= \left(\sum_{k=1}^K w_k (q_k(\boldsymbol{\theta}))^\alpha - \sum_{k=1}^K w_k + 1\right)^{1/\alpha} \\ &= \left(\sum_{k=1}^K w_k (q_k(\boldsymbol{\theta}))^\alpha\right)^{1/\alpha}. \end{aligned}$$

We conclude that the solution to (22) when $f(x) = f_\alpha(x)$ is $q(\boldsymbol{\theta}) = c\big(\sum_{k=1}^K w_k (q_k(\boldsymbol{\theta}))^\alpha\big)^{1/\alpha}$, where $c = 1/\int_\Theta \big(\sum_{k=1}^K w_k (q_k(\boldsymbol{\theta}))^\alpha\big)^{1/\alpha} \mathrm{d}\boldsymbol{\theta}$.

[9] Jensen's inequality [198, Th. 3.3] asserts that for a pdf $\varphi(\cdot)$, a measurable function $\zeta(\cdot)$, and a convex function $\psi(\cdot)$ we have that $\int \psi(\zeta(\boldsymbol{\theta}))\varphi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \geq \psi\big(\int \zeta(\boldsymbol{\theta})\varphi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\big)$, with equality if and only if the function ζ is constant almost everywhere.

Appendix H
Characterization of the Reverse α-divergence

We will show that $\mathcal{D}_\alpha(\varphi \| q_k) = \mathcal{D}_{\alpha^*}(q_k \| \varphi)$, where $\alpha^* = 1 - \alpha$. To this end, we will use (27) with $f(x) = f_\alpha(x) = \frac{x^\alpha - 1}{\alpha(\alpha-1)}$. By $f^*(x) = x f(1/x)$, we have

$$\begin{aligned} f_\alpha^*(x) &= x \frac{x^{-\alpha} - 1}{\alpha(\alpha-1)} \\ &= \frac{x^{-\alpha+1} - x}{\alpha(\alpha-1)} \\ &= \frac{x^{-(\alpha-1)} - 1}{\alpha(\alpha-1)} - \frac{1}{\alpha(\alpha-1)}(x-1) \\ &= f_{\alpha^*}(x) - \frac{1}{\alpha(\alpha-1)}(x-1). \end{aligned}$$

Thus, up to the additive term $-\frac{1}{\alpha(\alpha-1)}(x-1)$, the function $f_\alpha^*(x)$ is equal to $f_{\alpha^*}(x)$. Now, by [164, Prop. 1], an f-divergence does not change if $f(x)$ is replaced by $f(x) + c(x-1)$ for an arbitrary $c \in \mathbb{R}$. Hence, f_α^* and f_{α^*} result in the same f-divergence, and (27) together with (24) implies

$$\begin{aligned} \mathcal{D}_\alpha(\varphi \| q_k) &= \mathcal{D}_{f_\alpha}(\varphi \| q_k) \\ &= \mathcal{D}_{f_\alpha^*}(q_k \| \varphi) \\ &= \mathcal{D}_{f_{\alpha^*}}(q_k \| \varphi) \\ &= \mathcal{D}_{\alpha^*}(q_k \| \varphi). \end{aligned}$$

Appendix I
Proof of Theorem 16 (Constrained Minimization of the Weighted Average of Squared L_2 Distances)

We want to find

$$q = \arg\min_{\varphi \in \mathcal{P}} \sum_{k=1}^K w_k \|q_k - \varphi\|_2^2. \quad (163)$$

To this end, we note that

$$\begin{aligned} \min_{\varphi \in \mathcal{P}} \sum_{k=1}^K w_k \|q_k - \varphi\|_2^2 &= \min_{\varphi \in \mathcal{P}} \int_\Theta \sum_{k=1}^K w_k \big(q_k(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta})\big)^2 \mathrm{d}\boldsymbol{\theta} \\ &\geq \int_\Theta \min_{\varphi(\boldsymbol{\theta}) \geq 0} \left\{\sum_{k=1}^K w_k \big(q_k(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta})\big)^2\right\} \mathrm{d}\boldsymbol{\theta}. \end{aligned} \quad (164)$$

For each fixed $\boldsymbol{\theta}$, the function value $\varphi(\boldsymbol{\theta})$ that achieves the minimum $\min_{\varphi(\boldsymbol{\theta})\geq 0}\sum_{k=1}^{K} w_k\big(q_k(\boldsymbol{\theta})-\varphi(\boldsymbol{\theta})\big)^2$ is easily seen to be

$$\varphi^*(\boldsymbol{\theta})=\sum_{k=1}^{K} w_k q_k(\boldsymbol{\theta})\,.$$

Because $\varphi^*\in\mathcal{P}$ (due to $(w_1,\ldots,w_K)\in\mathcal{S}_K$), we have that

$$\begin{aligned}\sum_{k=1}^{K} w_k\|q_k-\varphi^*\|_2^2 &\geq \min_{\varphi\in\mathcal{P}}\sum_{k=1}^{K} w_k\|q_k-\varphi\|_2^2\\ &\overset{(164)}{\geq}\int_{\Theta}\sum_{k=1}^{K} w_k\big(q_k(\boldsymbol{\theta})-\varphi^*(\boldsymbol{\theta})\big)^2\mathrm{d}\boldsymbol{\theta}\\ &=\sum_{k=1}^{K} w_k\|q_k-\varphi^*\|_2^2. \end{aligned}\qquad(165)$$

Thus, all inequalities in (165) are actually equalities. In particular,

$$\sum_{k=1}^{K} w_k\|q_k-\varphi^*\|_2^2=\min_{\varphi\in\mathcal{P}}\sum_{k=1}^{K} w_k\|q_k-\varphi\|_2^2,$$

i.e., $q=\varphi^*$ solves (163).

Appendix J
Proof of Theorem 17 (Unconstrained Minimization of the Weighted Average of General Distances)

Let $\chi_\varphi(\boldsymbol{\theta})\triangleq\chi(\varphi(\boldsymbol{\theta}))$ and $\chi_{q_k}(\boldsymbol{\theta})\triangleq\chi(q_k(\boldsymbol{\theta}))$. We want to find

$$\tilde{q}=\arg\min_{\varphi}\sum_{k=1}^{K} w_k\|\chi_{q_k}-\chi_\varphi\|_2^2.\qquad(166)$$

To this end, we first derive

$$\chi^*=\arg\min_{\chi}\sum_{k=1}^{K} w_k\|\chi_{q_k}-\chi\|_2^2.\qquad(167)$$

Following the same steps as in Appendix I with q_k replaced by χ_{q_k} and φ replaced by χ, it is easy to see that

$$\chi^*(\boldsymbol{\theta})=\sum_{k=1}^{K} w_k\chi(q_k(\boldsymbol{\theta}))\,.\qquad(168)$$

Because $\chi(q_k(\boldsymbol{\theta}))\in(a,b)$, the convex combination $\sum_{k=1}^{K} w_k\chi(q_k(\boldsymbol{\theta}))$ is again in (a,b). Thus, $\chi^*(\boldsymbol{\theta})$ is in the range of χ and we can define

$$\varphi^*(\boldsymbol{\theta})\triangleq\chi^{-1}\big(\chi^*(\boldsymbol{\theta})\big).\qquad(169)$$

