

Série des Documents de Travail

n° 2017-49

**An instrumental approach to the value of
information**

**M.DE LARA¹
O.GOSSNER²**

Les documents de travail ne reflètent pas la position du CREST et n'engagent que leurs auteurs.
Working papers do not reflect the position of CREST but only the views of the authors.

¹ CERMICS, ENPC; Université Paris Est. E-mail : delara@cermics.enpc.fr

² CREST; Ecole Polytechnique; LSE. E-mail : olivier.gossner@polytechnique.edu

An instrumental approach to the value of information

Michel DE LARA Olivier GOSSNER

October 25, 2017

Abstract

We consider an agent who acquires information on a state of nature from an information structure before facing a decision problem. How much information is worth depends jointly on the decision problem and on the information structure. We represent the decision problem by the set of possible payoffs indexed by states of nature. We establish and exploit the duality between this set on one hand and the value of information function, which maps beliefs to expected payoffs under optimal actions at these beliefs, on the other. We then derive global estimates of the value of information of any information structure from local properties of the value function and of the set of optimal actions taken at the prior belief only.

1 Introduction

The value of a piece of information to an economic agent depends on many factors. It depends on the information at hand, on the agent's prior on the state of nature, on the decisions available to that agent as well as the the agent's preferences. In a trial, for instance, not all pieces of information have the same value, as evidence is valuable only insofar as it can affect the final outcome. This value also hinges on the prior belief that the defendant is guilty, as this belief plays a role on the impact of an evidence on conviction. In a framework of portfolio optimisation, the value of information for an investor depends on the set of assets, hence on the decisions available. Finally, the agent's degree of risk aversion, hence his preferences, also influences the value of information.

All these elements tend to be intrinsically tied and separating the influence of one of them from that of others is not straightforward. The dependency of the value of information on agent's decisions and preferences is certainly a major reason why most information rankings are either uniform among agents or restrict to certain classes of agents. Blackwell (1953)'s comparison of experiments, for instance, is uniform ; it states that an information structure is more informative than another one if all agents, no matter their available choices and preferences, weakly prefer the former to the latter. Lehmann (1988); Persico (2000); Cabrales, Gossner, and Serrano (2013) are example of papers that build information rankings based on restricted sets of decision problems.

The flip side of this approach is that information rankings are silent as to the dependency of the value of fixed piece information on the agent's preferences and available choices. They do not tell us what makes information more or less valuable to an arbitrary agent, and neither can they identify the agents who value a given piece of information more than others. If we want to answer this type of questions, we need to examine carefully how information, priors, decisions and preferences come to play.

To start with, the effect of priors and evidence on beliefs is well understood. Given a prior belief, and after receiving some information, the agent forms a posterior belief. The law of iterated expectation implies that the posterior beliefs average out to the prior belief. In fact, it is often useful to represent information acquisition as a distribution of posterior beliefs averaging to the prior belief. This representation, first introduced by Bohnenblust, Shapley, and Sherman (1949), was subsequently used in several applications, such as repeated games with incomplete information with zero-sum (Aumann and Maschler, 1967), non zero-sum (Aumann and Hart, 1986), and Bayesian persuasion (Kamenica and Gentzkov, 2011).

Let us now turn to the role of available decisions and preferences. Taking these elements as fixed, for every posterior belief, the agent makes a decision that maximises her expected utility. Though this decision process, to each (posterior) belief of the agent corresponds an expected utility at this belief. This map from beliefs to expected payoffs, called the value function, depends uniquely on decisions and preferences. The value of a piece of information for an agent is precisely the difference in expected utilities from having or not having the information at hand. Given that posteriors average out to the prior, the value of information is thus expressed as a convexity gap for the value function around the prior. Given a prior and the posteriors, the value of information can be read directly from the value function.

Thus, the value function fully captures the agent's preferences for information.

The key to an agent's preferences for information is thus to understand the properties the value function inherits from available decision and preferences. In this paper, we clarify and exploit the duality relationship between decisions and preferences on the one hand (using the proper representation) and the value function on the other. As it appears, this connexion is formally the same as the one that allows to derive profit functions as a function of prices from production sets in the classical Hotelling Lemma's framework (Hotelling, 1932). In particular, the (sub)gradient of the value function at any belief can be represented as the set of optimal decisions at this belief. The second derivative of the value function, which quantifies the intensity of the convexity gap at the belief, is thus naturally represented as the impact of beliefs on optimal decisions. Put simply, the local value of information is appropriately measured by how much beliefs impact choices around the prior.

Consider the insurance case detailed as section 5 below. An insurance company charges both a premium and a fixed fee for every contract. We study the insuree's choice of indemnity as a function of her perception of her own risk level. In our example, it is optimal not to insure for small perceived risk levels, whereas for higher levels of risk aversion, the optimal indemnity is positive and depends smoothly on the perceived risk level. At the threshold between those two situations, the decision maker is indifferent between no insurance and a positive insurance level. Now we assume that the insuree may obtain a small piece of information on the risk of loss. If her perceived risk is low, then this information is too small to trigger a choice of insurance, and as the information has no effect on choices, it has no value. For higher perceived risk levels, the demand for insurance changes smoothly with the perceived risk of loss and it follows from the envelope theorem that the choice of insurance with no information is first-order optimal even in the presence of a small information. The value of information is then a second order degree only of the information received. For an agent who is indifferent between no insurance and a strictly positive indemnity, a small piece of information has a high impact on the decision. A small information towards a lesser risk is enough to break the indifference towards no insurance, and a small information towards an increased breaks it towards a positive insurance level. For such agent, the value of information is highest, and so is the impact of information on decisions. Our results make provide a foundation for the above intuitions on the value of information for an agent.

In the paper we express the value of information according to the influence it has on decisions. We provide three upper and lower bounds on the value of information that apply to the three example cases above. We generalise the typology of the insurance example not only from this particular decision problems to arbitrary ones, but also from small information to all information structures, whether or not all posterior beliefs are close to the prior.

In a first upper and lower bounds, we characterise information with positive value. We show that information has positive value when, at least one of the optimal actions at the prior becomes suboptimal for some of the posteriors. We thus define the confidence set at \bar{p} as the set of posterior beliefs for which all optimal actions at \bar{p} remain optimal. We show that information has positive value if and only if posterior beliefs fall outside of the confidence set with positive probability. We provide corresponding lower and upper bounds to the value of

information.

In the second bounds, we express the fact that the value of information is maximal when it impacts actions the most, which happens when information breaks indifferences between several choices. We show that, when this is the case, the value of information can be suitably measured by an expected norm-1 distance between the prior and the posterior.

Finally, our third bounds applies to cases in which the agent’s optimal choice is a smooth function of her belief around the prior. We show that, in this situation, the value function is also smooth around the prior, and the value of information is essentially a quadratic function of the expected distance between the prior and the posterior.

In a finite decision problem such as such as shopping behavior (McFadden, 1973) or residential location (McFadden, 1978), at any given prior, the agent either has an optimal action that is locally constant, or is indifferent between several optimal choices. The first and second upper and lower bounds are particularly useful in finite choice problems. The third bounds typically applies in decision decision problems with a continuum of choices such as scoring rules (Brier, 1950) or investments decisions (Arrow, 1971). In certain decision problems, such as the insurance problem of Section 5, the behavior of optimal choice as a function of the belief depends on the range of parameters, and the appropriate bounds apply accordingly.

The paper is organized as follows. Section 2 presents the model and introduces the duality between actions and the value of information. The main results are presented in Section 3. Section 4 is devoted to applications to the question of marginal value of information, and Section 5 to our insurance example. Finally, Section 6 presents the related literature, and Section 7 concludes.

2 Model

We consider the classical question of an agent who faces a decision problem under imperfect information on a state of nature.

2.1 Information and action

The set of states of nature is a finite set K . We identify the set Σ of signed measures on K with \mathbb{R}^K . The agent holds a prior belief \bar{p} with full support in the set

$$\Delta = \Delta(K) \subset \Sigma \tag{1}$$

of probability distributions over K . We identify Δ with the simplex of \mathbb{R}^K .

