

Constriction for Sets of Probabilities

Michele Caprio

PRECISE Center, Computer and Information Science, University of Pennsylvania, USA

CAPRIO@SEAS.UPENN.EDU

Teddy Seidenfeld

Philosophy and Statistics, Carnegie Mellon University, USA

TEDDY@STAT.CMU.EDU

Abstract

Given a set of probability measures \mathcal{P} representing an agent’s knowledge on the elements of a sigma-algebra \mathcal{F} , we can compute upper and lower bounds for the probability of any event $A \in \mathcal{F}$ of interest. A procedure generating a new assessment of beliefs is said to constrict A if the bounds on the probability of A after the procedure are contained in those before the procedure. It is well documented that (generalized) Bayes’ updating does not allow for constriction, for all $A \in \mathcal{F}$. In this work, we show that constriction can take place with and without evidence being observed, and we characterize these possibilities.

Keywords: constriction, dilation, sets of probabilities, evidence, conditioning, forgetting

1. Introduction

Call $\Delta(\Omega, \mathcal{F})$ the space of countably additive probability measures on a measurable space (Ω, \mathcal{F}) of interest and let $\mathcal{P} \subseteq \Delta(\Omega, \mathcal{F})$ be a set of probability measures. Then $\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A)$ is called the lower probability of A , and its conjugate $\overline{P}(A) = 1 - \inf_{P \in \mathcal{P}} P(A^c) = \sup_{P \in \mathcal{P}} P(A)$ is called the upper probability of A . They are two of the main building blocks of the literature known as *imprecise probability theory* [33]. Like in measure theory, where if outer and inner measures of a set coincide, then we say that the set has a measure, if upper and lower probabilities coincide, then they are an ordinary probability measure, and \mathcal{P} is a singleton. The parallel we just drew between imprecise probability theory and measure theory is not merely heuristic in nature: [33, Section 3.1.5] shows that the natural extension of a coherent lower probability from an algebra to the power set is the corresponding inner measure.

The reasons for studying imprecise probabilities are discussed at length in [1, 33] and references therein; in this work we focus especially on the motivations expressed in [12, 25]. There, the authors point out how specifying sets of probabilities – and thus their “boundary elements”, namely lower and upper probabilities – accounts for the ambiguity faced by the agent carrying out the analysis. This means

that since the agent does not know the true data generating process governing the experiment of interest, they may want to take advantage of the flexibility of IP theory and specify a set of probability measures to represent their ignorance. The set will be “wider”, that is, the difference between $\overline{P}(A)$ and $\underline{P}(A)$ will be larger for all $A \in \mathcal{F}$, the higher the uncertainty faced by the agent.

The aim of this paper is to study the constriction phenomenon that takes place after a given procedure.

Definition 1 Consider an event of interest $A \in \mathcal{F}$, a generic set of probability measures $\mathcal{P} \subseteq \Delta(\Omega, \mathcal{F})$, and denote by \underline{P} and \overline{P} the lower and upper probabilities associated with \mathcal{P} , respectively. Call \mathfrak{J} a generic procedure that produces a new assessment of beliefs, and denote by $\underline{P}^{\mathfrak{J}}$ and $\overline{P}^{\mathfrak{J}}$ the lower and upper probabilities resulting from such procedure, respectively.¹ Then, we say that procedure \mathfrak{J} (strictly) constricts A , in symbols $\mathfrak{J} \leftarrow A$, if $\underline{P}^{\mathfrak{J}}(A) > \underline{P}(A)$ and $\overline{P}^{\mathfrak{J}}(A) < \overline{P}(A)$. We say that \mathfrak{J} weakly constricts A if one of the two inequalities is weak.

Trivially, if \mathfrak{J} strictly constricts A , then \mathfrak{J} weakly constricts A . Constriction is sometimes called contraction [7, 15, 19]. We prefer constriction – as denoted in [20] – because contraction is used in the belief revision literature to denote an instance of corrigibility for full beliefs, which happens when an agent gives up some current evidence by moving to a logically weaker body of evidence [24]. In addition when we say that \mathfrak{J} produces a new assessment of beliefs, we mean that procedure \mathfrak{J} outputs (a set of) probabilities that represent the belief of the agent around the elements of \mathcal{F} . This should not be confused with AGM theory [19] where procedures generate a new set of full beliefs (sets of sentences). We keep the same terminology as, given the context, no confusion arises.

Our interest for constriction stems from surprising results involving the opposite phenomenon, called *dilation*, which

¹We use Hebrew letter \mathfrak{J} to denote the procedure because the Hebrew word for procedure, הנהל (pronounced *nohal*), begins with \mathfrak{J} . In addition, Latin, Greek, and Cyrillic letters p, P, π, Π – that could be associated with the word “procedure” – are usually associated with probabilities and partitions, while Greek letter ϖ can be easily confused with ω , which we will use to denote an element of the state space Ω of interest.

was first observed in the context of (generalized) Bayes' updating of \mathcal{P} [30]. We remark that \mathcal{P} need not be closed or convex. Pick any $A \in \mathcal{F}$, and call $\mathcal{P}(A) := \{P(A) : P \in \mathcal{P}\}$. Let $X : \Omega \rightarrow \mathbb{R}$ be a \mathcal{P} -measurable random variable, that is, let it be P -measurable for all $P \in \mathcal{P}$. Call then \mathcal{I} a generic index set, and let $\mathbf{X} := \{X = x_i\}_{i \in \mathcal{I}}$ be the sample space of measurable events associated with X . Denote by $\mathcal{P}(A | x_i) := \{P(A | X = x_i) : P \in \mathcal{P}\}$ the set of conditional probabilities of event A , given $X = x_i$. In order to avoid issues with conditional probability given a P -null event, $P \in \mathcal{P}$, we assume that the P 's in \mathcal{P} agree on those outcomes $\{X = x_i\}$ that are P -null. Let us denote by $E = X^{-1}(x) \subseteq \Omega$ the evidence collected after experiment \mathbf{X} . A weaker notion of constriction is the following.

Definition 2 *The (generalized) Bayes' updating procedure, denoted by $\mathfrak{I} = (B, E)$,² strictly pointwise constricts A if, for each x_i in a set of \mathcal{P} -probability 1 (that is, a set of P -probability 1, for all $P \in \mathcal{P}$), $\underline{P}(A) < \underline{P}(A | X = x_i)$ and $\overline{P}(A) > \overline{P}(A | X = x_i)$. It weakly pointwise constricts A if one of the two inequalities is weak.*

Generalized Bayes' updating merely pointwise constricts A if it pointwise constricts A , but $\inf_{i \in \mathcal{I}} \underline{P}(A | X = x_i) = \underline{P}(A)$ and $\sup_{i \in \mathcal{I}} \overline{P}(A | X = x_i) = \overline{P}(A)$. Strict and weak mere pointwise constriction are defined similarly to before.

To see that pointwise constriction is weaker than constriction, notice that we obtain strict constriction if $\inf_{i \in \mathcal{I}} \underline{P}(A | X = x_i) > \underline{P}(A)$ and $\sup_{i \in \mathcal{I}} \overline{P}(A | X = x_i) < \overline{P}(A)$, and weak constriction if one of the latter two inequalities is weak.

Define now, for all $P \in \mathcal{P}$, $\mathbf{X}_P^{A+} := \{x \in \mathbf{X} : P(A | X = x) > P(A)\}$ and $\mathbf{X}_P^{A-} := \{x \in \mathbf{X} : P(A | X = x) < P(A)\}$. The following lemma comes immediately from the law of conditional expectations.

Lemma 3 *Pick any $A \in \mathcal{F}$. Then, for all $P \in \mathcal{P}$, $P(\mathbf{X}_P^{A+}) > 0$ if and only if $P(\mathbf{X}_P^{A-}) > 0$.*

We then have two propositions that give sufficient conditions that preclude even weak constriction when collecting evidence in the form of an experiment to learn the value of the random variable X and using generalized Bayes' updating.