This implies

$$\chi_{\varphi^*}(\boldsymbol{\theta})=\chi(\varphi^*(\boldsymbol{\theta}))=\chi^*(\boldsymbol{\theta})\,.\qquad(170)$$

We claim that $\tilde{q}$ defined in (166) equals φ^*. Indeed, we have for any φ

$$\begin{aligned}\sum_{k=1}^{K} w_k\|\chi_{q_k}-\chi_\varphi\|_2^2 &\geq \min_{\chi}\sum_{k=1}^{K} w_k\|\chi_{q_k}-\chi\|_2^2\\ &\overset{(167)}{=}\sum_{k=1}^{K} w_k\|\chi_{q_k}-\chi^*\|_2^2\\ &\overset{(170)}{=}\sum_{k=1}^{K} w_k\|\chi_{q_k}-\chi_{\varphi^*}\|_2^2,\end{aligned}$$

from which we conclude that φ^* achieves the minimum in (166) and thus equals $\tilde{q}$. We then obtain the optimal nonnormalized pooling function as

$$\tilde{q}(\boldsymbol{\theta})=\varphi^*(\boldsymbol{\theta})\overset{(169)}{=}\chi^{-1}\big(\chi^*(\boldsymbol{\theta})\big)\overset{(168)}{=}\chi^{-1}\Bigg(\sum_{k=1}^{K} w_k\chi(q_k(\boldsymbol{\theta}))\Bigg).$$

Appendix K
Proofs of the Fusion Rule for a Scalar Parameter

A. Proof of Theorem 21

For $d_\theta=1$, the local observation likelihood functions from (71) are given by

$$\ell_k(\theta)\propto\exp\bigg(-\frac{\theta^2\mathbf{h}_k^{\intercal}\boldsymbol{\Sigma}_{kk}^{-1}\mathbf{h}_k}{2}+\theta\mathbf{h}_k^{\intercal}\boldsymbol{\Sigma}_{kk}^{-1}\mathbf{h}_k t_k\bigg),\qquad(171)$$

where $t_k=\mathbf{v}_k^{\intercal}\mathbf{y}_k=\mathbf{h}_k^{\intercal}\boldsymbol{\Sigma}_{kk}^{-1}\mathbf{y}_k/(\mathbf{h}_k^{\intercal}\boldsymbol{\Sigma}_{kk}^{-1}\mathbf{h}_k)$ according to (82) and (83). Furthermore, the global likelihood function (81) can be rewritten as

$$\begin{aligned}\lambda(\theta)&\propto\exp\bigg(-\frac{(\mathbf{t}-\mathbf{1}_K\theta)^{\intercal}\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{t}-\mathbf{1}_K\theta)}{2}\bigg)\\ &\propto\exp\bigg(-\frac{\theta^2}{2\widehat{\sigma}^2}+\theta\mathbf{1}_K^{\intercal}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{t}\bigg),\end{aligned}\qquad(172)$$

where $\widehat{\sigma}^2=1/(\mathbf{1}_K^{\intercal}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_K)$. The relation (84) follows from

$$\begin{aligned}&\prod_{k=1}^{K}(\ell_k(\theta))^{w_k}\\ &\overset{(171)}{\propto}\exp\Bigg(\sum_{k=1}^{K} w_k\bigg(-\frac{\theta^2\mathbf{h}_k^{\intercal}\boldsymbol{\Sigma}_{kk}^{-1}\mathbf{h}_k}{2}+\theta\mathbf{h}_k^{\intercal}\boldsymbol{\Sigma}_{kk}^{-1}\mathbf{h}_k t_k\bigg)\Bigg)\\ &\overset{(85)}{=}\exp\bigg(-\frac{\sum_{k=1}^{K}\theta^2\mathbf{1}_K^{\intercal}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{e}_k}{2}+\sum_{k=1}^{K}\theta\mathbf{1}_K^{\intercal}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{e}_k t_k\bigg)\\ &\overset{(a)}{=}\exp\bigg(-\frac{\theta^2}{2\widehat{\sigma}^2}+\theta\mathbf{1}_K^{\intercal}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{t}\bigg)\\ &\overset{(172)}{\propto}\lambda(\theta)\,,\end{aligned}$$

where we used in (a) that $\sum_{k=1}^{K}\mathbf{e}_k=\mathbf{1}_K$ and $\sum_{k=1}^{K}\mathbf{e}_k t_k=\mathbf{t}$. Finally, the fusion rule for the posteriors in (86) easily follows from (84):

$$\begin{aligned}p(\theta\,|\,\mathbf{t})&\propto p(\theta)\lambda(\theta)\\ &\overset{(84)}{\propto}p(\theta)\prod_{k=1}^{K}(\ell_k(\theta))^{w_k}\\ &\propto p(\theta)\prod_{k=1}^{K}\bigg(\frac{\pi_k(\theta)}{p(\theta)}\bigg)^{w_k}\\ &=(p(\theta))^{1-\sum_{k=1}^{K}w_k}\prod_{k=1}^{K}(\pi_k(\theta))^{w_k}\,.\end{aligned}$$

B. Calculation of the Weights in Example 3

We will show expression (92) for w_k. The vectors $\mathbf{v}_k^\intercal$ in (83) are given as

$$\mathbf{v}_k^\intercal = \frac{1}{\mathbf{1}_{r_0+r_k}^\intercal \mathbf{1}_{r_0+r_k}} \mathbf{1}_{r_0+r_k}^\intercal = \frac{1}{r_0+r_k}\mathbf{1}_{r_0+r_k}^\intercal,$$

and, in turn, the matrix $\widetilde{\mathbf{\Sigma}}$ in (79) is given by the entries

$$\begin{aligned}\widetilde{\Sigma}_{kk'} &= \mathbf{v}_k^\intercal \mathbf{\Sigma}_{kk'} \mathbf{v}_{k'} \\ &= \frac{1}{r_0+r_k}\mathbf{1}_{r_0+r_k}^\intercal \mathbf{\Sigma}_{kk'} \frac{1}{r_0+r_{k'}}\mathbf{1}_{r_0+r_{k'}} \\ &= \frac{r_0}{(r_0+r_k)(r_0+r_{k'})}\end{aligned}$$

for $k \neq k'$ and

$$\widetilde{\Sigma}_{kk} = \frac{1}{r_0+r_k}.$$

It is easily verified that we can rewrite $\widetilde{\mathbf{\Sigma}}$ as the following sum of a diagonal matrix and a rank one matrix

$$\widetilde{\mathbf{\Sigma}} = \begin{pmatrix} \frac{r_1}{(r_0+r_1)^2} & & \\ & \ddots & \\ & & \frac{r_K}{(r_0+r_K)^2}\end{pmatrix} + \begin{pmatrix}\frac{1}{r_0+r_1} \\ \vdots \\ \frac{1}{r_0+r_K}\end{pmatrix} r_0 \begin{pmatrix}\frac{1}{r_0+r_1} & \cdots & \frac{1}{r_0+r_K}\end{pmatrix}.$$