A *decision problem* is given by an arbitrary compact convex choice set D and by a continuous payoff function $g: D \times K \rightarrow \mathbb{R}$. The convexity of D is justified by allowing the agent to randomise over decisions. Consistent with Blackwell (1953)’s framework, we define the set of *actions* as the compact convex subspace of \mathbb{R}^K given by:

$$A = \{(g(d, k))_{k \in K}, d \in D\} . \tag{2}$$

The scalar product between a vector $v \in \mathbb{R}^K$ and a signed measure $s \in \mathbb{R}^K$ is

$$\langle s, v \rangle = \sum_{k \in K} s_k v_k . \quad (3)$$

Under belief $p \in \Delta$, the decision maker chooses a decision $d \in D$ that maximises $\sum_k p_k g(d, k)$, or equivalently $a \in A$ that maximises $\langle p, a \rangle$, and the corresponding expected payoff is *expected payoff* $\max_{a \in A} \langle p, a \rangle \in \mathbb{R}$. We define the *value function* $V_A : \Delta \rightarrow \mathbb{R}$ by:

$$V_A(p) = \max_{a \in A} \langle p, a \rangle , \quad \forall p \in \Delta . \quad (4)$$

The value function $V_A : \Delta \rightarrow \mathbb{R}$ is convex — as the supremum of the family of linear maps $\langle \cdot, a \rangle$ for $a \in A$ — and continuous — as its domain is the whole convex set Δ (Hiriart-Uruty and Lemaréchal, 1993, p.175).

We follow Bohnenblust, Shapley, and Sherman (1949); Blackwell (1953) and describe a *statistical experiment* as a distribution of posterior beliefs that average to the prior belief. Equivalently, given the prior belief \bar{p} , we define an *information structure* as a random variable \mathbf{q} with values in Δ describing the agent's posterior beliefs, defined over a probability space Ω equipped with a probability \mathbb{P} , and such that:

$$\mathbb{E} \mathbf{q} = \int_{\Omega} \mathbf{q}(\omega) d\mathbb{P}(\omega) = \bar{p} , \quad (5)$$

where \mathbb{E} denotes the expectation.

Given the action set A and the information structure \mathbf{q} , the *value of information* $\mathbf{VoI}_A(\mathbf{q})$ is the difference between the expected payoff for an agent who receives information according to \mathbf{q} and one whose prior is \bar{p} . It is given by:

$$\mathbf{VoI}_A(\mathbf{q}) = \mathbb{E} V_A(\mathbf{q}) - V_A(\bar{p}) . \quad (6)$$

2.2 Duality between actions and the value function

Given a belief $p \in \Delta$, we let $A^*(p) \subset A$ be the *set of optimal actions at belief p*, given by

$$A^*(p) = \arg \max_{a' \in A} \langle p, a' \rangle = \{a \in A \mid \forall a' \in A, \langle p, a' \rangle \leq \langle p, a \rangle\} . \quad (7)$$

The set $A^*(p)$ is nonempty, closed, convex (as A is convex and compact), and is a subset of the *boundary* ∂A of A .

Conversely, an outside observer can make inferences on the agent's beliefs from observed actions. For an action $a \in A$, the set $\Delta_A^*(a)$ of *beliefs revealed by action a (beliefs revealed at a)* is the set of all beliefs for which a is an optimal action, it is given by:

$$\Delta_A^*(a) = \{p \in \Delta \mid \forall a' \in A, \langle p, a' \rangle \leq \langle p, a \rangle\} . \quad (8)$$

The set $\Delta_A^*(a)$ is the intersection with Δ of the so-called *normal cone* $N_A(a)$, that will be defined in (33). As the normal cone is nonempty for any $a \in \partial A$ (Hiriart-Uruty and

Lemaréchal, 1993, p.137), the set $\Delta_A^*(a)$ is nonempty when a lies in the “North-East” part of the boundary of A .

Obviously, given $a \in A$ and $p \in \Delta$, $a \in A^*(p)$ iff $p \in \Delta_A^*(a)$, as both express that action a is optimal under belief p .

In § 3.1, we rely on the definitions of the set $A^*(p)$ of optimal actions at belief p and of the set $\Delta_A^*(a)$ of beliefs revealed by action a to characterise the information structures with positive value. We show that an information structure has positive value if and only if, with positive probability, one of the optimal actions at the prior \bar{p} is not optimal under the realization of the posterior belief \mathbf{q} .

There is a natural and useful relationship between the set $A^*(p)$ of optimal actions at belief p in (7) and the value function V_A in (4); namely, $a \in A^*(p)$ if and only if a belongs to the subgradient of V_A at p (Hiriart-Uruty and Lemaréchal, 1993, p.220). Since any proper convex function on \mathbb{R}^n is differentiable (in the classical sense) at a point if and only if its subgradient at a point is a singleton (Hiriart-Uruty and Lemaréchal, 1993, p.251), we immediately observe that V_A is differentiable at p iff there is a unique optimal action at p . In this paper, we exploit the relationship between differentiability properties of the value function V_A at \bar{p} with the value of information \mathbf{VoI}_A for an agent with prior \bar{p} . In particular, in § 3.2, we provide bounds on the value of information when V_A is non-differentiable at the prior \bar{p} .

The following example illustrates the duality relationship between the set A of actions and the value function V_A .

Example 1 Consider two states of Nature, $K = \{1, 2\}$, decisions consisting of $D = \{d_1, d_2, d_3, d_4\}$ and their mixtures, and payoffs given by Table 1.

	$k = 1$	$k = 2$
d_1	3	0
d_2	2	2
d_3	0	$\frac{5}{2}$
d_4	0	0

Table 1: Table of payoffs

In this case, A is the convex hull of the four points $(3, 0)$, $(2, 2)$, $(0, \frac{5}{2})$ and $(0, 0)$. The value function V_A expressed as a function of the probability p of state 2 is the maximum of the following three linear maps: $3(1 - p)$, 2 , and $\frac{5}{2}p$. Action $(3, 0)$ is optimal for $p \leq 1/3$, $(2, 2)$ is optimal for $p \in [1/3, 4/5]$ and $(0, 5/2)$ is optimal for $p \geq 4/5$. Both set A and function V_A are represented in Figure 2.2.

At $p = 4/5$, the optimal actions are $(2, 2)$, $(0, 5/2)$, and their mixtures. At this point, the mapping V_A is not differentiable. However, its subgradient — which can be visualized as the set of straight lines that are below V_A and tangent to it at $p = 4/5$ — is still well defined and corresponds precisely to the optimal actions $A^*(4/5)$, i.e. the convex hull of $\{(0, 5/2), (3, 3)\}$.

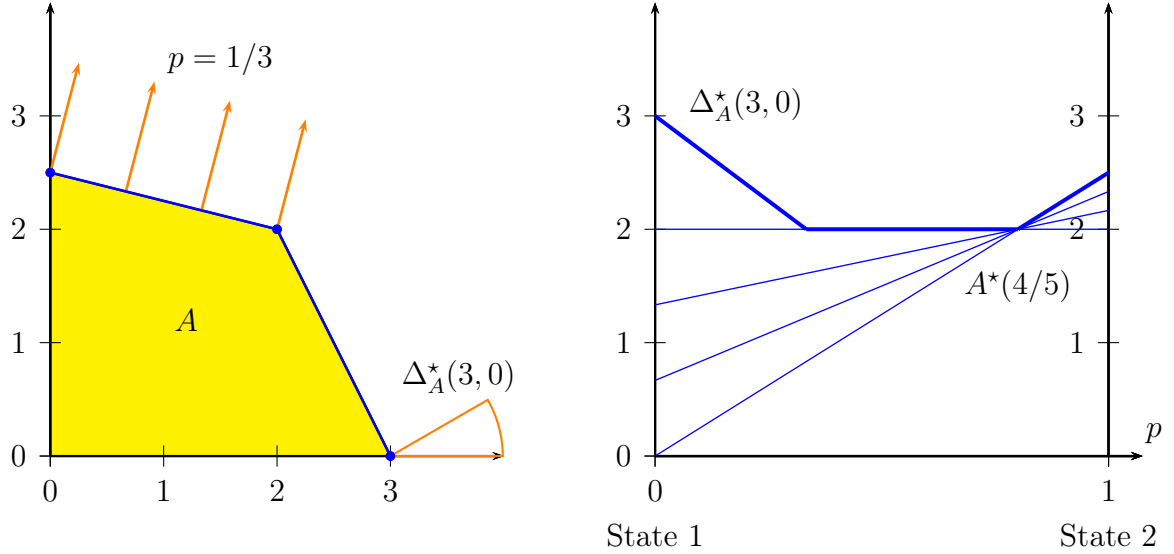


Figure 1: The set A on the left and the function V_A on the right. Each of the four arrows on the left represents an action a such that $p = 4/5$ belongs to the set $\Delta_A^*(a)$ of beliefs revealed by action a . On the right part, these four actions (each attached to an arrow) can be seen as four elements of the subgradient of the value function V_A at $p = 4/5$. The set $\Delta_A^*(3, 0) = [0, \frac{1}{3}]$ can be visualized both as the normal cone at $(3, 0)$ on the left part and as the range of values of probabilities p for which $(3, 0)$ is optimal on the right.