Proposition 4 *Pick any $A \in \mathcal{F}$. If $\mathcal{P}(A)$ is closed in the Euclidean topology, then no experiment \mathbf{X} is such that (B, E) weakly pointwise constricts A .*

Proposition 5 *Pick any $A \in \mathcal{F}$. No simple experiment \mathbf{X} is such that (B, E) weakly pointwise constricts A . That is,*

²In $\mathfrak{I} = (B, E)$, letter B denotes generalized Bayes' updating, and $E = X^{-1}(x)$ is the conditioning set. Throughout the paper, we refer to generalized Bayes' updating simply as "conditioning", while other techniques are referred to as "updating rules".

if X is a simple random variable (i.e. if the index set \mathcal{I} for the sample space \mathbf{X} is finite with P -probability 1, for all $P \in \mathcal{P}$), then (B, E) does not weakly pointwise constrict A .

The next example shows that if $\mathcal{P}(A)$ is open and X is not simple, we can have strict mere pointwise constriction. Together with Propositions 4 and 5, this exhausts the possible cases for generalized Bayes' updating.

Example 1 *Let A be a (measurable) event. For $0.4 < x < 0.6$ stipulate that $P_x(A) = x$, and let $\mathcal{P}(A) = \{P_x(A)\}$. So, $\underline{P}(A) = 0.4$ and $\overline{P}(A) = 0.6$, but $\mathcal{P}(A)$ is an open set. Let f be a well ordering of the rational numbers in the open interval $(0.4, 0.6)$, denoted as the set $\mathbb{Q}_{(0.4, 0.6)}$. So letting $\mathbb{N} = \{1, 2, \dots\}$ denote the natural numbers, we have $f : \mathbb{Q}_{(0.4, 0.6)} \leftrightarrow \mathbb{N}$ is a 1-1 (onto) function. Denote by $q_n = f^{-1}(n)$. Again, $q_n \in \mathbb{Q}_{(0.4, 0.6)}$. Let N be a P_X -measurable random variable where for each $0.4 < x < 0.6$, the likelihood ratio satisfies*

$$\frac{P_x(N = n | A)}{P_x(N = n | A^c)} = \frac{(1-x)q_n}{x(1-q_n)}. \quad (1)$$

Note that (1) constraints the distribution $P_x(N)$ without defining it. But (1) is coherent since, for each $0.4 < x < 0.6$, there are infinitely many values of q_n for which the likelihood ratio is greater than 1, and infinitely many values q_n for which the ratio is less than 1.³ By a trivial application of Bayes' Theorem,

$$\frac{P_x(A | N = n)}{P_x(A^c | N = n)} = \frac{q_n}{1 - q_n},$$

which is constant over $\mathcal{P}(A)$. That is, with respect to set $\mathcal{P}(A)$, the family of conditional probabilities of A , given $N = n$, is determinate despite the fact that the family of unconditional probabilities of A is indeterminate. Thus, for each $n \in \mathbb{N}$,

$$0.4 < \underline{P}(A | N = n) = q_n = \overline{P}(A | N = n) < 0.6$$

and Bayes' updating, given $N = n$, strictly merely pointwise constricts A . Note well that the strict constriction is merely pointwise as $\inf_{n \in \mathbb{N}} \underline{P}(A | N = n) = 0.4 = \underline{P}(A)$, and $\sup_{n \in \mathbb{N}} \overline{P}(A | N = n) = 0.6 = \overline{P}(A)$.

Let us now give an example of dilation, borrowed from [30]. It illustrates how, using generalized Bayes' updating, imprecise probabilities for an event A increase imprecision, for each possible outcome of an experiment.

³Here we see where the condition that $\mathcal{P}(A)$ is an open set is necessary. Condition (1) is incoherent when $P_x(A) = 0.4$ or $P_x(A) = 0.6$. Then, for all values of q_n , the likelihood ratio (1) would have values only to one side of 1, in contradiction with the law of total probability.

Example 2 Suppose we flip a fair coin twice so that the flips may be dependent. Denote by H_i and T_i outcome “heads” and “tails”, respectively, in tosses $i \in \{1, 2\}$. Let

$$\mathcal{P} := \left\{ P : P(H_1) = P(H_2) = \frac{1}{2}, P(H_1 \cap H_2) = p \right\}_{p \in [0, \frac{1}{2}]}$$

Now, suppose we flip the coin; we have $P(H_2) = 1/2$, but

$$0 = \underline{P}^B(H_2 | H_1) < \underline{P}(H_2) = \frac{1}{2} = \bar{P}(H_2) < \bar{P}^B(H_2 | H_1)$$

and

$$0 = \underline{P}^B(H_2 | T_1) < \underline{P}(H_2) = \frac{1}{2} = \bar{P}(H_2) < \bar{P}^B(H_2 | T_1),$$

where $\bar{P}^B(H_2 | H_1) = \bar{P}^B(H_2 | T_1) = 1$. As we can see, we start with a precise belief about the second toss and, no matter what the outcome of the first toss is, we end up having vacuous beliefs about the second toss.

The fact that Bayes’ rule of conditioning – arguably the most popular beliefs updating procedure – can give rise to dilation is one motivation for exploring updating techniques that instead admit constriction. That is the focus of our work.

The paper is divided as follows. Section 2 studies procedures that allow constriction to take place when no new evidence is collected. Theorems 8 and 9 are the main results and give very general conditions for procedures to give the opportunity for constriction in the absence of new collected evidence. In section 3, each individual in a group applies a (convex) personal pooling rule with “precise” inputs from the others in order to form their revised opinion. The process iterates until the individual opinions merge to a fixed point. Because the pooling rules are convex, the fixed point is a constriction of the original set of opinions. Section 4 studies constriction when evidence is collected and non-Bayesian updating rules are used to revise the agent’s beliefs. For a countably additive probability, given a generic partition \mathcal{E} of Ω , conditioning does not allow constriction for all $E \in \mathcal{E}$. So the only way of obtaining constriction for all $E \in \mathcal{E}$ is to intentionally forget the experiment associated with \mathcal{E} . But if we are able to make assumptions about the nature of \underline{P} , we can give conditions for constriction to take place for all $E \in \mathcal{E}$. Section 5 concludes our work. We study opportunities for constriction when we forego the assumption of countably additive probabilities in Appendix A, and we prove our results in Appendix B.

2. Constricting Without Evidence

In this section, we study procedures that give the opportunity for constriction when no data are collected.

2.1. Coherent Extension of a Precise Probability

Recall that, for de Finetti, a probability measure P is coherent if for any finite collection $\{A_i\}_{i=1}^n$ of nonempty subsets of a state space Ω of interest, we have that $\sup_{\omega \in \Omega} \sum_{i=1}^n c_i [I_{A_i}(\omega) - P(A_i)] \geq 0$, for all $c_1, \dots, c_n \in \mathbb{R}$, where I_{A_i} denotes the indicator function for set A_i . De Finetti’s Fundamental Theorem of Probability [8, Section 3.10] is the following.

Theorem 6 Call Ω the state space of interest. Given the probabilities $P(A_i)$ of a finite number of events $A_1, \dots, A_n \subseteq \Omega$, the probability $P(A_{n+1})$ of a further event A_{n+1}

1. either turns out to be determined if A_{n+1} is linearly dependent on the A_i ’s;
2. or can be assigned, coherently, any value in a closed interval $[p', p'']$.

More precisely, p' is the greatest lower bound (GLB) $\sup P(X)$ of the evaluations from below of the $P(X)$ given by the random quantities X linearly dependent on the A_i ’s for which we certainly have $X \leq A_{n+1}$.⁴ The same can be said for p'' (replacing \sup by \inf , maximum by minimum, A'_{n+1} by A''_{n+1} , and changing the direction of the inequalities, etc. It is the least upper bound of evaluations from above).

Notice that $[p', p'']$ can be an illusory restriction, for example if $p' = 0$ and $p'' = 1$. The interpretation to this result is the following. Suppose we express our subjective beliefs around events A_1, \dots, A_n via a precise probability distribution P . The fact that P is precise is a crucial tenet of de Finetti’s subjective probability theory. Then, if we want to coherently extend our beliefs to a new event A_{n+1} of interest, either we can do that “for free” if A_{n+1} is a linear combination of the other events, or we have an interval $[p', p'']$ within which to select the value to assign to $P(A_{n+1})$. De Finetti himself does not say specifically how to choose a value within $[p', p'']$. The takeaway seems to be along the lines of “you should be able to think hard enough to come up with a precise number $p \in [p', p'']$ to attach to $P(A_{n+1})$ ”.