By the matrix inversion lemma [199, eq. (0.7.4.2)], we can hence calculate $\widetilde{\mathbf{\Sigma}}^{-1}$ as

$$\widetilde{\mathbf{\Sigma}}^{-1} = \begin{pmatrix} \frac{(r_0+r_1)^2}{r_1} & & \\ & \ddots & \\ & & \frac{(r_0+r_K)^2}{r_K}\end{pmatrix} - \left(\sum_{k=0}^{K}\frac{1}{r_k}\right)^{-1} \begin{pmatrix}\frac{r_0+r_1}{r_1} \\ \vdots \\ \frac{r_0+r_K}{r_K}\end{pmatrix} \begin{pmatrix}\frac{r_0+r_1}{r_1} & \cdots & \frac{r_0+r_K}{r_K}\end{pmatrix}.$$

To calculate the weights w_k in (85), we have to sum over the kth column of $\widetilde{\mathbf{\Sigma}}^{-1}$ and divide by $\mathbf{1}_{r_0+r_k}^\intercal \mathbf{1}_{r_0+r_k} = r_0 + r_k$, i.e.,

$$\begin{aligned} w_k &= \frac{1}{r_0+r_k}\left(\frac{(r_0+r_k)^2}{r_k} - \frac{\sum_{k'=1}^{K}\frac{(r_0+r_k)(r_0+r_{k'})}{r_k r_{k'}}}{\sum_{k'=0}^{K}\frac{1}{r_{k'}}}\right) \\ &= \frac{r_0+r_k}{r_k} - \frac{\sum_{k'=1}^{K}\frac{r_0+r_{k'}}{r_k r_{k'}}}{\sum_{k'=0}^{K}\frac{1}{r_{k'}}} \\ &= \frac{r_0+r_k}{r_k} - \frac{\frac{K}{r_k} + \frac{r_0}{r_k}\sum_{k'=1}^{K}\frac{1}{r_{k'}}}{\sum_{k'=0}^{K}\frac{1}{r_{k'}}} \\ &= \frac{r_0+r_k}{r_k} - \frac{\frac{K-1}{r_k} + \frac{r_0}{r_k}\sum_{k'=0}^{K}\frac{1}{r_{k'}}}{\sum_{k'=0}^{K}\frac{1}{r_{k'}}} \\ &= \frac{r_0}{r_k} + 1 - \frac{\frac{K-1}{r_k}}{\sum_{k'=0}^{K}\frac{1}{r_{k'}}} - \frac{r_0}{r_k} \\ &= 1 - \frac{K-1}{r_k}\left(\sum_{k'=0}^{K}\frac{1}{r_{k'}}\right)^{-1}. \end{aligned}$$

Appendix L
Proofs of the Fusion Rule for a Vector Parameter

A. Proof of Theorem 23

We can rewrite (81) as

$$\begin{aligned}\lambda(\boldsymbol{\theta}) &\propto \exp\left(-\frac{((\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})\boldsymbol{\theta} - \mathbf{t})^\intercal \widetilde{\mathbf{\Sigma}}^{-1}((\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})\boldsymbol{\theta} - \mathbf{t})}{2}\right) \\ &\propto \exp\left(-\frac{\boldsymbol{\theta}^\intercal(\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\mathbf{\Sigma}}^{-1}(\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})\boldsymbol{\theta}}{2} + \boldsymbol{\theta}^\intercal(\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\mathbf{\Sigma}}^{-1}\mathbf{t}\right) \\ &\overset{(99)}{=} \exp\left(-\frac{\boldsymbol{\theta}^\intercal \widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\theta}}{2} + \boldsymbol{\theta}^\intercal(\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\mathbf{\Sigma}}^{-1}\mathbf{t}\right).\end{aligned} \tag{173}$$

Furthermore, from (71), we see that

$$\ell_k(\boldsymbol{\theta}) \propto \exp\left(-\frac{\boldsymbol{\theta}^\intercal \mathbf{H}_k^\intercal \mathbf{\Sigma}_{kk}^{-1}\mathbf{H}_k\boldsymbol{\theta}}{2} + \boldsymbol{\theta}^\intercal \mathbf{H}_k^\intercal \mathbf{\Sigma}_{kk}^{-1}\mathbf{H}_k\mathbf{t}_k\right),$$

where $\mathbf{t}_k = \mathbf{V}_k\mathbf{y}_k$. Thus, we have

$$\begin{aligned}\prod_{k=1}^{K}\ell_k(\mathbf{W}_k\boldsymbol{\theta}) &\propto \exp\left(-\frac{\boldsymbol{\theta}^\intercal\left(\sum_{k=1}^{K}\mathbf{W}_k^\intercal\mathbf{H}_k^\intercal\mathbf{\Sigma}_{kk}^{-1}\mathbf{H}_k\mathbf{W}_k\right)\boldsymbol{\theta}}{2} + \boldsymbol{\theta}^\intercal\sum_{k=1}^{K}\mathbf{W}_k^\intercal\mathbf{H}_k^\intercal\mathbf{\Sigma}_{kk}^{-1}\mathbf{H}_k\mathbf{t}_k\right) \\ &\overset{(98)}{=} \exp\left(-\frac{\boldsymbol{\theta}^\intercal\left(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{G}\right)\boldsymbol{\theta}}{2} + \boldsymbol{\theta}^\intercal\sum_{k=1}^{K}\mathbf{W}_k^\intercal\mathbf{H}_k^\intercal\mathbf{\Sigma}_{kk}^{-1}\mathbf{H}_k\mathbf{t}_k\right) \\ &= \frac{1}{\xi_0(\boldsymbol{\theta})}\exp\left(-\frac{\boldsymbol{\theta}^\intercal\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\theta}}{2} + \boldsymbol{\theta}^\intercal\sum_{k=1}^{K}\mathbf{W}_k^\intercal\mathbf{H}_k^\intercal\mathbf{\Sigma}_{kk}^{-1}\mathbf{H}_k\mathbf{t}_k\right),\end{aligned} \tag{174}$$

with $\xi_0(\boldsymbol{\theta})$ as defined in (97). By comparing (173) and (174), we see that (95) holds, provided that

$$(\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\mathbf{\Sigma}}^{-1}\mathbf{t} = \sum_{k=1}^{K}\mathbf{W}_k^\intercal\mathbf{H}_k^\intercal\mathbf{\Sigma}_{kk}^{-1}\mathbf{H}_k\mathbf{t}_k. \tag{175}$$

Inserting (96) into the right-hand side of (175), we obtain

$$\begin{aligned}\sum_{k=1}^{K}\mathbf{W}_k^\intercal\mathbf{H}_k^\intercal\mathbf{\Sigma}_{kk}^{-1}\mathbf{H}_k\mathbf{t}_k &= (\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\mathbf{\Sigma}}^{-1}\sum_{k=1}^{K}(\mathbf{e}_k \otimes \mathbf{I}_{d_\theta})\mathbf{t}_k \\ &= (\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\mathbf{\Sigma}}^{-1}\mathbf{t},\end{aligned}$$

concluding the proof of (95).