The set $\Delta_A^*(3, 3)$ of beliefs revealed by action $(3, 3)$ consists of the range $p \in [4/5, 1]$, and it can be seen on the right part that, for this range of probabilities, the action $(3, 0)$ is optimal and that V_A is linear and equal to $3(1 - p)$.

3 On the value of information

In this section, we relate the geometry of the set A of actions with the behavior of the agent around the prior \bar{p} , with differentiability properties of the value function V_A at the prior \bar{p} , and with the value of information \mathbf{VoI}_A . This approach allows us to derive bounds on the value of information that depend on how information impacts actions.

First, in §3.1, we consider information that does not allow to eliminate optimal actions. We introduce the *confidence set* as the set of posterior beliefs at which all optimal actions at the prior remain optimal. We show that information is valuable if and only, with positive probability, it can lead to a posterior outside this set.

Second, in §3.2, we then consider the somewhat opposite case of tie-breaking information. This corresponds to situations in which the agent is indifferent between several actions, and

the information allows her to select among them. We show that the value of information can be written as an expected distance between the prior and the posterior, provided that posterior beliefs move in these tie-breaking directions.

These two first approaches are suitable in finite decision problems where the value function is piecewise linear. In the third approach, in §3.3, we look at situations in which the optimal action is locally unique around the prior and depends on information impacts actions in a continuous and differentiable way. There, we show that the value of information can essentially be measured as an expected square distance from the prior to the posterior. This approach is particularly adapted to cases in which the space of actions is sufficiently rich, and where small changes of beliefs lead to corresponding change changes of actions.

3.1 Valuable information

Our first task is to make precise the idea that useful information is information that affects optimal choices. Since there are potentially many optimal actions at a prior \bar{p} and at a posterior p , there are in principle many ways to formalize this idea.

We say that a belief p is in the *confidence set* $\Delta_A^c(\bar{p})$ of \bar{p} when all optimal actions at \bar{p} are also optimal at p :

$$\Delta_A^c(\bar{p}) = \bigcap_{a \in A^*(\bar{p})} \Delta_A^*(a). \quad (9)$$

Another way to look at this notion is to consider an observer who sees choices by the decision maker: $p \in \Delta_A^c(\bar{p})$ when none of the actions chosen by the agent at \bar{p} would lead the observer to refute the possibility that the agent has belief p .

The notion of confidence set allows for a characterisation of valuable information as follows.

Proposition 1 *Valuable Information*

$$\begin{aligned} \mathbf{VoI}_A(\mathbf{q}) = 0 & \quad \text{iff} \quad \exists a^* \in A, \quad a^* \in A^*(\mathbf{q}) \text{ a.s.} \\ & \quad \text{iff} \quad \mathbf{q} \in \Delta_A^c(\bar{p}) \text{ a.s.} \end{aligned}$$

It is relatively straightforward to see that if all posteriors remain in the confidence set, information is valueless. In fact, when this is the case, the same action is optimal for all the posteriors, which means that the agent can play this action, while ignoring the new information, and obtain the same value. The proposition shows that the converse result also holds: the value of information is positive whenever posteriors fall outside of the confidence set with some positive probability.

More can be said about estimates on the value of information. In order to do so, we introduce an ε -neighborhood of the confidence set. For $\varepsilon > 0$, let

$$\Delta_{A,\varepsilon}^c(\bar{p}) = \{q \in \Delta \mid d(q, \Delta_A^c(\bar{p})) < \varepsilon\}. \quad (10)$$

This leads us to a first estimate of the value of information.

Theorem 1 For every ε , there exist positive constants $c_{A,\varepsilon}$ and C_A s.t., for every \mathbf{q} ,

$$C_A \mathbb{P}\{\mathbf{q} \notin \Delta_A^c(\bar{p})\} \geq \mathbf{VoI}_A(\mathbf{q}) \geq c_{A,\varepsilon} \mathbb{P}\{\mathbf{q} \notin \Delta_{A,\varepsilon}^c(\bar{p})\}. \quad (11)$$

On a en fait le résultat plus fort suivant qui devrait être la nouvelle forme du théorème:

Theorem 2 For every ε , there exist positive constants $c_{A,\varepsilon}$ and C_A s.t., for every \mathbf{q} ,

$$C_A \mathbb{E} d(A^*(\bar{p}), \mathbf{q}) \geq \mathbf{VoI}_A(\mathbf{q}) \geq c_{A,\varepsilon} \mathbb{P}\{\mathbf{q} \notin \Delta_{A,\varepsilon}^c(\bar{p})\}. \quad (12)$$

where by definition:

$$d(A^*(\bar{p}), q) = \inf_{p \in A^*(\bar{p})} \|p - q\|.$$

ce second résultat doit suivre du premier par la convexité de \mathbf{VoI} .

Michel : attention $p \in A^(\bar{p})$ n'a pas de sens*

Michel : voici ce que je suis capable de montrer

$$\left(\inf_{a' \in A^*(\bar{p})} \sup_{a \in A} \|a - a'\| \right) \mathbb{E} \left[\mathbf{1}_{\mathbf{q} \notin \Delta_A^c(\bar{p})} \|\mathbf{q} - \bar{p}\| \right] \geq \mathbf{VoI}_A(\mathbf{q}) \geq c_{A,\varepsilon} \mathbb{P}\{\mathbf{q} \notin \Delta_{A,\varepsilon}^c(\bar{p})\}. \quad (13)$$

3.2 Indifferences

We now consider situations in which information impacts the most actions. Those are situations of indifference in which, at the prior belief \bar{p} , the agent is *undecided* between several optimal actions. A small piece of information can then be enough to break this indifference. As shown by the following proposition, the value function then exhibits a *kink* at \bar{p} .

Proposition 2 The two conditions are equivalent:

- the set $A^*(\bar{p})$ of optimal actions at the prior belief \bar{p} contains more than one element;
- the value function V_A is non-differentiable (in the standard sense) at the prior belief \bar{p} .

At such beliefs \bar{p} , the convexity gap of the value function V_A is maximal in the directions in which it is non-differentiable. This allows us to derive a second bound on the value of information. Cases of indifference are typical of situations with a finite number of action choices.

We call *indifference kernel* $\Sigma_A^i(\bar{p})$ at \bar{p} the vector space of signed measures

$$\Sigma_A^i(\bar{p}) = [A^*(\bar{p}) - A^*(\bar{p})]^\perp. \quad (14)$$

In other words, beliefs in the indifference kernel $\Sigma_A^i(\bar{p})$ do not break any of the ties in $A^*(\bar{p})$.

We have the inclusion

$$\Delta_A^c(\bar{p}) \subset \Sigma_A^i(\bar{p}) \cap \Delta. \quad (15)$$

Theorem 3 *Bounds on the VoI for the undecided agent.*

There exist positive constants c_A and C_A such that, for every information structure \mathbf{q} ,

$$C_A \mathbb{E} \|\mathbf{q} - \bar{p}\| \geq \mathbf{VoI}_A(\mathbf{q}) \geq \mathbf{VoI}_{A^*(\bar{p})}(\mathbf{q}) \geq c_A \mathbb{E} \|\mathbf{q} - \bar{p}\|_{\Sigma_A^i(\bar{p})}, \quad (16)$$

where $\|\cdot\|_{\Sigma_A^i(\bar{p})}$ is a semi-norm with kernel $\Sigma_A^i(\bar{p})$, the indifference kernel in (14).

Recall that a semi-norm on the signed measures Σ on K , identified with \mathbb{R}^K , is a mapping $\|\cdot\| : \mathbb{R}^K \rightarrow \mathbb{R}_+$ which satisfies the requirements of a norm, except that the vector subspace $\{s \in \mathbb{R}^K \mid \|s\| = 0\}$ — called the *kernel* of the semi-norm $\|\cdot\|$ — is not necessarily reduced to zero.