Denote by $\mathfrak{d} = \text{deFin}$ the procedure of choosing any value in $[p', p'']$ to assign to the probability of event A_{n+1} . Then, the following holds.

Theorem 7 Suppose – in the notation of Theorem 6 – that $p' \neq p''$. Then, $\text{deFin} \leftrightarrow A_{n+1}$ if $P^{\text{deFin}}(A_{n+1}) \in (p', p'')$; the constriction is weak if $P^{\text{deFin}}(A_{n+1}) \in \{p', p''\}$.

Recall that, for Walley, a lower probability measure \underline{P} is coherent if for any finite collection $\{A_i\}_{i=0}^n$ of nonempty

⁴This inequality has to be interpreted as $X(\omega) \leq I_{A_{n+1}}(\omega)$, for all $\omega \in \Omega$.

subsets of a state space Ω of interest, we have that

$$\sup_{\omega \in \Omega} \left[\sum_{i=1}^n (I_{A_i}(\omega) - \underline{P}(A_i)) - s (I_{A_0}(\omega) - \underline{P}(A_0)) \right] \geq 0,$$

for all $s, n \in \mathbb{Z}_+$. In [33, Section 3.1] the author gives the imprecise probabilities (IP) counterpart of Theorem 6. That is, Walley presents a method to extend coherently lower and upper probabilities $\underline{P}(A_i), \overline{P}(A_i)$ from a finite collection of sets $\{A_1, \dots, A_n\} \subseteq 2^\Omega$ to any other $A_{n+1} \subseteq \Omega$. However, this result is not intended to prompt constriction, in contrast with de Finetti's Fundamental Theorem. The central idea in IP theory is to be "comfortable" working with sets of probabilities, and not being forced to select a precise value inside the set.

Notice that Theorems 6 and 7, and the results in [33, Section 3.1], are given in de Finetti's and Walley's frameworks, respectively. They both rely on the finitely additive probabilities. Because we consider finitely many events, though, this distinction is immaterial. That being said, it is important to point out that de Finetti's and Walley's extensions procedures apply also starting from an arbitrary (possibly infinite) set of events, where the distinction between finite and countable additivity matters, see for example Appendix A.

Let $\#$ denote the cardinality operator, $\text{Conv}(H)$ the convex hull of a generic set H , and $\text{ex}[K]$ the extreme points of a generic convex set K . We can generalize Theorem 7 to the following.

Theorem 8 *Suppose a generic procedure \mathfrak{I} generates a set $\mathcal{P} \subseteq \mathcal{A}(\Omega, \mathcal{F})$ of probabilities on (Ω, \mathcal{F}) such that $\#\mathcal{P} \geq 2$, and then prescribes a way of selecting one element $P^\star = \underline{P}^\mathfrak{I} = \overline{P}^\mathfrak{I}$ from $\text{Conv}(\mathcal{P})$. Assume that $\emptyset \neq \text{ex}[\text{Conv}(\mathcal{P})] = \{P_j^{\text{ex}}\}_{j \in \mathcal{J}}$, where \mathcal{J} is a generic index set. Then, we have that*

- if $P^\star \in \text{ex}[\text{Conv}(\mathcal{P})]$, then there may exist a collection $\{\tilde{A}\} \subseteq \mathcal{F}$ for which \mathfrak{I} weakly constricts \tilde{A} . In addition, $\mathfrak{I} \leftarrow A$, for all $A \in \mathcal{F} \setminus \{\tilde{A}\}$;
- if instead $P^\star = \sum_{j \in \mathcal{J}} \alpha_j P_j^{\text{ex}}$, $\alpha_j > 0$ for all j , then $\mathfrak{I} \leftarrow A$, for all $A \in \mathcal{F}$.

We can also give a topological version of Theorem 8; call $\partial_X H$ and $\text{int}_X H$ the boundary and the interior of a generic set H in X , respectively.

Theorem 9 *Endow $[0, 1]$ with the Euclidean topology, and call $\mathcal{B}([0, 1])$ the Borel sigma-algebra on $[0, 1]$. Fix a generic $A \in \mathcal{F}$, and assume that $\mathcal{P}(A) := \{P(A) : P \in \mathcal{P}\} \subseteq \mathcal{B}([0, 1])$ and that $\#\mathcal{P}(A) \geq 2$. Then,*

- if $\mathcal{P}(A)$ is closed in the Euclidean topology and $P^\star(A) \in \partial_{\mathcal{B}([0,1])} \mathcal{P}(A)$, then \mathfrak{I} weakly constricts A ;

- if instead $P^\star(A) \in \text{int}_{\mathcal{B}([0,1])} \mathcal{P}(A)$, then $\mathfrak{I} \nleftarrow A$.

Remark 10 *Notice that the assumption that $\mathcal{P}(A) \subseteq \mathcal{B}([0, 1])$ is verified in the case that \mathcal{P} is convex.*

It is immediate to see how Theorem 7 is a special case of Theorems 8 and 9. Another procedure that fits the requirement of Theorems 8 and 9 is *Halmos' extension* [17, Exercise 48.4], [5, Section 4.13]. Consider two generic measurable spaces (X, \mathcal{X}) and (Y, \mathcal{Y}) . Let $\mu \in \mathcal{A}(X, \mathcal{X})$ and, for all $x \in X$, $\nu_x \in \mathcal{A}(Y, \mathcal{Y})$. Suppose further that for all $B \in \mathcal{Y}$, $\nu_\bullet(B) : X \rightarrow [0, 1]$ is \mathcal{X} -measurable. Then,

- (i) map $E \mapsto \nu_x(\{y \in Y : (x, y) \in E\})$ is \mathcal{X} -measurable, for all $E \in \mathcal{X} \times \mathcal{Y}$;
- (ii) map $\pi : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$,

$$E \mapsto \pi(E) := \int_X \nu_x(\{y \in Y : (x, y) \in E\}) \mu(dx)$$

is a probability measure on $\mathcal{X} \times \mathcal{Y}$.

Suppose now that there exists a set $A \subseteq Y$ such that its inner and outer measures do not coincide, that is, for all $x \in X$, $\nu_{\star x}(A) \neq \nu_x^\star(A)$. Then, consider $A' = X \times A$. We have that

$$\pi_{A'} \equiv \pi(A') = \int_X \nu_x(\{y \in Y : (x, y) \in A'\}) \mu(dx),$$

and $\pi_{A'} \in [\nu_{\star x}(A), \nu_x^\star(A)]$. So Halmos' extension prescribes a way to extend a countable additive probability measure on \mathcal{Y} to another countably additive probability measure on $\mathcal{X} \times \mathcal{Y}$ that gives a well defined measure to a \mathcal{Y} -non-measurable set A . This value belongs to the interval whose endpoints are the inner and outer measures of A , respectively. It is immediate to see, then, how Halmos' extension satisfies the conditions of Theorems 8 and 9.

Remark 11 *Before going on, we need to mention a noteworthy difference between $\mathfrak{I} = \text{deFin}$ and extension theorems from measure theory (\grave{a} la Halmos). The relevant contrast is that for the Fundamental Theorem (applied to probability), de Finetti uses as his domains linear spans of, e.g. indicator functions. And for the measure theorists, the extension of probabilities to a larger ring of sets uses (countable) sums of indicators defined: in the finite case from an algebra, and in the infinite case from a sigma-algebra. Two additional meaningful differences are (1) for finite structures, de Finetti does not require that probabilities are defined over an algebra, whereas, the others do; (2) for infinite structures, as de Finetti does not require countable additivity, his inner and outer approximations are by finite sums of indicators. By contrast, the measure theorists require countably additive probabilities, and so they use countable sums for constructing inner and outer measure approximations.*

We now briefly present three procedures that fit the requirements of Theorems 8 and 9:

- (i) *convex pooling* [32, Section 2], in which the opinions of k agents (expressed via precise probabilities P_j , $j \in \{1, \dots, k\}$) are first pooled in a convex way, thus forming a set

$$\mathcal{P} = \left\{ P \in \Delta(\Omega, \mathcal{F}) : P = \sum_{j=1}^k \zeta_j P_j \right\},$$

where $\zeta_j \geq 0$, for all $j \in \{1, \dots, k\}$ and $\sum_{j=1}^k \zeta_j = 1$, from which a unique pooled opinion P^* is selected;

- (ii) *Jaynes' MaxEnt* [21], in which, given a set of constraints \mathbf{C} , the set of probabilities of interest to the researcher is $\mathcal{P} = \{P \in \Delta(\Omega, \mathcal{F}) : P \text{ satisfies } \mathbf{C}\}$, and P^* is selected by maximizing the Shannon entropy in \mathcal{P} ;
- (iii) *generalized fiducial inference* (GFI) [18], in which a set of data-dependent measures on the parameter space Ω – called generalized fiducial distributions – is defined by carefully inverting a deterministic data-generating equation without the use of Bayes' theorem. Mathematically, we can write $\mathcal{P} = \{P \in \Delta(\Omega, \mathcal{F}) : P \text{ satisfies [18, Equation (2)]}\}$. As pointed out in [18, Remark 4], P^* is then selected by choosing the appropriate norm to endow the sample space. In [18, Section 1], the authors point out how, while GFI is different philosophically from Dempster-Shafer theory [10] and inferential models [26], the resulting solutions of these three methods are often mathematically closely related to one another.