Finally, the fusion rule (100) easily follows from (95):

$$\begin{aligned}p(\boldsymbol{\theta}\,|\,\mathbf{t}) &\propto p(\boldsymbol{\theta})\lambda(\boldsymbol{\theta}) \\ &\propto p(\boldsymbol{\theta})\xi_0(\boldsymbol{\theta})\prod_{k=1}^{K}\ell_k(\mathbf{W}_k\boldsymbol{\theta}) \\ &\propto p(\boldsymbol{\theta})\xi_0(\boldsymbol{\theta})\prod_{k=1}^{K}\frac{\pi_k(\mathbf{W}_k\boldsymbol{\theta})}{p(\mathbf{W}_k\boldsymbol{\theta})}.\end{aligned}$$

B. Proof of Corollary 24

We start directly from $p(\boldsymbol{\theta}\,|\,\mathbf{t}) \propto p(\boldsymbol{\theta})\lambda(\boldsymbol{\theta})$. By (173) and our choice of prior $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \propto \exp\big(-\frac{\boldsymbol{\theta}^\intercal \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}}{2} + \boldsymbol{\theta}^\intercal \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\big)$, we have that

$$
\begin{aligned}
p(\boldsymbol{\theta})\lambda(\boldsymbol{\theta}) &\propto \exp\Bigg(-\frac{\boldsymbol{\theta}^\intercal \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\theta} + \boldsymbol{\theta}^\intercal \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}}{2} \\
&\qquad + \boldsymbol{\theta}^\intercal (\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{t} + \boldsymbol{\theta}^\intercal \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\Bigg) \\
&= \exp\Bigg(-\frac{\boldsymbol{\theta}^\intercal \big(\widehat{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}_0^{-1}\big)\boldsymbol{\theta}}{2} \\
&\qquad + \boldsymbol{\theta}^\intercal \big((\mathbf{1}_K \otimes \mathbf{I}_{d_\theta})^\intercal \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{t} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\big)\Bigg) \\
&= \exp\Bigg(-\frac{\boldsymbol{\theta}^\intercal \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\theta}}{2} + \boldsymbol{\theta}^\intercal \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1\Bigg) \\
&\propto \exp\Bigg(-\frac{(\boldsymbol{\theta} - \boldsymbol{\mu}_1)^\intercal \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_1)}{2}\Bigg), \qquad (176)
\end{aligned}
$$

with $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ given by (101) and (102), respectively. Expression (176) is proportional to the pdf of a Gaussian with mean $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}_1$.

References

[1] S. Bandyopadhyay and S.-J. Chung, "Distributed Bayesian filtering using logarithmic opinion pool for dynamic sensor networks," *Automatica*, vol. 97, pp. 7–17, 2018.

[2] D. Clark, S. Julier, R. Mahler, and B. Ristić, "Robust multi-object sensor fusion with unknown correlations," in *Proc. Sens. Signal Process. Def. (SSPD 2010)*, London, UK, 2010.

[3] J. Hu, L. Xie, and C. Zhang, "Diffusion Kalman filtering based on covariance intersection," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 891–902, 2011.

[4] W. Maddern and P. Newman, "Real-time probabilistic fusion of sparse 3d lidar and dense stereo," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS 2016)*, Daejeon, South Korea, 2016, pp. 2181–2188.

[5] K. Da, T. Li, Y. Zhu, H. Fan, and Q. Fu, "Recent advances in multisensor multitarget tracking using random finite set," *Front. Inform. Technol. Electron. Eng.*, vol. 22, pp. 5–24, 2021.

[6] C. Fantacci, B.-N. Vo, B.-T. Vo, G. Battistelli, and L. Chisci, "Robust fusion for multisensor multiobject tracking," *IEEE Signal Process. Letters*, vol. 25, no. 5, pp. 640–644, 2018.

[7] F. Meyer, O. Hlinka, H. Wymeersch, E. Riegler, and F. Hlawatsch, "Distributed localization and tracking of mobile networks including noncooperative objects," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 1, pp. 57–71, 2015.

[8] M. Üney, D. E. Clark, and S. J. Julier, "Distributed fusion of PHD filters via exponential mixture densities," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 521–531, 2013.

[9] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Proc. Adv. Neural Inf. Process. Syst. (NIPS 2017)*, vol. 30, 2017.

[10] H. Lee, S. Hong, and E. Kim, "Neural network ensemble with probabilistic fusion and its application to gait recognition," *Neurocomputing*, vol. 72, no. 7-9, pp. 1557–1564, 2009.

[11] Q. Lu, G. Karanikolas, Y. Shen, and G. B. Giannakis, "Ensemble Gaussian processes with spectral features for online interactive learning with scalability," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 1910–1920.

[12] A. T. Thorgeirsson and F. Gauterin, "Probabilistic predictions with federated learning," *Entropy*, vol. 23, no. 1, p. 41, 2021.

[13] F. Alam, R. Mehmood, I. Katib, N. N. Albogami, and A. Albeshri, "Data fusion and IoT for smart ubiquitous environments: A survey," *IEEE Access*, vol. 5, pp. 9533–9554, 2017.

[14] E. Kats, J. Goldberger, and H. Greenspan, "A soft STAPLE algorithm combined with anatomical knowledge," in *Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI 2019)*. Shenzhen, China: Springer, 2019, pp. 510–517.

[15] B. Kolosz, S. Grant-Muller, and K. Djemame, "Modelling uncertainty in the sustainability of intelligent transport systems for highways using probabilistic data fusion," *Environ. Model. Softw.*, vol. 49, pp. 78–97, 2013.

[16] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Comput. Electron. Agr.*, vol. 151, pp. 61–69, 2018.

[17] G.-B. Li, L.-L. Yang, Y. Xu, W.-J. Wang, L.-L. Li, and S.-Y. Yang, "A combined molecular docking-based and pharmacophore-based target prediction strategy with a probabilistic fusion method for target ranking," *J. Mol. Graph. Model.*, vol. 44, pp. 278–285, 2013.

[18] A. H. Murphy and R. L. Winkler, "Probability forecasting in meteorology," *J. Am. Stat. Assoc.*, vol. 79, no. 387, pp. 489–500, 1984.

[19] R. Marty, V. Fortin, H. Kuswanto, A.-C. Favre, and E. Parent, "Combining the Bayesian processor of output with Bayesian model averaging for reliable ensemble forecasting," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 64, no. 1, pp. 75–92, 2015.

[20] J. Mitchell and S. G. Hall, "Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation," *Oxf. Bull. Econ. Stat.*, vol. 67, pp. 995–1033, 2005.

[21] E. Moral-Benito, "Model averaging in economics: An overview," *J. Econ. Surv.*, vol. 29, no. 1, pp. 46–75, 2015.

[22] S. Barak, A. Arjmand, and S. Ortobelli, "Fusion of multiple diverse predictors in stock market," *Inform. Fusion*, vol. 36, pp. 90–102, 2017.

[23] C. Genest and J. V. Zidek, "Combining probability distributions: A critique and an annotated bibliography," *Stat. Sci.*, vol. 1, no. 1, pp. 114–135, 1986.