The Theorem shows that a lower bound of the value of information is the expectation of a semi-norm of the distance between the prior belief and the posterior belief. To understand the role of the kernel $\Sigma_A^i(\bar{p})$ of this semi-norm, let us first consider the set of beliefs in this set. A posterior q is in $\Sigma_A^i(\bar{p}) = [A^*(\bar{p}) - A^*(\bar{p})]^\perp$ if and only if, for any two optimal actions $a, a' \in A^*(\bar{p})$, $\langle q, a \rangle = \langle q, a' \rangle$. In words, posteriors that do not break any of the ties in $A^*(\bar{p})$ might not be valuable to the agent. But, on the other hand, the Theorem tells us that all other directions — i.e., those that allow to break at least one of the ties in $A^*(\bar{p})$ — are valuable to the agent, and furthermore, in these directions, the value of information behaves like an expected distance from the prior to the posterior.

3.3 Flexible decisions

Finally, we consider the case in which there is a unique optimal action for each belief in the range considered, and this action depends in a certain smooth way on the belief. More precisely, we assume that around the prior, optimal actions depend on a 1-1 way on the belief in a certain differentiable way. This assumption is met when, for instance, the decision problem faced by the agent is a scoring rule, or an investment problem as in Cabrales, Gossner, and Serrano (2013).

Proposition 3 *Suppose that the action set A has boundary ∂A which is a C^2 submanifold of \mathbb{R}^K . The three following conditions are equivalent:*

1. *The set-valued mapping*

$$A^* : \Delta \rightrightarrows A, \quad p \mapsto A^*(p) \quad (17)$$

is a mapping¹ which is a local diffeomorphism at the prior belief \bar{p} ;

2. *The set $A^*(\bar{p})$ of optimal actions at the prior belief \bar{p} is reduced to a singleton at which the curvature of the action set A is positive;*
3. *The value function V_A is twice differentiable at the prior belief \bar{p} and the Hessian is definite positive.*

¹Meaning that the set $A^*(p)$ is a singleton for all $p \in \Delta$, in which case we identify singleton set and single element.

Theorem 4 *Bounds on the VoI for the flexible agent*

If the d.m. is flexible at \bar{p} , there exist positive constants c_A and C_A such that, for every information structure \mathbf{q} ,

$$C_A \mathbb{E} \|\mathbf{q} - \bar{p}\|^2 \geq \mathbf{VoI}_A(\mathbf{q}) \geq c_A \mathbb{E} \|\mathbf{q} - \bar{p}\|^2 . \quad (18)$$

4 Marginal value of information

Radner and Stiglitz (1984) study the question of the marginal value of information. They provide joint conditions on a parametrized family of information structures together with a decision problem such that, when the agent is close to receiving not information at all, the marginal value of information is null. Their result was subsequently generalized by Chade and Shlee (2002) and De Lara and Gilotte (2007) who also characterize joint conditions on parametrized information and a decision problem leading to zero marginal value of information.

In this section we show how our bounds on the value of information apply to the marginal value of information. In particular, we provide separate conditions on the decision problem and on the family of parametrized information structures that result in a null value of information. We then examine several parametrized families of information structures and rely on our main results to study how the marginal value of information varies depending on the decision problem faced.

Let $(\mathbf{q}^\theta)_{\theta>0}$ be a family of information structures. As in Radner and Stiglitz (1984), we are interested in the marginal value of information:

$$V^+ = \limsup_{\theta \rightarrow 0} \frac{1}{\theta} \mathbf{VoI}_A(\mathbf{q}^\theta) . \quad (19)$$

Proposition 4 *Assume either that*

1. $\mathbb{E} d(A^*(\bar{p}), \mathbf{q}^\theta) = o(\theta)$ *Question ‡ Olivier : trop rapide. Voir la nouvelle forme de la proposition sur "confident". Michel : voir mon commentaire sur la nouvelle forme de la proposition sur "confident".*
2. *The decision maker is flexible at \bar{p} and $\mathbb{E} \|\mathbf{q}^\theta - \bar{p}\|^2 = o(\theta)$ Michel : je suggère de remplacer par $\limsup_{\theta \rightarrow 0} \frac{1}{\theta} \mathbb{E} \|\mathbf{q}^\theta - \bar{p}\|^2 = 0$*

then $V^+ = 0$.

Michel : je ne comprends pas ce qui suit ; voir mon commentaire sur la nouvelle forme de la proposition sur "confident". The first condition is met automatically if $\mathbb{E} \|\mathbf{q}^\theta - \bar{p}\| = o(\theta)$. It is also met if, for instance, $A^*(\bar{p})$ has non-empty interior, and posteriors converge to the prior almost-surely.

The second case, when the decision maker is flexible, compares with the original Radner-Stiglitz assumptions. ² Michel : je suis quelque peu emb  t   car tant Radner and Stiglitz (1984) que Chade and Shlee (2002) font des hypoth  ses sur les actions optimales. En cons  quence, il est difficile de comparer avec ce que l'on fait.

Michel : Assumption (A0) in Radner and Stiglitz (1984) does not require that $A^*(q^\theta)$ be a singleton, for all θ . Therefore, flexibility is not required.

Michel : Chade and Shlee (2002) provide sufficient conditions that bear on the conditional distribution of the signal knowing the state of Nature. Is it easy to display their implication in term of posterior conditional distribution of the state of Nature knowing the signal?

Michel : Condition IIDV=0 in De Lara and Gilotte (2007) is equivalent to $\limsup_{\theta \rightarrow 0} \frac{1}{\theta} \mathbb{E} \|q^\theta - \bar{p}\| = 0$. We conclude thanks to la nouvelle forme de la proposition sur "confident" with the inequality that I propose

4.1 Examples

In this section we study the marginal value of information for several typical parametrized of information structures. In the first, information consists on the observation of a brownian motion with known variance and a drift that depends on the state of nature. In the second, information consists of the observation of a Poisson process whose probability of success depends on the state of nature. In these two well studied families in the learning literature, the natural parametrization of information is the length of the interval of time during which observation takes place. In our third example, the agent observes a binary signal and the marginal value of information depends on the asymptotic informativeness of these signals close to the situation without information.

In all three following examples we assume binary states of nature: $K = \{0, 1\}$. The d.m.'s prior belief on the state being 1 is denoted \bar{p} . Following the conditions under which we established bounds on the value of information, we call as "undecided" the case in which the decision problem faced by the d.m. is such that there is indifference between two actions at \bar{p} , "flexible" the case in which the optimal action is a smooth function of the belief in a neighborhood of \bar{p} , and "confident" the case in which there is a unique optimal action in an open interval of beliefs containing \bar{p} , and in this case we let (p_l, p_h) be the set of beliefs for which this action is the unique optimal one.

Our aim is to develop estimates of the marginal value of information. There are three possibilities: it can be infinite, null, or positive and finite. We denote these three cases by $V^+ = \infty$, $V^+ = 0$ and $V^+ \simeq 1$ respectively.

Example 2 (Brownian motion) *Frameworks in which agents observe a Brownian motion with known volatility and unknown drift include Bergemann and V  lim  ki (1997), Keller and*

²Michel, vu que c'est un r  sultat    la RS, c'est important de dire comment les hypoth  ses peuvent se comparer avec celles de RS et des autres papiers, Gilotte, de Lara, Chade Schlee etc. Je te laisse dire quelques mots car c'est toi le sp  cialiste.

Rady (1999), Bolton and Harris (1999) as well reputation models Faingold and Sannikov (2011).

Assume the d.m. observes the realisation of a Brownian motion with variance 1 and drift $k \in \{0, 1\}$

$$d\mathbf{Z}_t = kdt + d\mathbf{B}_t \quad (20)$$

for a small interval of time $\theta > 0$.

If we let \mathbf{q}^t be the posterior belief at time t , it is well known that \mathbf{q}^t follows an equation of the form:

$$d\mathbf{q}^t = \mathbf{q}^t(1 - \mathbf{q}^t)d\mathbf{w}_t$$

where \mathbf{w} is a standard Brownian process, cf. for instance Lemma 1 in Bolton and Harris (1999) or Lemma 2 in Faingold and Sannikov (2011).