3. Constricting Based on Convex Pooling

As an example of a procedure that allows agents to collect new evidence, but does not use conditioning to update an agent's beliefs, we present the famous model in [9]. There, the author supposes that there are k individuals, each having their own subjective probability distribution F_i for the unknown value of some parameter $\omega \in \Omega$.⁵ For agent i , the opinions of all the other $k - 1$ agents represent new evidence. Instead of conditioning on those, agent i pools their own opinion with that of the other agents. DeGroot shows that, repeating this process for all agent i , the group reaches (asymptotically) an agreement on a common subjective probability distribution. After updating

⁵Usually the elements of the parameter space Θ are denoted by θ , while the elements of the state space Ω by ω . Since the focus of DeGroot's model is the parameter space only, we used – just in this section – the $\omega \in \Omega$ notation for the parameter space to maintain the notation consistent with other sections.

their opinions, the probability distribution for every member of the group belongs to the set

$$\mathcal{P} = \left\{ F = \sum_{j=1}^k \zeta_j F_j \right\}, \quad (2)$$

where $\zeta_j \in [0, 1]$, for all $j \in \{1, \dots, k\}$, and $\sum_{j=1}^k \zeta_j = 1$. In particular, for all $i \in \{1, \dots, k\}$, we write that after the first (pooling) iteration, the updated probability measure for agent i , denoted by F_{i1} is given by $F_{i1} = \sum_{j=1}^k p_{ij} F_j$. This means that individual i weighs the opinion of all the agents, including themselves, via coefficients p_{i1}, \dots, p_{ik} representing the relative importance that agent i assigns to the opinion of the other members of the group. Because this is true for all agents, we can give a linear algebra notation to the updating process. Call \mathbf{P} the $k \times k$ stochastic matrix whose rows are given by probability vectors (p_{i1}, \dots, p_{ik}) , $i \in \{1, \dots, k\}$. Call then $\mathbf{F} = (F_1, \dots, F_k)^\top$; we have that $\mathbf{F}^{(1)} = \mathbf{P}\mathbf{F}$, where $\mathbf{F}^{(1)} := (F_{11}, \dots, F_{k1})^\top$. Of course this holds for all iterations, so in turn we have that $\mathbf{F}^{(n)} = \mathbf{P}^n \mathbf{F} = \mathbf{P}^n \mathbf{F}$, for all $n \in \mathbb{N}$. The members continue to make these revisions indefinitely or until $\mathbf{F}^{(n)} = \mathbf{F}^{(n-1)}$, for all $n \geq N$, for some $N \in \mathbb{N}$, so further revisions would not change the opinions of the members. The following is the main result of [9].

Theorem 12 *If there exists $n \in \mathbb{N}$ such that every element in at least one column of \mathbf{P}^n is positive, then a consensus is reached.*

That is, if the condition in Theorem 12 is satisfied, then there exists a $k \times 1$ dimensional vector $\pi = (\pi_1, \dots, \pi_k)$ (that is unique, as guaranteed by [9, Theorem 3]) whose elements are non-negative and sum up to 1, and such that $\pi \mathbf{P} = \pi$. In turn, this entails that if we call \mathbf{H} the $k \times k$ stochastic matrix whose rows are all the same and equal to π , we have that $\mathbf{F}^* = \mathbf{H}\mathbf{F}$, where $\mathbf{F}^* = (F_1^*, \dots, F_k^*)^\top$ such that $F^* = F_1^* = \dots = F_k^* = \sum_{j=1}^k \pi_j F_j$, where F^* is the common subjective distribution that is reached in the consensus. Notice that F^* belongs to \mathcal{P} in (2).

Call $\mathcal{P}_n := \text{Conv}(F_{1n}, \dots, F_{kn})$, for all $n \in \mathbb{N}$. In this example we have that $\mathcal{P}_n \subseteq \mathcal{P}_{n-1}$, for all $n \in \mathbb{Z}_+$, where \mathcal{P}_0 is set \mathcal{P} in equation (2). This means that the limit of sequence (\mathcal{P}_n) is set $\bigcap_{n \in \mathbb{Z}_+} \mathcal{P}_n$. If the condition in Theorem 12 is satisfied, then $\bigcap_{n \in \mathbb{Z}_+} \mathcal{P}_n = \{F^*\}$. Given a generic set $A \in \mathcal{F}$, DeGroot procedure $\mathfrak{d} = \text{DeGr}$ may only weakly constrict A in general, for example if F^* belongs to the extrema of $\mathcal{P} \equiv \mathcal{P}_0$. That is, $F^*(A) \geq \underline{F}(A)$ and $F^*(A) < \overline{F}(A)$, or $F^*(A) > \underline{F}(A)$ and $F^*(A) \leq \overline{F}(A)$. Nevertheless, there may exist $r, s \in \mathbb{Z}_+$, $r < s$, such that $\underline{F}_s(A) > \underline{F}_r(A)$ and $\overline{F}_s(A) < \overline{F}_r(A)$, where $\underline{F}_s(A) = \inf_{F \in \mathcal{P}_s} F(A)$, $\overline{F}_s(A) = \sup_{F \in \mathcal{P}_s} F(A)$, and similarly for \mathcal{P}_r .

DeGroot model is one of the possible examples of an agent collecting evidence and then revising their initial

opinion using a rule that is different from conditioning on the gathered data. We showed that there is at least one such procedure in which constriction can take place if all the gathered information is used.

Notice that, in contrast with deGroot's model for consensus, in the light of Proposition 4, there is no opportunity for constriction in Aumann's important "Agreeing to Disagree" model [2]. We outline the reason why here. Identify the group's initial IP set of probabilities $\mathcal{P}(A)$ about the event of interest A , with the lower and upper probabilities $\underline{P}(A)$ and $\overline{P}(A)$, taken with respect to minimum and maximum of the precise individual opinions about A that results after each agent learns their "private" information. That is, $\mathcal{P}(A)$ obtains at round 1 in Aumann's process. Since there are finitely many agents in the group, this IP set is closed. In Aumann's model, the agents then iteratively share their individual, precise probabilities about A . At each subsequent round after the first, they use Bayes' updating to revise their individual probability of A , given the new, shared evidence of what they learn about the other's probability of A . This procedure amounts to (iteratively) using generalized Bayes' updating for the closed IP set $\mathcal{P}(A)$ given the updated individual precise probabilities for event A . By Aumann's theorem, after finitely many rounds the process reaches a fixed point where a consensus $P^*(A)$ is reached. But if $\underline{P}(A) < \overline{P}(A)$, then Proposition 4 establishes that it cannot be that the consensus opinion, $P^*(A)$, always satisfies $\underline{P}(A) < P^*(A) < \overline{P}(A)$.

4. Constricting Based on Non-Bayesian Updating

Suppose the results of an experiment induce a partition $\mathcal{E} = \{E_j\}$ of the state space of interest Ω . Then, if we retain the assumption that probability measures are countably additive, conditioning on E_j does not allow for constriction, for all $E_j \in \mathcal{E}$. So in general we have that constriction for all $E_j \in \mathcal{E}$ can take place only if we intentionally forget the whole experiment that induces partition \mathcal{E} . But if we are able to make assumptions on the nature of lower probability \underline{P} associated with the set \mathcal{P} of probabilities representing the agent's beliefs, and if we consider updating procedures that are alternative to Bayes' conditioning, we have opportunities for constriction.