[24] S. G. Hall and J. Mitchell, "Combining density forecasts," *Int. J. Forecast.*, vol. 23, no. 1, pp. 1–13, 2007.

[25] T. Gneiting, "Editorial: Probabilistic forecasting," *J. R. Stat. Soc. Ser. A Stat. Soc.*, pp. 319–321, 2008.

[26] M. Stone, "The opinion pool," *Ann. Math. Statist.*, vol. 32, no. 4, pp. 1339–1342, 1961.

[27] T. Bailey, S. Julier, and G. Agamennoni, "On conservative fusion of information with unknown non-Gaussian dependence," in *Proc. Int. Conf. Inf. Fusion (FUSION 2012)*, Singapore, Singapore, 2012.

[28] C. Genest, "A characterization theorem for externally Bayesian groups," *Ann. Statist.*, vol. 12, no. 3, pp. 1100–1105, 1984.

[29] M. B. Hurley, "An information theoretic justification for covariance intersection and its generalization," in *Proc. Int. Conf. Inf. Fusion (FUSION 2002)*, Annapolis, MD, 2002.

[30] N. Lehrer, O. Tslil, and A. Carmi, "Log-linear Chernoff fusion for distributed particle filtering," in *Proc. Int. Conf. Inf. Fusion (FUSION 2019)*, Ottawa, ON, Canada, 2019.

[31] M. Gunay, U. Orguner, and M. Demirekler, "Chernoff fusion of Gaussian mixtures based on sigma-point approximation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 6, pp. 2732–2746, 2016.

[32] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. American Control Conf.*, Albuquerque, NM, 1997.

[33] K.-C. Chang, C.-Y. Chong, and S. Mori, "Analytical and computational evaluation of scalable distributed fusion algorithms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 4, pp. 2022–2034, 2010.

[34] I. Urteaga, M. F. Bugallo, and P. M. Djurić, "Sequential Monte Carlo methods under model uncertainty," in *2016 IEEE Stat. Signal Process. Workshop (SSP)*, Palma de Mallorca, Spain, 2016.

[35] O. Hlinka, F. Hlawatsch, and P. M. Djurić, "Consensus-based distributed particle filtering with distributed proposal adaptation," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3029–3041, 2014.

[36] V. Savic, H. Wymeersch, and S. Zazo, "Belief consensus algorithms for fast distributed target tracking in wireless sensor networks," *Signal Processing*, vol. 95, pp. 149–160, 2014.

[37] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

[38] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, 2008.

[39] O. Punska, "Bayesian approaches to multi-sensor data fusion," Master's thesis, University of Cambridge, Cambridge, UK, 1999.

[40] G. Battistelli and L. Chisci, "Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, Mar. 2014.

[41] Z. Deng, P. Zhang, W. Qi, J. Liu, and Y. Gao, "Sequential covariance intersection fusion Kalman filter," *Inf. Sci.*, vol. 189, pp. 293–309, 2012.

[42] M. Tang, Y. Rong, J. Zhou, and X. R. Li, "Information geometric approach to multisensor estimation fusion," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 279–292, 2018.

[43] G. Li, G. Battistelli, L. Chisci, W. Yi, and L. Kong, "Distributed multi-view multi-target tracking based on CPHD filtering," *Signal Process.*, vol. 188, p. 108210, 2021.

[44] T. Li, X. Wang, Y. Liang, and Q. Pan, "On arithmetic average fusion and its application for distributed multi-Bernoulli multitarget tracking," *IEEE Trans. Signal Process.*, vol. 68, pp. 2883–2896, 2020.

[45] T. Li and F. Hlawatsch, "A distributed particle-PHD filter using arithmetic-average fusion of Gaussian mixture parameters," *Inform. Fusion*, vol. 73, pp. 111–124, 2021.

[46] W. Yi, G. Li, and G. Battistelli, "Distributed multi-sensor fusion of PHD filters with different sensor fields of view," *IEEE Trans. Signal Process.*, vol. 68, pp. 5204–5218, 2020.

[47] L. Gao, G. Battistelli, and L. Chisci, "Fusion of labeled RFS densities with minimum information loss," *IEEE Trans. Signal Process.*, vol. 68, pp. 5855–5868, 2020.

[48] A. K Gostar, T. Rathnayake, R. Tennakoon, A. Bab-Hadiashar, G. Battistelli, L. Chisci, and R. Hoseinnezhad, "Cooperative sensor fusion in centralized sensor networks using Cauchy-Schwarz divergence," *Signal Process.*, vol. 167, p. 107278, 2020.

[49] S. Li, G. Battistelli, L. Chisci, W. Yi, B. Wang, and L. Kong, "Computationally efficient multi-agent multi-object tracking with labeled random finite sets," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 260–275, 2019.

[50] R. P. S. Mahler, "Optimal/robust distributed data fusion: A unified approach," in *Aerosense 2000: Signal and Image Processing (Proceedings of SPIE)*, vol. 4052, Orlando, FL, USA, 2000.

[51] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020.

[52] M. Yurochkin, M. Agarwal, S. Ghosh, K. Greenewald, N. Hoang, and Y. Khazaeni, "Bayesian nonparametric federated learning of neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML 2019)*, Long Beach, CA, 2019.

[53] R. T. Clemen, "Combining forecasts: A review and annotated bibliography," *Int. J. Forecast.*, vol. 5, no. 4, pp. 559–583, 1989.

[54] J. S. Armstrong, "Combining forecasts," in *Principles of Forecasting*, J. S. Armstrong, Ed. Boston, MA: Springer, 2001, pp. 417–439.

[55] K. F. Wallis, "Combining density and interval forecasts: A modest proposal," *Oxf. Bull. Econ. Stat.*, vol. 67, no. s1, pp. 983–994, 2005.

[56] R. L. Winkler, Y. Grushka-Cockayne, K. C. Lichtendahl Jr, and V. R. R. Jose, "Probability forecasts and their combination: A research perspective," *Decis. Anal.*, vol. 16, no. 4, pp. 239–260, 2019.

[57] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: A tutorial," *Stat. Sci.*, vol. 14, no. 4, pp. 382–401, 1999.

[58] T. M. Fragoso, W. Bertoli, and F. Louzada, "Bayesian model averaging: A systematic review and conceptual classification," *Int. Stat. Rev.*, vol. 86, no. 1, pp. 1–28, 2018.

[59] D. Posada, "jModelTest: Phylogenetic model averaging," *Mol. Biol. Evol.*, vol. 25, no. 7, pp. 1253–1256, 2008.

[60] D. Darriba, G. L. Taboada, R. Doallo, and D. Posada, "jModelTest 2: More models, new heuristics and parallel computing," *Nat. Methods*, vol. 9, no. 8, pp. 772–772, 2012.

[61] M. F. J. Steel, "Model averaging and its use in economics," *J. Econ. Lit.*, vol. 58, no. 3, pp. 644–719, 2020.