Thus, for small values of θ ,

$$\begin{aligned} \mathbb{E} \|\mathbf{q}^\theta - p\| &\sim \sqrt{\theta} \\ \mathbb{E} \|\mathbf{q}^\theta - p\|^2 &\sim \theta \end{aligned}$$

It follows from Theorems 2-4 that the marginal value of information is characterized, depending on the decision problem faced, as:

1. In the confident case, $V^+ = 0$
2. In the flexible case, $V^+ \simeq 1$
3. In the undecided case, $V^+ = \infty$

Example 3 (Exponential learning) Exponential learning plays a central role in models of strategic experimentation such as Keller, Rady, and Cripps (2005). Assume the d.m. observes a Poisson process with intensity ρ_k during a small interval of time $\theta > 0$, where $\rho_0 < \rho_1$. The probability of two success is negligible compared to the probability of one success (of order θ^2 compared to θ). A success leads to a posterior that converges, as $\theta \rightarrow 0$, from below to

$$q^+ = \frac{\bar{p}\rho_1}{\bar{p}\rho_1 + (1 - \bar{p})\rho_0} > \bar{p} \quad (21)$$

and happens with probability $\sim \theta$. In the absence of success, the posterior belief converges to \bar{p} as $\theta \rightarrow 0$.

In particular:

$$\begin{aligned} \mathbb{E} \|\mathbf{q}^\theta - p\| &\sim \theta \\ \mathbb{E} \|\mathbf{q}^\theta - p\|^2 &\sim \theta \end{aligned}$$

Hence the following estimates on the marginal value of information:

1. In the confident case,

(a) $V^+ \simeq 1$ if $q^+ > p_h$

(b) $V^+ \simeq 0$ if $q^+ \leq p_h$

2. In the flexible case, $V^+ \simeq 1$

3. In the undecided case, $V^+ \simeq 1$

Example 4 (Equally likely signals) In our third example, we consider binary and equally signals, which lead to a “split” of beliefs around the prior \bar{p} . Depending on the precision of these signals as a function of θ , the posterior beliefs are $p \pm \theta^\alpha$ for a certain parameter $\alpha > 0$. Lower values of α correspond to more spread out beliefs around the prior, hence to more accurate information.

In this case we easily compute:

$$\mathbb{E} \|\mathbf{q}^\theta - p\| = \theta^\alpha \quad (22)$$

$$\mathbb{E} \|\mathbf{q}^\theta - p\|^2 = \theta^{2\alpha} \quad (23)$$

Here again, the marginal value of information is deduced from Theorems 2–4.

1. In the confident case, $V^+ = 0$

2. In the flexible case,

(a) $V^+ = \infty$ if $\alpha < \frac{1}{2}$

(b) $V^+ \simeq 1$ if $\alpha = \frac{1}{2}$

(c) $V^+ = 0$ if $\alpha > \frac{1}{2}$

3. In the undecided case,

(a) $V^+ = \infty$ if $\alpha < 1$

(b) $V^+ \simeq 1$ if $\alpha = 1$

(c) $V^+ = 0$ if $\alpha > 1$

The following Table 2 summarises the marginal value of information in all previous examples.

In all the cases except one, the marginal value of information is completely determined by the local behaviour of the value function around the prior. For the Poisson case, the marginal value of information is 0 or positive, depending on whether the observation of a success is sufficient to lead to a decision reversal.

The marginal value of information is always weakly higher in the smooth case than in the kink case, and weakly higher in the kink case than in other cases. In the flat case, the marginal value of information is null, except in the Poisson case with $p^+ > p_h$. This is driven by the fact that, in all other cases, posteriors are, with high probability, too close to the prior to lead to a decision reversal. In the kink situation, the marginal value of information

V^+	confident	flexible	undecided
Poisson	0 or 1	1	1
Brownian	0	1	∞
EL, $\alpha < \frac{1}{2}$	0	∞	∞
EL, $\alpha = \frac{1}{2}$	0	1	∞
EL, $\frac{1}{2} < \alpha < 1$	0	0	∞
EL, $\alpha = 1$	0	0	1
EL, $\alpha > 1$	0	0	0

Table 2: Marginal value of information in the different examples. EL stands for for the equally likely signals case, 1 represents positive marginal value of information.

is always positive or infinite, except for sufficiently uninformative binary signals ($\alpha > 1$). Finally, in the smooth case, the most representative of decision problems with a continuum of actions, the value of information is positive or infinite, except with quite uninformative binary signals ($\alpha > 1/2$).

5 An insurance example

In this example we study the incentives of an insurree to gather information on her risk level.

Example 5 *The model is drawn from the classical insurance framework (Bernoulli, 1738; Eeckhoudt, Gollier, and Schlesinger, 2005).*

An insurree faces the decision of partially or fully insuring a good of value V against the possibility of its total loss. Pricing is linear, for an indemnity I , the insurance company charges

$$P(I) = \alpha I + f , \quad (24)$$

for some parameters $\alpha, f > 0$. In exchange for the premium $P(I)$, the insurree gets compensated of an amount I by the insurance company in case of a loss. For the range of wealth considered, the insurree's utility function is considered to have constant absolute risk aversion ρ . The insurree's subjective perception that a loss may arise is p . The insurree chooses either not to insure, which is equivalent to insuring for an indemnity of 0 at zero cost, and obtains expected utility

$$U_0(p) = 1 - pe^{-\rho(-P(I)+I)} - (1 - p)e^{-\rho(V-P(I))} ,$$

or to insure for an indemnity $I > 0$ that maximizes the expected utility

$$U(p, I) = 1 - pe^{-\rho(-P(I)+I)} - (1 - p)e^{-\rho(V-P(I))} . \quad (25)$$

Assuming a positive level of insurance is taken, the problem's FOC gives a unique solution:

$$\hat{I}(p) = V - \frac{1}{\rho} \ln\left(\frac{1-p}{p} \frac{\alpha}{1-\alpha}\right) .$$

The question now becomes whether the level $\hat{I}(p)$ or no insurance is chosen.

Proposition 5 *There exists a threshold belief p^* such that*

1. for $p < p^*$, it is optimal not to insure,
2. for $p = p^*$, the insuree is indifferent between no insurance and insurance at a positive indemnity level $\hat{I}(p^*)$.
3. for $p > p^*$, it is optimal to insure at an indemnity level $\hat{I}(p)$.

Proof. It is optimal to insure at the level $\hat{I}(p)$ when $U(p, \hat{I}(p)) \geq U_0$. The difference

$$U(p, \hat{I}(p)) - U_0 = p + (1 - p)e^{-\rho V} - pe^{-\rho(-P(\hat{I}(p)) + \hat{I}(p))} - (1 - p)e^{-\rho(V - P(\hat{I}(p)))}$$

is increasing in p and negative for $p = 0$. Hence $U(p, \hat{I}(p)) \geq U_0$ is equivalent to $p \geq p^*$ for a certain $p^* > 0$. ■

Now we assume that the insuree has access to a small piece of information concerning her probability of loss. Once informed, she discovers that probability q of a loss is in fact either $p - \varepsilon$ or $p + \varepsilon$, where both possibilities are equally likely and ε is a small number. Let $V(q)$ be the utility of the insuree with beliefs q , once the optimal policy is chosen:

$$V(q) = \max \left\{ \max_{I \geq 0} U(q, I), U_0(p) \right\}. \quad (26)$$

The Value of Information in the decision problem is defined as the expected utility with the information minus the expected utility absent the information:

$$\mathbf{VoI}(\varepsilon) = \frac{1}{2}V(p + \varepsilon) + \frac{1}{2}V(p - \varepsilon) - V(p) \quad (27)$$

Note that $\mathbf{VoI}(\varepsilon)$ measures the value of information in terms of the utility, the equivalent measure in monetary terms would be $\frac{1}{\rho} \ln(1 + \mathbf{VoI}(\varepsilon))$. The following proposition characterizes the value of a small amount of information, in terms of the agent's optimal insurance behavior.

Proposition 6 *Depending on p , the value of information for small ε behaves as follows:*

Confident For $p < p^*$, $\mathbf{VoI}(\varepsilon) = 0$ for small ε ,

Undecided For $p = p^*$, $\mathbf{VoI}(\varepsilon) \sim \bar{C}\varepsilon$ for a constant $C^* > 0$,

Flexible For $p > p^*$, $\mathbf{VoI}(\varepsilon) \sim C(\rho)\varepsilon^2$ for a constant $C(p) > 0$.