4.1. Background

In this section, we give some background concepts that are needed to better understand the results that follow. Lower and upper probabilities (LP and UP, respectively) are a particular type of Choquet capacities.

Definition 13 *Given a measurable space (Ω, \mathcal{F}) with $\Omega \neq \emptyset$, we say that a set function $\nu : \mathcal{F} \rightarrow [0, 1]$ is a*

Choquet capacity if $\nu(\emptyset) = 0$, $\nu(\Omega) = 1$, and $\nu(A) \leq \nu(B)$ for all $A, B \in \mathcal{F}$ such that $A \subseteq B$.

Denote by $\mathcal{M} := \{P \in \Delta(\Omega, \mathcal{F}) : P(A) \geq \underline{P}(A), \forall A \in \mathcal{F}\}$ the set of (countably additive) probability measures compatible with \underline{P} [15], and assume it is nonempty and relatively compact. Notice also that \mathcal{M} is convex. The following are special cases of lower probabilities.

Definition 14 *LP \underline{P} is a Choquet capacity of order k , or k -monotone capacity, if for every collection $\{A, A_1, \dots, A_k\} \subseteq \mathcal{F}$ such that $A_i \subseteq A$, for all $i \in \{1, \dots, k\}$, we have*

$$\underline{P}(A) \geq \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{\#I-1} \underline{P}(\cap_{i \in I} A_i). \quad (3)$$

Its conjugate UP \overline{P} is called a k -alternating capacity because it satisfies that for every collection $\{A, A_1, \dots, A_k\} \subseteq \mathcal{F}$ such that $A \subseteq A_i$, for all $i \in \{1, \dots, k\}$,

$$\overline{P}(A) \leq \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{\#I-1} \overline{P}(\cup_{i \in I} A_i), \quad (4)$$

A special case of LP that we will use in the remainder of the paper are *convex LP's*, that are Choquet capacities of order 2; they satisfy $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B) - \underline{P}(A \cap B)$, for all $A, B \in \mathcal{F}$. Another special case of LP that we will use are belief functions.

Definition 15 *A LP \underline{P} is called a belief function if it is a Choquet capacity of order ∞ , i.e., if (3) holds for every k .*

Unique to a belief function is its intuitive interpretation as a random set object that realizes itself as subsets of Ω .

Definition 16 *If \underline{P} is a belief function, its associated mass function is the non-negative set function $m : \mathcal{F} \rightarrow [0, 1]$,*

$$A \mapsto m(A) := \sum_{B \subseteq A} (-1)^{\#(A-B)} \underline{P}(B), \quad (5)$$

where $A - B \equiv A \cap B^c$, and the subsets B of A have to belong to \mathcal{F} as well.

Properties of mass function m are the following

- (a) $m(\emptyset) = 0$;
- (b) $\sum_{B \subseteq \Omega} m(B) = 1$;
- (c) $\underline{P}(A) = \sum_{B \subseteq A} m(B)$, and is unique to \underline{P} .

Formula (5) is called the Möbius transform of \underline{P} [34]. A mass function m induces a precise probability distribution on \mathcal{F} , as the distribution of a random set. These concepts are further studied in [15]. Notice also that Definition 16 only applies if Ω is a finite set. A general definition and

Möbius characterization of belief functions on infinite sets has been given in [29].

To update a set of probabilities \mathcal{P} given a set $E \in \mathcal{F}$ is to replace set function \underline{P} with a version of the conditional set function $\underline{P}^\times(\cdot | E)$. The definition of \underline{P}^\times is precisely the job of the updating rule. Recall that we introduced generalized Bayes' rule of conditioning $\mathfrak{I} = (B, E)$ in section 1.

We now give the formal definitions of three additional updating rules for lower and upper probabilities. Generalized Bayes', geometric, and Dempster's rules are the ones that are most commonly used and studied in the literature, while Gärdenfors' rule is a very general updating mechanism that subsumes many other methods of belief revision. The reasons for why an agent endorses one instead of another are explored in [15, 31].

Definition 17 Let $\mathcal{P} \subseteq \Delta(\Omega, \mathcal{F})$ be closed and convex. Then, the conditional LP's and UP's according to the geometric rule are set functions $\underline{P}^G, \overline{P}^G$ such that, for all $A, E \in \mathcal{F}$

$$\begin{aligned} \underline{P}^{(G,E)}(A) &\equiv \underline{P}^G(A | E) := \frac{P(A \cap E)}{\underline{P}(E)} \quad \text{and} \\ \overline{P}^{(G,E)}(A) &\equiv \overline{P}^G(A | E) = 1 - \underline{P}^{(G,E)}(A^c), \end{aligned} \quad (6)$$

provided that $\underline{P}(E) > 0$.

So the main difference between generalized Bayes' and geometric updating procedures is that the former considers the infimum of the ratio of $P(A \cap E)$ and $P(E)$, while the latter considers the ratio of the infima. We introduce next Dempster's updating rule.

Definition 18 Call $E \in \mathcal{F}$ the collected evidence. Assume that \underline{P} is a belief function having mass function m and such that $\underline{P}(E) > 0$. Let \underline{P}_0 be a separate belief function whose associated mass function m_0 is such that $m_0(E) = 1$. The conditional belief function $\underline{P}^D(\cdot | E)$ is defined as

$$\underline{P}^{(D,E)}(A) \equiv \underline{P}^D(A | E) := \underline{P}(A) \oplus \underline{P}_0(E), \quad \forall A \in \mathcal{F},$$

where combination operator \oplus means that the mass function associated with $\underline{P}^D(\cdot | E)$ is

$$m^D(A | E) = \frac{\sum_{C \cap E = A} m(C)}{\sum_{C \cap E \neq \emptyset} m(C)}, \quad \forall A \in \mathcal{F}.$$

Consequently, Dempster's updating rule yields the following. If $\mathcal{P} = \mathcal{M}$, then the LP's and UP's according to Dempster's updating rule are set functions $\underline{P}^D, \overline{P}^D$ such that, for all $A, E \in \mathcal{F}$

$$\begin{aligned} \overline{P}^{(D,E)}(A) &\equiv \overline{P}^D(A | E) := \frac{\overline{P}(A \cap E)}{\overline{P}(E)} \quad \text{and} \\ \underline{P}^{(D,E)}(A) &\equiv \underline{P}^D(A | E) = 1 - \overline{P}^{(D,E)}(A^c), \end{aligned} \quad (7)$$

provided that $\overline{P}(E) > 0$.

If \underline{P} is a belief function and $\mathcal{P} = \mathcal{M}$, the geometric rule appears to be a natural dual to Dempster's rule. Operationally, though, they differ, as pointed out in [15, Section 2.2]. The main difference is that Dempster's rule requires \underline{P} to be a belief function, while geometric rule does not. A thorough comparison of geometric and Dempster's rules can be found in [11, 13]. In addition, an axiomatic extension of Dempster's conditioning rule to 2-monotone/2-alternating capacities has been given in [28].

Finally, we introduce the following.

Definition 19 Call $E \in \mathcal{F}$ the collected evidence. Assume that \underline{P} is a belief function having mass function m . Consider a function $\mathfrak{f} : \mathcal{F} \times \mathcal{F} \rightarrow [0, 1]$ having constraints

- (a) $\sum_{B \in \mathcal{F}} \mathfrak{f}(B, X) = 1$, for all $X \in \mathcal{F}$,
- (b) $B \subseteq E^c \implies \mathfrak{f}(B, X) = 0$,
- (c) $\mathfrak{f}(\emptyset, X) = 0$, for all $X \in \mathcal{F}$.