[62] C. F. Dormann, J. M. Calabrese, G. Guillera-Arroita, E. Matechou, V. Bahn, K. Bartoń, C. M. Beale, S. Ciuti, J. Elith, K. Gerstner, J. Guelat, P. Keil, J. J. Lahoz-Monfort, L. J. Pollock, B. Reineking, D. R. Roberts, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, S. N. Wood, R. O. Wüest, and F. Hartig, "Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference," *Ecol. Monogr.*, vol. 88, no. 4, pp. 485–504, 2018.

[63] F. E. Turkheimer, R. Hinz, and V. J. Cunningham, "On the undecidability among kinetic models: From model selection to model averaging," *J. Cereb. Blood Flow Metab.*, vol. 23, no. 4, pp. 490–498, 2003.

[64] J. M. Montgomery and B. Nyhan, "Bayesian model averaging: Theoretical developments and practical applications," *Political Anal.*, vol. 18, no. 2, pp. 245–270, 2010.

[65] W. Neiswanger, C. Wang, and E. P. Xing, "Asymptotically exact, embarrassingly parallel MCMC," in *Proc. Conf. Uncertain. Artif. Intell. (UAI 2014)*, Arlington, VA, 2014.

[66] X. Wang and D. B. Dunson, "Parallelizing MCMC via Weierstrass sampler," *arXiv preprint arXiv:1312.4605*, 2013.

[67] R. Bardenet, A. Doucet, and C. Holmes, "On Markov chain Monte Carlo methods for tall data," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 1515–1557, 2017.

[68] J. Rabin, G. Peyré, J. Delon, and M. Bernot, "Wasserstein barycenter and its application to texture mixing," in *Proc. Int. Conf. Scale Space Var. Methods Comput. Vis. (SSVM 2011)*, Ein-Gedi, Israel, 2011.

[69] S. Srivastava, V. Cevher, Q. Dinh, and D. Dunson, "WASP: Scalable Bayes via barycenters of subset posteriors," in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS 2015)*, San Diego, CA, 2015.

[70] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.

[71] L. Breiman, "Stacked regressions," *Mach. Learn.*, vol. 24, no. 1, pp. 49–64, 1996.

[72] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. V. Le, G. E. Hinton, and J. Dean, "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," in *Proc. Int. Conf. Learn. Represent. (ICLR 2017)*, Toulon, France, 2017.

[73] M. Hoang, N. Hoang, B. K. H. Low, and C. Kingsford, "Collective model fusion for multiple black-box experts," in *Proc. Int. Conf. Mach. Learn. (ICML 2019)*, Long Beach, CA, 2019.

[74] H. Liu, Y. Ong, X. Shen, and J. Cai, "When Gaussian process meets big data: A review of scalable GPs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4405–4423, 2020.

[75] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inform. Fusion*, vol. 16, pp. 3–17, 2014.

[76] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York, NY: Springer, 2002.

[77] K. F. Wallis, "Combining forecasts–forty years later," *Appl. Financial Econ.*, vol. 21, no. 1–2, pp. 33–41, 2011.

[78] R. M. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York, NY: Oxford University Press, 1991.

[79] R. T. Clemen and R. L. Winkler, "Combining probability distributions from experts in risk analysis," *Risk Anal.*, vol. 19, no. 2, pp. 187–203, 1999.

[80] F. Dietrich and C. List, "Probabilistic opinion pooling," in *The Oxford Handbook of Probability and Philosophy*. Oxford, UK: Oxford University Press, 2016.

[81] R. T. Stewart and I. O. Quintana, "Probabilistic opinion pooling with imprecise probabilities," *J. Philos. Log.*, vol. 47, no. 1, pp. 17–45, 2018.

[82] ——, "Learning and pooling, pooling and learning," *Erkenntnis*, vol. 83, no. 3, pp. 369–389, 2018.

[83] A. E. Abbas, "A Kullback-Leibler view of linear and log-linear pools," *Decis. Anal.*, vol. 6, no. 1, pp. 25–37, 2009.

[84] A. Garg, T. S. Jayram, S. Vaithyanathan, and H. Zhu, "Generalized opinion pooling," in *Proc. Int. Symp. Artif. Intell. Math. (ISAIM 2004)*, Fort Lauderdale, FL, 2004.

[85] K. Da, T. Li, Y. Zhu, H. Fan, and Q. Fu, "Kullback-Leibler averaging for multitarget density fusion," in *Proc. Int. Symp. Distrib. Comput. Artif. Intell. (DCAI 2019)*, Ávila, Spain, 2019.

[86] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," *SIAM J. Appl. Math.*, vol. 43, no. 2, pp. 904–924, 2011.

[87] R. L. Winkler, "The consensus of subjective probability distributions," *Manag. Sci.*, vol. 15, no. 2, pp. 61–75, 1968.

[88] P. A. Morris, "Combining expert judgments: A Bayesian approach," *Manag. Sci.*, vol. 23, no. 7, pp. 679–693, 1977.

[89] R. L. Winkler, "Combining probability distributions from dependent information sources," *Manag. Sci.*, vol. 27, no. 4, pp. 479–488, 1981.

[90] D. Lindley, "Reconciliation of probability distributions," *Oper. Res.*, vol. 31, no. 5, pp. 866–880, 1983.

[91] R. T. Clemen and R. L. Winkler, "Limits for the precision and value of information from dependent sources," *Oper. Res.*, vol. 33, no. 2, pp. 427–442, 1985.

[92] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[93] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, MA, USA: Artech House, 2003.

[94] C. Chong and S. Mori, "Convex combination and covariance intersection algorithms in distributed fusion," in *Proc. Int. Conf. Inform. Fusion*, Montréal, Canada, 2001.

[95] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT: YBS Publishing, 1995.

[96] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

[97] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA: Artech House, 2007.

[98] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion*. Storrs, CT: YBS Publishing, 2011.

[99] S. Challa, M. R. Morelande, D. Mušicki, and R. J. Evans, *Fundamentals of Object Tracking*. New York, NY, USA: Cambridge University Press, 2011.

[100] R. P. S. Mahler, *Advances in Statistical Multisource-Multitarget Information Fusion*. Boston, MA, USA: Artech House, 2014.

[101] W. Koch, *Tracking and Sensor Data Fusion*. Berlin, Germany: Springer, 2016.

[102] F. Meyer, T. Kropfreiter, J. L. Williams, R. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, "Message passing algorithms for scalable multitarget tracking," *Proc. IEEE*, vol. 106, no. 2, pp. 221–259, 2018.

[103] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. New York, NY: Springer, 2003.

[104] L. M. Kaplan, Y. Bar-Shalom, and W. D. Blair, "Assignment costs for multiple sensor track-to-track association," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 2, pp. 655–677, 2008.

[105] S. Maresca, P. Braca, J. Horstmann, and R. Grasso, "Maritime surveillance using multiple high-frequency surface-wave radars," *IEEE Trans. Geosc. Remote Sens.*, vol. 52, no. 8, pp. 5056–5071, 2014.

[106] M. Üney, S. Julier, D. Clark, and B. Ristić, "Monte Carlo realisation of a distributed multi-object fusion algorithm," in *Sensor Signal Processing for Defence (SSPD 2010)*, London, UK, Sep. 2010, pp. 1–5.