Proof. The confident and undecided cases are immediate consequences of Theorems 2 and 3, together with Proposition 6. In the flexible case, the optimal insurance level is given by $\hat{I}(p)$, and is differentiable with p with $\frac{\partial \hat{I}(p)}{\partial p} \neq 0$. The set optimal actions $A^*(p)$ is reduced to the singleton point

$$A^*(p) = (1 - \exp^{-\rho(V-P(\hat{I}(p)))}, 1 - \exp^{-\rho(\hat{I}(p)-P(\hat{I}(p)))}) ,$$

where by convention the first coordinate corresponds to no loss and the second corresponds to the loss. As $p \mapsto A^*(p)$ is a local diffeomorphism on $\hat{I}(p)$, it is a local diffeomorphism³ at p and Theorem 4 applies. ■

The results from the proposition are intuitive. First, a small piece of information is valueless if the agent is not taking insurance. For such agents, small information does not impact behavior as even bad news is not enough to trigger insurance purchase. For an undecided agent who is indifferent between no insurance and insurance at a positive level $I(p^*)$, a small piece of information will be enough to break the indifference and influence significantly her behavior, and this is the situation in which information is the most valuable. Finally, for an agent who takes a positive level of insurance, information may impact the level of insurance chosen. But, because the change of insurance level is itself of order ε , and that the insurance level $I(p^*)$ is ε -optimal at the posterior, the value of information is a second order in ε .

Figure 2 represents the set A and the corresponding value function. In the representation of A , the horizontal axis corresponds to the payoff without accident, and the vertical axis to the payoff in case of an accident. The red dot to the right corresponds to the choice of no insurance; it maximizes payoff in case of no loss. The blue curve corresponds to the set of payoffs that achieved by different coverage levels. Finally A is the convex hull of this set of points. For lower values of p , the value function is linear as the insuree chooses not to take insurance. At p^* which is approximately 0.334, the value function exhibits a kink, and the agent is indifferent between no insurance and a positive level of insurance. Finally, for larger values of p , the V function is twice continuously differentiable with positive second derivative.

6 Related Literature

The value of information in decision problems is a well-studied question in statistics. The central work in this area is Blackwell (1953) that defines a source of information α as *more informative* than another one, β , whenever all agents, independently of their preferences and decision problems faced, weakly prefer α to β . Blackwell (1953) characterises precisely this relationship in the following terms: α is more informative than β if and only if information from β can be obtained as a garbling of the information from α .

The requirement that all agents agree on their preferences between two statistical experiments is a powerful one. It implies that this ranking is incomplete, as many such pairs of experiments cannot be ranked according to this ordering. Some authors have been looking

³See footnote 1 on the meaning of A^* being a local diffeomorphism at \bar{p} .

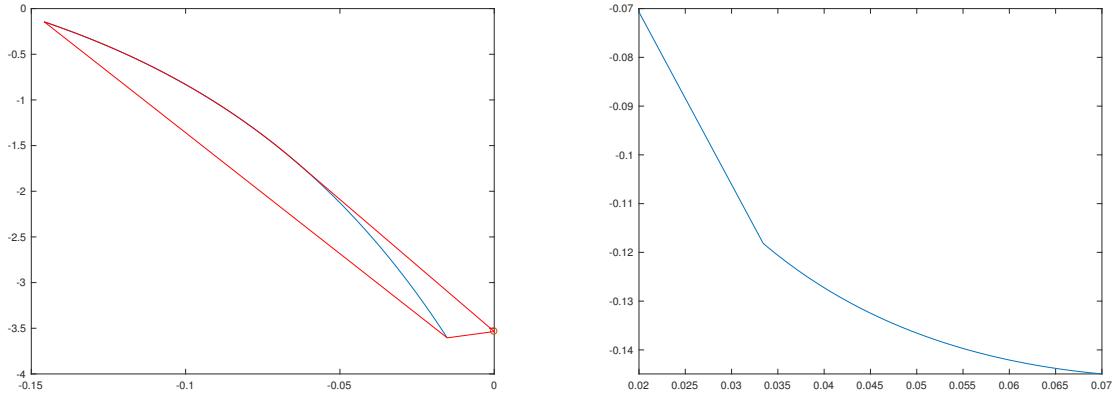


Figure 2: The set A on the left and the corresponding value function for the insurance example on the right. $\alpha = 0.08, V = 1000, p = 10$

at sub-classes of decision problems in order to obtain rankings that would be more complete than Blackwell's. For instance, Lehmann (1988); Persico (2000); Athey and Levin (2017) restrict attention to classes of decision problems that generate different classes of monotone decision rules. Focusing on investment decision problems, Cabrales, Gossner, and Serrano (2013) obtain and characterise a complete ranking of information sources based on an uniform criterion, while Cabrales, Gossner, and Serrano (2017) use a duality approach to characterise the value of an information purchase that consists of an information structure with a price attached to it.

The present work departs from this literature in the sense that we focus on the value of information for a given agent, instead of trying to measure the value of information independently of the agent. Gilboa and Lehrer (1991) and Azrieli and Lehrer (2008) characterise the possible preferences for information that any agent can have, letting the decision problem vary and the agent's preferences vary. Of course, the preferences of any given agent satisfies their conditions.

Radner and Stiglitz (1984); Chade and Shlee (2002); De Lara and Gilotte (2007) study the question of marginal value of information. They consider parametrised information structures, and derive general conditions on the pair consisting of the information structures and the decision problem under which the marginal value of information close to no information is zero. Our work contributes to this question by allowing to derive estimates on the value of information based on separate conditions on the decision problem and on the information structure. This is the approach we have taken in Section 4.

7 Conclusion

We formalized the natural duality between the set of available choices to a decision maker and the value function expressed as a function of her belief. This, in turn, allowed us to derive bounds on the value of any piece of information that are based solely on local properties on the agent's behavior around her prior. Finally, we have provided applications to the question of marginal value of information, as well as an insurance example.

References

- ARROW, K. J. (1971): "The value of and demand for information," in *Decision and Organization*, ed. by C. McGuire, and R. Radner, pp. 131–139, Amsterdam. North-Holland.
- ATHEY, S., AND J. LEVIN (2017): "The value of information in monotone decision problems," *Research in Economics*.
- AUBIN, J.-P. (1982): *Mathematical Methods of Game and Economic Theory*. North-Holland, Amsterdam, second edn.
- AUMANN, R., AND M. MASCHLER (1967): "Repeated games with incomplete information: A survey of recent results," in *Reports to the U.S. Arms Control and Disarmament Agency, ST-116*, pp. 287–403.
- AUMANN, R. J., AND S. HART (1986): "Bi-convexity and bi-martingales," *Israel Journal of Mathematics*, 54, 159–180.
- AZRIELI, Y., AND E. LEHRER (2008): "The value of a stochastic information structure," *Games and Economic Behavior*, 63, 679–693.
- BERGEMANN, D., AND J. VÄLIMÄKI (1997): "Market diffusion with two-sided learning," *The RAND Journal of Economics*, pp. 773–795.
- BERNOULLI, D. (1738): "Specimen theoriae novae de mensura sortis," in *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, vol. 5, pp. 175–192.
- BLACKWELL, D. (1953): "Equivalent comparison of experiments," *Annals of Mathematical Statistics*, 24, 265–272.
- BOHNENBLUST, H. F., L. S. SHAPLEY, AND S. SHERMAN (1949): "Reconnaissance in Game Theory," Discussion Paper RM-208, The RAND corporation.
- BOLTON, P., AND C. HARRIS (1999): "Strategic experimentation," *Econometrica*, 67(2), 349–374.
- BRIER, G. W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 78(1-3).