Then, the belief function $\underline{P}^{(I,E)}(\cdot) \equiv \underline{P}^I(\cdot | E)$ obtained according to Gärdenfors' generalized imaging updating rule (GGI) is such that its associated mass function is given by

$$m^I(A | E) = \sum_{X \in \mathcal{F}} \mathfrak{f}(A, X) m(X), \quad \forall A \in \mathcal{F}. \quad (8)$$

Since E is the collected information, we have that $E \sqcup E^c = \Omega$, where \sqcup denotes the disjoint union. Equation (8) tells us that, upon learning that the evidence collected is not in E^c , then the probabilities given by m to the (sub)events in E^c are transferred to the "closest" (sub)events in E according to function \mathfrak{f} . Constraint (a) is needed to ensure that the "probability bits" that are transferred are then normalized. Constraint (b) corresponds to the closed world assumption, which can be expressed as the statement "if the evidence collected is not in E^c , then it must belong to E ". Smets [31, Section C.6] drops requirement (b) because he works under the open world assumption, which negates the previous statement to symbolize that the agent may have specified the state space they work with incorrectly. So if the evidence collected is not in E^c , it may be in E but also in a superset of E that the agent did not consider at the beginning of the experiment. Constraint (c) is just a sanity check. Function \mathfrak{f} is a version of a conditional probability; it was first introduced in [16] and then generalized by [31]. The choice of function \mathfrak{f} informs how having collected evidence E influences our change of beliefs around A . GGI subsumes many other existing updating rules [31].

In the remainder of the paper, we write $\mathfrak{I} = (\times, E)$, $\times \in \{B, G, D, I\}$, to indicate the generalized Bayes', geometric, Dempster's, and Gärdenfors' updating rules, respectively, given collected evidence E . We write that $(\times, E) \leftarrow A$ if given evidence E , rule \times (strictly) constricts A , and

$(\times, \mathcal{E}) \leftrightarrow A$ if rule \times (strictly) constricts A for all elements of partition \mathcal{E} .

4.2. Intentional Forgetting

In this section we show the following claim. If we are not willing to assume that \underline{P} is at least convex – let alone a belief function – that is, if we can only use generalized Bayes’ and geometric rules to update our beliefs, then we cannot obtain constriction for all the elements E of a partition \mathcal{E} representing the results of an experiment of interest. In this very general case, we need to forget in order to constrict. We retain the assumptions that \mathcal{P} is closed and convex.

Lemma 20 *Let \mathcal{E} be a measurable and denumerable partition of Ω , and let $\times \in \{B, G\}$. Then, for any $A \in \mathcal{F}$ we have that*

$$\inf_{E \in \mathcal{E}} \underline{P}^\times(A | E) \leq \underline{P}(A) \quad \text{and} \quad \sup_{E \in \mathcal{E}} \overline{P}^\times(A | E) \geq \overline{P}(A).$$

An immediate consequence of Lemma 20 is the following.

Theorem 21 *Let \mathcal{E} be a measurable and denumerable partition of Ω . Then for $\times \in \{B, G\}$, we have that for all $A \in \mathcal{F}$, there exists $E \in \mathcal{E}$ such that $\mathfrak{I} = (\times, E)$ does not weakly constrict A .*

This result tells us that for any event $A \in \mathcal{F}$ of interest, we can never find an element of \mathcal{E} that (even weakly) constricts A . In turn, this implies that there exist $E_1, E_2 \in \mathcal{E}$, E_1 possibly different than E_2 , such that $[\underline{P}(A), \overline{P}(A)] \subset [\underline{P}^B(A | E_1), \overline{P}^B(A | E_1)]$ and $[\underline{P}(A), \overline{P}(A)] \subset [\underline{P}^G(A | E_2), \overline{P}^G(A | E_2)]$. As it appears clear, if the agent is unwilling to make any extra assumption on the nature of \mathcal{P} , then there is no opportunity for constriction to take place for all the elements of partition \mathcal{E} .

In [30, Theorem 2.3] the authors give sufficient conditions for dilation to take place for all $E \in \mathcal{E}$. Then, intentionally forgetting altogether the experiment that dilates A seems the only viable option to reach constriction. As the name suggests, intentional forgetting corresponds to an agent willingly forgetting pieces of information, for example because they are redundant, because they may be harmful, or because they are instructed to do so. If after collecting evidence E our current beliefs are encapsulated in lower and upper probabilities $\underline{P}^\times(\cdot | E)$ and $\overline{P}^\times(\cdot | E)$, respectively, then by forgetting we mean reversing the learning process so that our “updated” lower and upper probabilities becomes what used to be the lower and upper “priors”, i.e. $\underline{P}(\cdot)$ and $\overline{P}(\cdot)$, respectively.

The topic of forgetting is studied in statistics. In [27], for example, forgetting is intended in the sense of *stabilized forgetting*; with this we mean the following. Suppose that the agent is operating in an environment that is susceptible

to changes. Then, the agent’s response to surprising events depends on their beliefs about how likely the environment is to change. If it is volatile, a single unexpected event triggers forgetting of past beliefs and relearning of a new contingency. So at time t , the agent collects evidence E_t ; they use it to infer whether the environment has changed or not. In the former case, they erase their memory of past events and reset their prior belief to their initial prior knowledge. In the latter, they can learn a new posterior belief of the environment structure based on their previous belief. Another example is given by limited memory procedures. In [3, 4], the authors study bandit problems based on limited memory: working with restricted memory, data that are too old are forgotten.

Forgetting is studied in machine learning (ML) as well. In [6], for instance, the authors come up with an algorithm that features a forgetting factor which balances the relative importance of new data and past data and adjust the model to pay more attention to the new data when the concept drift is detected. In this framework, forgetting is intended as in past data progressively losing importance as new evidence is collected. Another example in ML where forgetting is crucial is *continual learning*: as data gets discarded and has a limited lifetime, the ability to forget what is not important and retain what matters for the future are the main issues that continual learning targets and focuses on [22].

Finally, and rather unsurprisingly, psychologists and cognitive scientists have thoroughly inspected the phenomenon of forgetting. The reference textbook that investigates intentional forgetting is [14]. The authors examine the effect on memory of instructions to forget in a wide variety of contexts. They point out how with the enormous number of information available nowadays, online forgetting of some information is necessary, and how often times replacing existing information with new information is mandatory (think of a person changing their phone number). Study on intentional forgetting stemmed from the phenomenon of *directed forgetting*: we are able to deal more effectively with large amounts of information by following instructions to treat some of the information as “to be forgotten” (e.g. evidence presented in a courtroom that, being inadmissible, is asked to be disregarded). In this way, interference is reduced and we are able to devote all of our resources to the remaining to-be-remembered information. It is easy to see how stabilized forgetting is a particular case of intentional forgetting, and so is the forgetting factor approach used in the machine learning literature.

Consider the problem of an agent that expresses their initial beliefs on (Ω, \mathcal{F}) via a set of probabilities $\mathcal{P} = \mathcal{M}$. For convenience, we write $\mathcal{P} \equiv \mathcal{P}_{E_0}$. As data become available, they update their beliefs using Bayes’ rule of conditioning for every element of \mathcal{P} . With this we mean the following. Suppose we collect evidence in the form of

$E_1 \subseteq \mathcal{F}$; then we update the elements of \mathcal{P} to obtain

$$\begin{aligned} \mathcal{P}_{E_1} &:= \left\{ P_{E_1} \in \Delta(\Omega, \mathcal{F}) : P_{E_1}(A) \equiv P(A | E_1) \right. \\ &= \left. \frac{P(E_1 | A)P(A)}{P(E_1)} \propto P(E_1 | A)P(A), \forall A \in \mathcal{F} \right\}, \end{aligned}$$

where $P \in \mathcal{P}$ represents the prior and $P(E_1 | \cdot)$ represents the likelihood. More in general, let the evidence collected up to time $t > 0$ be encapsulated in collection $\{E_k\}_{k=1}^t \subseteq \mathcal{F}$. Then, the agent's updated opinion is given by set

$$\begin{aligned} \mathcal{P}_{E_1 \dots E_t} &:= \left\{ P_{E_1 \dots E_t} : P_{E_1 \dots E_t}(A) \equiv P_{E_1 \dots E_{t-1}}(A | E_t) \right. \\ &= \left. \propto P(E_t | A, E_1, \dots, E_{t-1})P_{E_1 \dots E_{t-1}}(A), \forall A \in \mathcal{F} \right\}, \end{aligned}$$

where $\mathcal{P}_{E_1 \dots E_{t-1}} \ni P_{E_1 \dots E_{t-1}}(\cdot) \equiv P(\cdot | E_1 \dots E_{t-1})$ represents the "revised" prior (that is, the posterior computed at time $t - 1$) and $P(E_t | \cdot)$ represents the likelihood.