[107] R. Mahler, "Toward a theoretical foundation for distributed fusion," in *Distributed Data Fusion for Network-Centric Operations*, D. Hall, C.-Y. Chong, J. Llinas, and M. Liggins II, Eds. CRC Press, 2013, ch. 8.

[108] T. Li, H. Fan, J. García, and J. M. Corchado, "Second-order statistics analysis and comparison between arithmetic and geometric average fusion: Application to multi-sensor target tracking," *Information Fusion*, vol. 51, pp. 233–243, 2019.

[109] J. Y. Yu, M. Coates, and M. Rabbat, "Distributed multi-sensor CPHD filter using pairwise gossiping," in *2016 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2016, pp. 3176–3180.

[110] T. Li, J. M. Corchado, and S. Sun, "Partial consensus and conservative fusion of Gaussian mixtures for distributed PHD fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 5, pp. 2150–2163, 2019.

[111] L. Gao, G. Battistelli, and L. Chisci, "Multiobject fusion with minimum information loss," *IEEE Signal Process. Lett.*, vol. 27, pp. 201–205, 2020.

[112] C. Fantacci, B.-N. Vo, B.-T. Vo, G. Battistelli, and L. Chisci, "Consensus labeled random finite set filtering for distributed multi-object tracking," *ArXiv*, vol. abs/1501.01579, 2015.

[113] S. Li, W. Yi, R. Hoseinnezhad, G. Battistelli, B. Wang, and L. Kong, "Robust distributed fusion with labeled random finite sets," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 278–293, 2018.

[114] T. Kropfreiter and F. Hlawatsch, "A probabilistic label association algorithm for distributed labeled multi-Bernoulli filtering," in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, Rustenburg, South Africa, 2020, pp. 1–8.

[115] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 452–459, 2015.

[116] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[117] R. Krems, "Bayesian machine learning for quantum molecular dynamics," *Physical Chemistry Chemical Physics*, vol. 21, no. 25, pp. 13 392–13 410, 2019.

[118] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, "Leveraging uncertainty information from deep neural networks for disease detection," *Scientific reports*, vol. 7, no. 1, pp. 1–14, 2017.

[119] E. Begoli, T. Bhattacharya, and D. Kusnezov, "The need for uncertainty quantification in machine-assisted medical decision making," *Nature Machine Intelligence*, vol. 1, no. 1, pp. 20–23, 2019.

[120] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian SegNet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," *arXiv preprint arXiv:1511.02680*, 2015.

[121] J. Ching and K.-K. Phoon, "Constructing site-specific multivariate probability distribution model using Bayesian machine learning," *Journal of Engineering Mechanics*, vol. 145, no. 1, p. 04018126, 2019.

[122] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift," in *Advances in Neural Information Processing Systems*, 2019, pp. 13 991–14 002.

[123] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, 2021.

[124] H. Wang and D. Yeung, "Towards bayesian deep learning: A framework and some existing methods," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 12, pp. 3395–3408, 2016.

[125] A. G. Wilson and P. Izmailov, "Bayesian deep learning and a probabilistic perspective of generalization," *arXiv preprint arXiv:2002.08791*, 2020.

[126] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[127] H. Salimbeni and M. Deisenroth, "Doubly stochastic variational inference for deep gaussian processes," in *Advances in Neural Information Processing Systems*, 2017, pp. 4588–4599.

[128] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. New York, NY: Springer, 2009.

[129] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press, 2012.

[130] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, 2018.

[131] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of artificial intelligence research*, vol. 11, pp. 169–198, 1999.

[132] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and systems magazine*, vol. 6, no. 3, pp. 21–45, 2006.

[133] S. Rothe and D. Söffker, "Comparison of different information fusion methods using ensemble selection considering benchmark data," in *2016 19th International Conference on Information Fusion (FUSION)*. IEEE, 2016, pp. 73–78.

[134] S. Savazzi, M. Nicoli, and V. Rampa, "Federated learning with cooperating devices: A consensus approach for massive IoT networks," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4641–4654, 2020.

[135] M. Mohri, G. Sivek, and A. T. Suresh, "Agnostic federated learning," in *Proc. Int. Conf. Mach. Learn. (ICML 2019)*, Long Beach, CA, 2019.

[136] A. Vehtari, A. Gelman, T. Sivula, P. Jylänki, D. Tran, S. Sahai, P. Blomstedt, J. P. Cunningham, D. Schiminovich, and C. P. Robert, "Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data." *J. Mach. Learn. Res.*, vol. 21, pp. 17–1, 2020.

[137] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*, 3rd ed. Springer, 2016.

[138] R. J. Bermowitz and E. A. Zurndorfer, "Automated guidance for predicting quantitative precipitation," *Mon. Weather Rev.*, vol. 107, no. 2, pp. 122–128, 1979.

[139] A. S. Tay and K. F. Wallis, "Density forecasting: a survey," *J. Forecast.*, vol. 19, no. 4, pp. 235–254, 2000.

[140] W. Freiberger and U. Grenander, "On the formulation of statistical meteorology," *Revue de l'Institut International de Statistique*, pp. 59–86, 1965.

[141] E. S. Epstein, "Stochastic dynamic prediction," *Tellus*, vol. 21, no. 6, pp. 739–759, 1969.

[142] T. Gneiting and R. Ranjan, "Combining predictive distributions," *Electron. J. Stat.*, vol. 7, pp. 1747–1782, 2013.

[143] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, "Using Bayesian model averaging to calibrate forecast ensembles," *Mon. Weather Rev.*, vol. 133, no. 5, pp. 1155–1174, 2005.

[144] B. Glahn, M. Peroutka, J. Wiedenfeld, J. Wagner, G. Zylstra, B. Schuknecht, and B. Jackson, "MOS uncertainty estimates in an ensemble framework," *Mon. Weather Rev.*, vol. 137, no. 1, pp. 246–268, 2009.

[145] F. Bassetti, R. Casarin, and F. Ravazzolo, "Bayesian nonparametric calibration and combination of predictive distributions," *J. Am. Stat. Assoc.*, vol. 113, no. 522, pp. 675–685, 2018.

[146] M. P. Clements and D. I. Harvey, "Combining probability forecasts," *Int. J. Forecast.*, vol. 27, no. 2, pp. 208–223, 2011.

[147] V. Zarnowitz, "The new ASA-NBER survey of forecasts by economic statisticians," *Am. Stat.*, vol. 23, no. 1, pp. 12–16, 1969.

[148] J. Tastu, P. Pinson, P.-J. Trombe, and H. Madsen, "Probabilistic forecasts of wind power generation accounting for geographically dispersed information," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 480–489, 2013.

[149] S. Baran and S. Lerch, "Mixture EMOS model for calibrating ensemble forecasts of wind speed," *Environmetrics*, vol. 27, no. 2, pp. 116–130, 2016.