- CABRALES, A., O. GOSSNER, AND R. SERRANO (2013): “Entropy and the value of information for investors,” *American Economic Review*, 103, 360–377.
- CABRALES, A., O. GOSSNER, AND R. SERRANO (2017): “A normalized value for information purchases,” *Journal of Economic Theory*, 170, 266–288.
- CHADE, H., AND E. SHLEE (2002): “Another look at the Radner-Stiglitz Nonconcavity in the Value of Information,” *Journal of Economic Theory*, 107, 421–452.
- DE LARA, M., AND L. GILOTTE (2007): “A tight sufficient condition for Radner–Stiglitz nonconcavity in the value of information,” *Journal of Economic Theory*, 137(1), 696–708.
- ECKHOUDT, L., C. GOLLIER, AND H. SCHLESINGER (2005): *Economic and Financial Decisions under Risk*. Princeton University Press.
- FAINGOLD, E., AND Y. SANNIKOV (2011): “Reputation in Continuous-Time Games,” *Econometrica*, 79, 773–876.
- GILBOA, I., AND E. LEHRER (1991): “The value of information - an axiomatic approach,” *Journal of Mathematical Economics*, 20, 443–459.
- HIRIART-URURTY, J. B., AND C. LEMARÉCHAL (1993): *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, Berlin.
- HOTELLING, H. (1932): “Edgeworth’s taxation paradox and the nature of demand and supply functions,” *Journal of Political Economy*, 40(5), 577–616.
- KAMENICA, E., AND M. GENTZKOZ (2011): “Bayesian persuasion,” *American Economic Review*, 101, 2590–2615.
- KELLER, G., AND S. RADY (1999): “Optimal Experimentation in a Changing Environment,” *Review of Economic Studies*, 66(3), 475–507.
- KELLER, G., S. RADY, AND M. CRIPPS (2005): “Strategic experimentation with exponential bandits,” *Econometrica*, 73(1), 39–68.
- LEHMANN, E. L. (1988): “Comparing location experiments,” *The Annals of Statistics*, 16(2), 521–533.
- MCFADDEN, D. (1973): “Conditional logit analysis of qualitative choice behavior,” in *Frontiers in Econometrics*, ed. by P. Zarembka. Academic Press.
- (1978): “Modelling the choice of residential location,” in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. North-Holland.
- PERSICO, N. (2000): “Information acquisition in auctions,” *Econometrica*, 68(1), 135–148.

RADNER, R., AND J. STIGLITZ (1984): "A nonconcavity in the value of information," in *Bayesian Models of Economic Theory*, ed. by M. Boyer, and R. Kihlstrom, pp. 33–52, Amsterdam. Elsevier.

SCHNEIDER, R. (2014): *Convex bodies: the Brunn-Minkowski theory*. Cambridge University Press, second edn.

A Convex analysis and geometry background

Let $A \subset \mathbb{R}^K$ be nonempty, convex and compact subset of \mathbb{R}^K , called the *action set*. We denote by Σ the set of *signed measures on K* , identified with \mathbb{R}^K . The set Δ of probability distribution on K is a convex subset of the set Σ , identified with the simplex of \mathbb{R}^K .

A.1 Convex analysis

Support function. The *support function* σ_A of the action set A is defined by

$$\sigma_A(s) = \sup_{a \in A} \langle s, a \rangle, \quad \forall s \in \Sigma. \quad (28)$$

The value function $V_A : \Delta \rightarrow \mathbb{R}$ in (4) is the restriction of the support function σ_A to probability distributions:

$$V_A(p) = \sigma_A(p), \quad \forall p \in \Delta. \quad (29)$$

It is well known that σ_A is convex (as the supremum of the family of linear maps $\langle \cdot, a \rangle$ for $a \in A$). As the action set A is compact, $\sigma_A(s)$ takes finite values, hence its domain is Σ , hence σ_A is continuous.

(Exposed) face. For any signed measure $s \in \Sigma$, we let

$$F_A(s) = \arg \max_{a' \in A} \langle s, a' \rangle = \{a \in A \mid \forall a' \in A, \langle s, a' \rangle \leq \langle s, a \rangle\} \subset A \quad (30)$$

be the set of maximizers of $a \mapsto \langle s, a \rangle$ over A . We call $F_A(s)$ the *(exposed) face of A in the direction $s \in \Sigma$* . As the action set A is convex and compact, the face $F_A(s)$ of A in the direction s is nonempty, for any $s \in \Sigma$, and the face is a subset of the *boundary* ∂A of A :

$$F_A(s) \subset \partial A, \quad \forall s \in \Sigma. \quad (31)$$

The set $A^*(p)$ of optimal actions under belief p in (7) coincides with the (exposed) face $F_A(p)$ of A in the direction p in (30):

$$A^*(p) = F_A(p), \quad \forall p \in \Delta. \quad (32)$$

Normal cone. For any payoff vector a in A , we define

$$N_A(a) = \{s \in \Sigma \mid \forall a' \in A, \langle s, a' \rangle \leq \langle s, a \rangle\} \subset \Sigma. \quad (33)$$

We call $N_A(a)$ the *normal cone* to the closed convex set A at $a \in A$. Notice that $N_A(a)$ is made of signed measures, that are not necessarily beliefs. The set $\Delta_A^*(a)$ of beliefs compatible with optimal action a in (8) is related to the normal cone $N_A(a)$ at a in (33) by:

$$\Delta_A^*(a) = N_A(a) \cap \Delta, \quad \forall a \in A. \quad (34)$$

Conjugate subsets of actions and beliefs. Exposed face F_A and normal cone N_A are conjugate as follows:

$$s \in \Sigma \text{ and } a \in F_A(s) \iff a \in A \text{ and } s \in N_A(a). \quad (35)$$

Some results.

Proposition 7 *Let $C \subset \mathbb{R}^K$ be a nonempty convex set.*

1. *Let $X \subset C$ be nonempty. Let $Y \subset \mathbb{R}^K$ be nonempty. Then*

$$X \subset \bigcap_{y \in Y} F_C(y) \iff Y \subset \bigcap_{x \in X} N_C(x) \iff \sigma_C(y) = \langle y, x \rangle, \quad \forall x \in X, \quad \forall y \in Y. \quad (36)$$

2. *Let y_1, \dots, y_n belong to \mathbb{R}^K . If there exists $s_1 > 0, \dots, s_n > 0$ with $s_1 + \dots + s_n = 1$ such that*

$$\sigma_C(s_1 y_1 + \dots + s_n y_n) = s_1 \sigma_C(y_1) + \dots + s_n \sigma_C(y_n), \quad (37)$$

then

$$F_C(s_1 y_1 + \dots + s_n y_n) = F_C(y_1) \cap \dots \cap F_C(y_n) = \bigcap_{y \in \overline{\text{co}}(y_1, \dots, y_n)} F_C(y). \quad (38)$$

3. *Let $Y \subset \mathbb{R}^K$ be nonempty convex. The support function σ_C is linear on $\text{ri}(Y)$ if and only if*

$$F_C(y) = F_C(y') \neq \emptyset, \quad \forall (y, y') \in \text{ri}(Y)^2. \quad (39)$$

4. *Let $y \in \mathbb{R}^K$ be such that $F_C(y) \neq \emptyset$. Then, we have*

$$\sigma_C(y') - \sigma_C(y) \geq \sigma_{F_C(y)}(y' - y), \quad \forall y' \in \mathbb{R}^K. \quad (40)$$

5. *The function σ_{C-C} is a semi-norm with kernel $[C - C]^\perp$.*

Proof. Exposed face and normal cone are conjugate as follows (Hiriart-Uruty and Lemaréchal, 1993, p.220):

$$x \in C \text{ and } \langle y, x \rangle = \sigma_C(y) \iff x \in \underset{x' \in C}{\text{argmax}} \langle y, x' \rangle \quad (41a)$$

$$\iff x \in F_C(y) \quad (41b)$$

$$\iff x \in C \text{ and } y \in N_C(x). \quad (41c)$$

1. We have

$$\begin{aligned}
X \subset \bigcap_{y \in Y} F_C(y) &\iff x \in F_C(y), \forall x \in X, \forall y \in Y \\
&\iff y \in N_C(x), \forall x \in X, \forall y \in Y \quad \text{by (41) as } X \subset C \\
&\iff Y \subset \bigcap_{x \in X} N_C(x)
\end{aligned}$$

2. We suppose that $F_C(s_1y_1 + \dots + s_ny_n) \neq \emptyset$, else there is nothing to show. Let us put $\bar{y} = s_1y_1 + \dots + s_ny_n$ and let $\bar{x} \in F_C(\bar{y})$. By assumption, we have

$$\langle \bar{y}, \bar{x} \rangle = \sigma_C(\bar{y}) = s_1\sigma_C(y_1) + \dots + s_n\sigma_C(y_n). \quad (42)$$