Pick any $k \in \mathbb{N}$ such that $1 \leq k \leq t$. Assume that $\mathcal{P}_{E_1 \dots E_{t-k}}$ and $\mathcal{P}_{E_1 \dots E_{t-k} \dots E_t}$ are both convex and closed. Fix an event $A \in \mathcal{F}$ of interest and define $\mathcal{P}_{\star E_1 \dots E_{t-k}}(A) := \{P_{E_1 \dots E_{t-k}} \in \mathcal{P}_{E_1 \dots E_{t-k}} : P_{E_1 \dots E_{t-k}}(A) = \underline{P}_{E_1 \dots E_{t-k}}(A)\}$ and $\mathcal{P}_{\star^* E_1 \dots E_{t-k}}(A) := \{P_{E_1 \dots E_{t-k}} \in \mathcal{P}_{E_1 \dots E_{t-k}} : P_{E_1 \dots E_{t-k}}(A) = \bar{P}_{E_1 \dots E_{t-k}}(A)\}$. For a generic $P \in \Delta(\Omega, \mathcal{F})$, and generic $A, B \in \mathcal{F}$, define notion of dependence d_P and sets induced by its value by

$$\begin{aligned} d_P(A, B) &:= P(A \cap B) - P(A)P(B), \\ \Sigma^+(A, B) &:= \{P \in \Delta(\Omega, \mathcal{F}) : d_P(A, B) > 0\}, \\ \Sigma^-(A, B) &:= \{P \in \Delta(\Omega, \mathcal{F}) : d_P(A, B) < 0\}. \end{aligned}$$

Call now $\mathfrak{I} = (\text{IF}_{\times}, \mathcal{E})$ the procedure of intentionally forgetting evidence \mathcal{E} (where \mathcal{E} can be an element of \mathcal{F} , a whole partition \mathcal{E} of state space Ω , or a collection $\{\mathcal{E}\}$ of partitions) after having updated endorsing rule \times . Then, the following gives sufficient conditions for intentional forgetting to induce constriction.

Theorem 22 *Fix an event $A \in \mathcal{F}$ of interest and let the agent endorse any rule $\times \in \{B, G\}$. If*

$$\mathcal{P}_{\star E_1 \dots E_{t-k}}(A) \cap \Sigma^-(A, E_{t-k+1} \cap \dots \cap E_t) \neq \emptyset$$

and

$$\mathcal{P}_{\star^* E_1 \dots E_{t-k}}(A) \cap \Sigma^+(A, E_{t-k+1} \cap \dots \cap E_t) \neq \emptyset,$$

then forgetting $E_{t-k+1} \cap \dots \cap E_t$ strictly constricts A , in symbols $(\text{IF}_{\times}, E_{t-k+1} \cap \dots \cap E_t) \leftarrow A$.

If this holds for all elements E_s of partition \mathcal{E}_s , $s \in \{t - k + 1, \dots, t\}$, we write

$$(\text{IF}_{\times}, \mathcal{E}_{t-k+1}, \dots, \mathcal{E}_t) \leftarrow A,$$

so we can forget all the experiments that took place after time $t - k$. If we let $k = t$, then we obtain stabilized forgetting as in [27]. Notice that for stabilized forgetting subscript $E_1 \dots E_{t-k}$ in Theorem 22 is substituted by E_0 . We also have the following.

Corollary 23 *Fix an event $A \in \mathcal{F}$ of interest and let the agent endorse any rule $\times \in \{B, G\}$. If*

$$\mathcal{P}_{\star E_k \dots E_t}(A) \cap \Sigma^-(A, E_1 \cap \dots \cap E_{k-1}) \neq \emptyset$$

and

$$\mathcal{P}_{\star^* E_k \dots E_t}(A) \cap \Sigma^+(A, E_1 \cap \dots \cap E_{k-1}) \neq \emptyset,$$

then forgetting $E_1 \cap \dots \cap E_{k-1}$ strictly constricts A , in symbols $(\text{IF}_{\times}, E_1 \cap \dots \cap E_{k-1}) \leftarrow A$.

In this case, we obtain the machine learning version of forgetting [6]. The agent intentionally forgets data collected before time k . More formally, the authors weight the evidence E_t collected at each time t by a coefficient depending on t that goes to 0 the farther time t is from present time T , that is, it goes to 0 as $|t - T|$ grows to infinity. Evidence that is old enough gets severely discounted, to the point that for practical purpose we can consider it as being forgotten, and so the result in Corollary 23 applies.

Notice that intentional forgetting is always a viable way of inducing constriction, as long as the selected updating rule induces dilation first. In this section we focused on $\times \in \{B, G\}$ because generalized Bayes' and geometric rules are the most general ones we presented (they do not require the lower probability of interest to be a belief function).

4.2.1. LEVI-NEUTRALITY

A particular type of forgetting is the one inspired by Levi's work on corrigible infallibility, see e.g. [23]. Suppose that at time t an agent is equipped with a body of beliefs \mathcal{K}_t regarding the events in \mathcal{F} , that is, a collection of logical predicates that describe the beliefs of the agent at time t .⁶ For the purpose of studying constriction, we focus on the (relatively compact and convex) set $\mathcal{P}_{\mathcal{K}_t}$ of probability measures representing the agent's beliefs at time t induced by \mathcal{K}_t .⁷ Consider a generic event $H \subseteq \Omega$; every element $P_{\mathcal{K}_t}$ of $\mathcal{P}_{\mathcal{K}_t}$ has the following two properties:

⁶Not to be confused with belief functions.

⁷In particular, we assume $\mathcal{P}_{\mathcal{K}_t} = \mathcal{M}_{\mathcal{K}_t} := \{P \in \Delta(\Omega, \mathcal{F}) : P(A) \geq \underline{P}_{\mathcal{K}_t}(A), \forall A \in \mathcal{F}\}$.

- if \mathcal{K}_t rules out H , that is, if given the beliefs at time t event H is considered impossible – written $\mathcal{K}_t \mapsto \neg H$ –, then $P_{\mathcal{K}_t}(A | H)$ is not well defined, for all $A \in \mathcal{F}$;
- if instead \mathcal{K}_t does not rule out H , then $P_{\mathcal{K}_t}(A | H)$ is well defined, for all $A \in \mathcal{F}$, and represents the agent’s beliefs around the plausibility of A if H obtains.

The set $\mathcal{P}_{\mathcal{K}_t}$ is compatible with lower probability $\underline{P}_{\mathcal{K}_t}$. Suppose then that at time $t + 1$ the agent collects evidence E . Let \mathcal{K}_{t+1} be the new body of beliefs; abusing notation, we write $\mathcal{K}_t \cup \{E\}$. Then, suppose that the agent endorses either of generalized Bayes’ or geometric rules, $\times \in \{B, G\}$; the updated beliefs of the agent are encapsulated in $\underline{P}_{\mathcal{K}_t}^\times(\cdot | E)$. If (according to \times) E dilates an event A' of interest, then the agent can neglect E to obtain constriction; they run “reverse conditioning”. We say that the agent is Levi-neutral towards E . Their infallible beliefs, encapsulated in \mathcal{K}_{t+1} , are subject to being corrected, and we have $\mathcal{K}_{t+2} = \mathcal{K}_t$. The body of beliefs contracts: \mathcal{K}_{t+1} loses element E and goes back to what used to be at time t . We write $(\text{LN}_\times, E) \leftrightarrow A'$ to denote that being Levi-neutral towards E – after having updated endorsing rule \times – constricts A' . If this holds for all elements of a partition \mathcal{E} of Ω representing the possible outcomes of an experiment of interest, we write $(\text{LN}_\times, \mathcal{E}) \leftrightarrow A'$.

4.3. Assumptions on the Nature of \underline{P}

If the agent is willing to make some assumptions on the type of lower probability \underline{P} that represents their beliefs, then we can have constriction for all $E \in \mathcal{E}$ without resorting to intentional forgetting. As the proverb goes, there is no free lunch. The following is Theorem 5.9 in [15].