[150] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecast.*, vol. 32, no. 3, pp. 914–938, 2016.
[151] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1548–1568, 2018.
[152] K. Doubleday, S. Jascourt, W. Kleiber, and B.-M. Hodge, "Probabilistic solar power forecasting using Bayesian model averaging," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 325–337, 2020.
[153] C. Genest, "Pooling operators with the marginalization property," *Can. J. Stat.*, vol. 12, no. 2, pp. 153–163, 1984.
[154] C. Genest, K. J. McConway, and M. J. Schervish, "Characterization of externally Bayesian pooling operators," *Ann. Stat.*, vol. 14, no. 2, pp. 487–501, 1986.
[155] P. S. Bullen, *Handbook of Means and Their Inequalities*. Dordrecht, The Netherlands: Springer, 2013.
[156] K. J. McConway, "Marginalization and linear opinion pools," *J. Am. Stat. Assoc.*, vol. 76, no. 374, pp. 410–414, 1981.
[157] R. Laddaga, "Lehrer and the consensus proposal," *Synthese*, vol. 36, no. 4, pp. 473–477, 1977.
[158] A. Madansky, "Externally Bayesian groups," RAND Corporation, Santa Monica, CA, Tech. Rep., 1964.
[159] M. J. Rufo, J. Martin, and C. J. Pérez, "Log-linear pool to combine prior distributions: A suggestion for a calibration-based approach," *Bayesian Anal.*, vol. 7, no. 2, pp. 411–438, 2012.
[160] C. Genest, "A conflict between two axioms for combining subjective distributions," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 46, no. 3, pp. 403–405, 1984.
[161] I. Csiszár, "Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten," *Magyer Tud. Akad. Mat. Kutató Int. Közl.*, vol. 8, pp. 85–108, 1963.
[162] ——, "Information-type measures of difference of probability distributions and indirect observation," *Stud. Sci. Math. Hung.*, vol. 2, pp. 299–318, 1967.
[163] I. Vajda, "On the f-divergence and singularity of probability measures," *Period. Math. Hung.*, vol. 2, no. 1–4, pp. 223–234, 1972.
[164] I. Sason, "On f-divergences: Integral representations, local behavior, and inequalities," *Entropy*, vol. 20, no. 5, 2018.
[165] V. Veeravalli, T. Basar, and H. Poor, "Minimax robust decentralized detection," *IEEE Trans. Inform. Theory*, vol. 40, no. 1, pp. 35–40, 1994.
[166] M. Fauß, A. M. Zoubir, and H. V. Poor, "Minimax robust detection: Classic results and recent advances," *IEEE Trans. Signal Process.*, vol. 69, pp. 2252–2283, 2021.
[167] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4394–4412, 2006.
[168] T. Van Erven and P. Harremos, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.
[169] K. Dedecius and P. M. Djurić, "Sequential estimation and diffusion of information over networks: A Bayesian approach with exponential family of distributions," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1795–1809, 2016.
[170] H. Zhu and R. Rohwer, "Information geometric measurements of generalisation," Aston University, Birmingham, UK, Tech. Rep., 1995.
[171] T. Minka, "Divergence measures and message passing," Microsoft Research, Cambridge, UK, Tech. Rep., 2005.
[172] A. Cichocki, S. Cruces, and S. Amari, "Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization," *Entropy*, vol. 13, no. 1, pp. 134–170, 2011.
[173] K. Pearson, "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Lond. Edinb. Dubl. Phil. Mag.*, vol. 50, no. 302, pp. 157–175, 1900.
[174] D. P. Malladi and J. L. Speyer, "A new approach to multiple model adaptive estimation," in *Proc. IEEE Conf. Decis. Control (CDC 1997)*, San Diego, CA, 1997.
[175] J. Wang and M. R. Taaffe, "Multivariate mixtures of normal distributions: Properties, random vector generation, fitting, and as models of market daily changes," *INFORMS J. Comput.*, vol. 27, no. 2, pp. 193–203, 2015.
[176] P. D. Hanlon and P. S. Maybeck, "Multiple-model adaptive estimation using a residual correlation Kalman filter bank," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 393–406, 2000.
[177] C. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY: Springer, 2013.
[178] A. Hinrichs, J. Prochno, and M. Ullrich, "The curse of dimensionality for numerical integration on general domains," *J. Complex.*, vol. 50, pp. 25–42, 2019.
[179] C. Genest and K. J. McConway, "Allocating the weights in the linear opinion pool," *J. Forecast.*, vol. 9, no. 1, pp. 53–73, 1990.
[180] M. H. DeGroot and J. Mortera, "Optimal linear opinion pools," *Manag. Sci.*, vol. 37, no. 5, pp. 546–558, 1991.
[181] R. T. Clemen, "Comment on Cooke's classical method," *Reliab. Eng. Syst. Saf.*, vol. 93, no. 5, pp. 760–765, 2008.
[182] T. Heskes, "Selecting weighting factors in logarithmic opinion pools," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS 1997)*, Denver, CO, 1997.
[183] L. M. De Carvalho, D. A. M. Villela, F. C. Coelho, and L. S. Bastos, "Combining probability distributions: Extending the logarithmic pooling approach," *arXiv preprint arXiv:1502.04206*, 2020.
[184] D. W. Bunn, "Two methodologies for the linear combination of forecasts," *J. Oper. Res. Soc.*, vol. 32, no. 3, pp. 213–222, 1981.
[185] D. W. Bunn and E. Kappos, "Synthesis or selection of forecasting models," *Eur. J. Oper. Res.*, vol. 9, no. 2, pp. 173–180, 1982.
[186] R. E. Barlow, R. W. Mensing, and N. G. Smiriga, "Combination of experts' opinions based on decision theory," in *Reliability and Quality Control*, A. P. Basu, Ed. Amsterdam, The Netherlands: Elsevier, 1986, pp. 9–19.
[187] M. H. DeGroot, "Reaching a consensus," *J. Am. Stat. Assoc.*, vol. 69, no. 345, pp. 118–121, 1974.
[188] A. Carvalho and K. Larson, "A consensual linear opinion pool," in *Proc. Int. Jt. Conf. Artif. Intell. (IJCAI 2013)*, Beijing, China, 2013.
[189] M. J. Rufo, C. J. Pérez, and J. Martín, "A Bayesian approach to aggregate experts' initial information," *Electron. J. Stat.*, vol. 6, pp. 2362–2382, 2012.
[190] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA: Duxbury, 2002.
[191] L. D. Brown, *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory*. Hayward, CA: Institute of Mathematical Statistics, 1986, vol. 9.
[192] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*. Boca Raton, FL: CRC Press, 1992.
[193] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 67, 2016.
[194] Y.-H. Kim and B. Pass, "Wasserstein barycenters over Riemannian manifolds," *Adv. Math.*, vol. 307, pp. 640–683, 2017.
[195] S. Ghosal and A. Van der Vaart, *Fundamentals of nonparametric Bayesian inference*. Cambridge, UK: Cambridge University Press, 2017, vol. 44.
[196] J. Aczél, *Lectures on Functional Equations and Their Applications*. New York, NY: Academic Press, 1966.
[197] R. G. Bartle, *The Elements of Integration and Lebesgue Measure*. New York, NY: Wiley, 1995.
[198] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1986.
[199] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge, UK: Cambridge Univ. Press, 2013.