Theorem 24 *Let $\mathcal{E} = \{E, E^c\}$ be the partition associated with the outcomes of the experiment of interest. Assume that \underline{P} is a belief function such that $\underline{P}(E), \underline{P}(E^c) > 0$, and consider any event $A \in \mathcal{F}$. Then, if \mathcal{E} dilates A under the Geometric rule, then it must constrict A under Dempster’s rule. Similarly, if \mathcal{E} dilates A under Dempster’s rule, then it must constrict A under the Geometric rule.*

The proof of Theorem 24 only requires that \underline{P} is convex, but we need the assumption that \underline{P} is in fact a belief function otherwise we would not be able to use Dempster’s rule (see Definition 18). As we can see, Dempster’s and geometric rule contradict each other.

Assuming that \underline{P} is a belief function allows us to use updating rules that are otherwise inaccessible.

Theorem 25 *Let E be the evidence collected by the agent, and assume that \underline{P} is a belief function having mass function m such that $\underline{P}(E) > 0$. Consider any event $A \in \mathcal{F}$. We have*

that $(I, E) \leftrightarrow A$ if and only if

$$\sum_{B \subseteq A} \left[\sum_{X \in \mathcal{F}} \tilde{f}(B, X)m(X) - m(B) \right] > 0 \quad \text{and}$$

$$\sum_{B \subseteq A^c} \left[\sum_{X \in \mathcal{F}} \tilde{f}(B, X)m(X) - m(B) \right] > 0.$$

If the conditions in Theorem 25 hold for every elements of partition \mathcal{E} , we write $(I, \mathcal{E}) \leftrightarrow A$.

The main point of this section is that if we are willing to formulate an assumption on the nature of lower probability \underline{P} associated with set \mathcal{P} representing our beliefs, then we are able to find constriction by using updating rules that in general do not allow for constriction for all $E \in \mathcal{E}$, like the geometric rule. We can also use entirely new updating techniques like Dempster’s rule or GGI that are otherwise inapplicable.

5. Conclusion

In this paper, we show that, when updating an agent’s opinions, there are at least three settings for constricting sets of probabilities (representing the beliefs), namely when belief revision is performed without evidence, when it is based on convex pooling, and when it is based on non-Bayesian updating. Also, we provide examples of procedures for every such framework.

This is just the first step towards a deeper study of the constricting phenomenon, that we will carry over in the next future. In particular, we plan to find more instances in which constricting is possible, and to find a trait d’union linking these settings.

Acknowledgments

We would like to thank three anonymous referees for their helpful and constructive comments. Michele Caprio would like to acknowledge funding from ARO MURI W911NF2010080.

Author Contributions

Both of the authors contributed equally to this paper.

References

- [1] Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes (editors). *Introduction to Imprecise Probabilities*. Wiley Series in Probability and Statistics. Chichester, UK : Wiley, 2014.

- [2] Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.
- [3] Dorian Baudry, Yoan Russac, and Olivier Cappé. On limited-memory subsampling strategies for bandits. In Marina Meila and Tong Zhang, editors, *ICML 2021: Proceedings of the 38th International Conference on Machine Learning*, pages 727–737. PMLR, 2021. URL <http://proceedings.mlr.press/v139/audry21b/audry21b.pdf>.
- [4] Donald A. Berry and Bert Fristedt. *Bandit problems – Sequential Allocation of Experiments*. Monographs on Statistics and Applied Probability. Dordrecht : Springer, 1985.
- [5] Patrick Billingsley. *Probability and Measure*. New York, NY : Wiley, second edition, 1995.
- [6] Weipeng Cao, Zhong Ming, Zhiwu Xu, Jiyong Zhang, and Qiang Wang. Online sequential extreme learning machine with dynamic forgetting factor. *IEEE Access*, 7:179746–179757, 2019.
- [7] Michele Caprio and Ruobin Gong. Dynamic precise and imprecise probability kinematics. Available on [arXiv:2110.04382](https://arxiv.org/abs/2110.04382), 2022.
- [8] Bruno de Finetti. *Theory of Probability*, volume 1. New York : Wiley, 1974.
- [9] Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.
- [10] Arthur P. Dempster. The Dempster–Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, 48(2):365–377, 2008.
- [11] Didier Dubois and Henri Prade. Updating with belief functions, ordinal conditional functions and possibility measures. In Piero P. Bonissone, Max Henrion, Laveen N. Kanal, and John F. Lemmer, editors, *UAI '90: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, MIT, Cambridge, MA, USA, July 27-29, 1990*, pages 311–330. Elsevier, 1990.
- [12] Daniel Ellsberg. Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669, 1961.
- [13] Itzhak Gilboa and David Schmeidler. Updating ambiguous beliefs. *Journal of Economic Theory*, 59(1):33–49, 1993.
- [14] Jonathan M. Golding and Colin M. MacLeod (editors). *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ : Lawrence Erlbaum Associates, 1998.
- [15] Ruobin Gong and Xiao-Li Meng. Judicious judgment meets unsettling updating: dilation, sure loss, and Simpson’s paradox. *Statistical Science*, 36(2):169–190, 2021.
- [16] Peter Gärdenfors. *Knowledge in flux: Modeling the dynamics of epistemic states*. Boston, MA : The MIT press, 1988.
- [17] Paul R. Halmos. *Measure Theory*. Graduate Texts in Mathematics. New York, NY : Springer, 1950.
- [18] Jan Hannig, Hari Iyer, Randy C. S. Lai, and Thomas C. M. Lee. Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111(515):1346–1361, 2016.
- [19] Sven Ove Hansson. Logic of Belief Revision. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Palo Alto, CA : Metaphysics Research Lab, Stanford University, Spring 2022 edition, 2022.
- [20] Timothy Herron, Teddy Seidenfeld, and Larry Wasserman. Divisive conditioning: Further results on dilation. *Philosophy of Science*, 64(3):411–444, 1997.
- [21] Hiremagalur K. Kesavan. Jaynes’ maximum entropy principle. In Christodoulos A. Floudas and Panos M. Pardalos, editors, *Encyclopedia of Optimization*. Boston, MA : Springer, 2008.
- [22] Timothée Lesort, Vincenzo Lomonaco, Andrei Stoian, Davide Maltoni, David Filliat, and Natalia Díaz-Rodríguez. Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. *Information Fusion*, 58:52–68, 2020.
- [23] Isaac Levi. Why indeterminate probability is rational. *Journal of Applied Logic*, 7(4):364–376, 2009.
- [24] Isaac Levi. How infallible but corrigible full belief is possible. In Horacio Arló-Costa, Vincent F. Hendricks, and Johan van Benthem, editors, *Readings in Formal Epistemology: Sourcebook*, pages 247–267. Cham, CH : Springer International Publishing, 2016.
- [25] Massimo Marinacci and Luigi Montrucchio. *Introduction to the Mathematics of Ambiguity*. Uncertainty in Economic Theory. New York, NY : Routledge, 2004.
- [26] Ryan Martin and Chuanhai Liu. *Inferential Models: Reasoning with Uncertainty*, volume 145 of *Monograph on Statistics and Applied Probability*. Boca Raton, FL : CRC Press, 2015.

- [27] Vincent Moens and Alexandre Zénon. Learning and forgetting using reinforced Bayesian change detection. *PLoS Computational Biology*, 4(15):1–41, 2019.
- [28] Davide Petturiti and Barbara Vantaggi. Conditional submodular Choquet expected values and conditional coherent risk measures. *International Journal of Approximate Reasoning*, 113:14–38, 2019.
- [29] Davide Petturiti and Barbara Vantaggi. *How to Assess Coherent Beliefs: A Comparison of Different Notions of Coherence in Dempster-Shafer Theory of Evidence*, pages 161–185. Springer International Publishing, Cham, 2022.
- [30] Teddy Seidenfeld and Larry Wasserman. Dilation for sets of probabilities. *The Annals of Statistics*, 21(3): 1139–1154, 1993.
- [31] Philippe Smets. About updating. In *Proceedings of the Seventh international conference on Uncertainty in Artificial Intelligence*, pages 378–385, 1991.
- [32] Rush T. Stewart and Ignacio Ojea Quintana. Probabilistic opinion pooling with imprecise probabilities. *Journal of Philosophical Logic*, 47(1):17–45, 2018.
- [33] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*, volume 42 of *Monographs on Statistics and Applied Probability*. London, UK : Chapman and Hall, 1991.
- [34] Roland R. Yager and Liping Liu (editors). *Classic works of the Dempster-Shafer theory of belief functions*, volume 39 of *Studies in Fuzziness and Soft Computing*. New York, NY : Springer, 2008.