Now, as $\bar{x} \in F_C(\bar{y}) \subset C$, we have that

$$\langle y_1, \bar{x} \rangle \leq \sigma_C(y_1), \dots, \langle y_n, \bar{x} \rangle \leq \sigma_C(y_n), \quad (43)$$

by definition of σ_C . If one of these inequalities were strict, we would deduce the strict inequality $\langle \bar{y}, \bar{x} \rangle < s_1\sigma_C(y_1) + \dots + s_n\sigma_C(y_n)$, which is false. Therefore, we have

$$\langle y_1, \bar{x} \rangle = \sigma_C(y_1), \dots, \langle y_n, \bar{x} \rangle = \sigma_C(y_n), \quad (44)$$

that is,

$$\bar{x} \in F_C(y_1) \cap \dots \cap F_C(y_n). \quad (45)$$

As we have seen that $F_C(y_1) \cap \dots \cap F_C(y_n) = \bigcap_{y \in \overline{\text{co}}(y_1, \dots, y_n)} F_C(y)$, and as $\bigcap_{y \in \overline{\text{co}}(y_1, \dots, y_n)} F_C(y) \subset F_C(\bar{y})$, we conclude that

$$F_C(s_1y_1 + \dots + s_ny_n) = F_C(y_1) \cap \dots \cap F_C(y_n) = \bigcap_{y \in \overline{\text{co}}(y_1, \dots, y_n)} F_C(y). \quad (46)$$

3. Suppose that the support function σ_C is linear on $\text{ri}(Y)$. Let $(y, y') \in \text{ri}(Y)^2$. There is an open segment inside $\text{ri}(Y)$ that contains the pair $\{y, y'\}$. By item 2 and (38) — using the endpoints of the open segment — we obtain that $F_C(y) = F_C(y')$. Why is it $\neq \emptyset$?

Suppose that $F_C(y) = F_C(y') \neq \emptyset$. Then, for any $y' \in \text{ri}(Y)$ and $x' \in F_C(y')$, we have $\sigma_C(y) = \langle x', y \rangle$ for all $y \in \text{ri}(Y)$.

4. The subdifferential of the support function σ_C of the (nonempty) closed convex set $C \subset \mathbb{R}^K$ at $y \in \mathbb{R}^K$ is (Aubin, 1982, p.107), (Hiriart-Uruty and Lemaréchal, 1993, p.258)

$$\partial\sigma_C(y) = F_C(y) = \underset{x \in C}{\text{argmax}} \langle y, x \rangle, \quad \forall y \in \mathbb{R}^K. \quad (47)$$

If $\partial\sigma_C(y) \neq \emptyset$, then (40) is a consequence of the definition of the subdifferential.

5. The support function σ_{C-C} is homogeneous and sublinear, and it is nonnegative since $0 \in C - C$. As a consequence, σ_{C-C} is a semi-norm. The kernel is easily calculated.

■

A.2 Geometric convex analysis

The nonempty, convex and compact set $A \subset \mathbb{R}^K$ is called a *convex body* of \mathbb{R}^K (Schneider, 2014, p. 8).

Regular points and smooth bodies. We say that a point $a \in A$ is *smooth* or *regular* (Schneider, 2014, p. 83) if the normal cone $N_A(a)$ is reduced to a half-line. The *set of regular points* is denoted by $\text{reg}(A)$:

$$a \in \text{reg}(A) \iff \exists s \in \Sigma, \quad s \neq 0, \quad N_A(a) = \mathbb{R}_+ s. \quad (48)$$

Notice that a regular point a necessarily belongs to the boundary ∂A of A : $\text{reg}(A) \subset \partial A$. The body A is said to be *smooth* if all boundary points of A are regular ($\text{reg}(A) = \partial A$); in that case, it can be shown that its boundary ∂A is a C^1 submanifold of \mathbb{R}^K (Schneider, 2014, Theorem 2.2.4, p. 83).

Spherical image map of A . We denote by $S^{|K|-1}$ the unit sphere of the signed measures Σ on K (identified with \mathbb{R}^K with its canonical scalar product):

$$S^{|K|-1} = \{s \in \Sigma, \quad \|s\| = 1\}. \quad (49)$$

By (48), we have that

$$a \in \text{reg}(A) \iff \exists! s \in S^{|K|-1}, \quad N_A(a) = \mathbb{R}_+ s. \quad (50)$$

If a point $a \in A$ is regular, the unique outer normal unitary vector to A at a is denoted by $n_A(a)$, so that $N_A(a) = \mathbb{R}_+ n_A(a)$. The mapping

$$n_A : \text{reg}(A) \rightarrow S^{|K|-1}, \quad \text{where } \text{reg}(A) \subset \partial A \quad (51)$$

is called the *spherical image map of A* , or the *Gauss map*, and is continuous (Schneider, 2014, p. 88). We have

$$a \in \text{reg}(A) \Rightarrow N_A(a) = \mathbb{R}_+ n_A(a) \quad \text{where } n_A(a) \in S^{|K|-1}. \quad (52)$$

Reverse spherical image map of A . We say that a unit signed measure $s \in S^{|K|-1}$ is *regular* (Schneider, 2014, p. 87) if the (exposed) face $F_A(s)$ of A in the direction s , as defined in (30), is reduced to a singleton. The *set of regular unit signed measures* is denoted by $\text{regn}(A)$:

$$s \in \text{regn}(A) \iff s \in S^{|K|-1} \quad \text{and} \quad \exists! a \in A, \quad F_A(s) = \{a\}. \quad (53)$$

For a regular unit signed measure $s \in S^{|K|-1}$, we denote by $f_A(s)$ the unique element of $F_A(s)$, so that $F_A(s) = \{f_A(s)\}$. The mapping

$$f_A : \text{regn}(A) \rightarrow \partial A, \quad \text{where } \text{regn}(A) \subset S^{|K|-1} \quad (54)$$

is called the *reverse spherical image map of A* , and is continuous (Schneider, 2014, p. 88). We have

$$s \in \text{regn}(A) \Rightarrow F_A(s) = \{f_A(s)\}. \quad (55)$$

Bodies with C^2 surface.

Proposition 8 ((Schneider, 2014, p. 113)) *If the body A has boundary ∂A which is a C^2 submanifold of \mathbb{R}^K , then*

- *all points $a \in \partial A$ are regular ($\text{reg}(A) = \partial A$),*
- *the spherical image map n_A in (51) is defined over the whole boundary ∂A and is of class C^1 ,*
- *the spherical image map n_A has the reverse spherical image map f_A as right inverse, that is,*

$$n_A \circ f_A = \text{Id}_{\text{regn}(A)}. \quad (56)$$

Proof. The first two items can be found in (Schneider, 2014, p. 113). Now, we prove that $n_A \circ f_A = \text{Id}_{\text{regn}(A)}$. As $f_A : \text{regn}(A) \rightarrow \partial A$ by (54), and as $n_A : \partial A \rightarrow S^{|K|-1}$ by (51) since $\text{reg}(A) = \partial A$, the mapping $n_A \circ f_A : \text{regn}(A) \rightarrow S^{|K|-1}$ is well defined. Let $s \in \text{regn}(A)$. By (55), we have that $F_A(s) = \{f_A(s)\}$ and by (52), we have that $N_A(f_A(s)) = \mathbb{R}_+ n_A(f_A(s))$. From (35) — stating that exposed face and normal cone are conjugate — we deduce that $s \in \mathbb{R}_+ n_A(f_A(s))$. As $s \in S^{|K|-1}$, we conclude that $s = n_A(f_A(s))$ by (51). ■

Weingarten map. Let $a \in \text{reg}(A)$ be a regular point such that the spherical image map n_A in (51) is differentiable at a , with differential denoted by $T_a n_A$. The *Weingarten map* (Schneider, 2014, p. 113)

$$T_a n_A : T_a \partial A \rightarrow T_{n_A(a)} S^{|K|-1} \quad (57)$$

linearly maps the tangent space $T_a \partial A$ of the boundary ∂A at point a into the tangent space $T_{n_A(a)} S^{|K|-1}$ of the sphere $S^{|K|-1}$ at $n_A(a)$. The eigenvalues of the Weingarten map at a are called the *principal curvatures* of A at a (Schneider, 2014, p. 114); they are nonnegative (Schneider, 2014, p. 115). By definition, the body A has *positive curvature* at a if all principal curvatures at a are positive or, equivalently, if the Weingarten map is of maximal rank at a (Schneider, 2014, p. 115).

Reverse Weingarten map. Let $s \in \text{regn}(A)$ be a regular unit signed measure such that the reverse spherical image map f_A in (54) is differentiable at s , with differential denoted by $T_s f_A$. The *reverse Weingarten map*

$$T_s f_A : T_s S^{|K|-1} \rightarrow T_{f_A(s)} \partial A \quad (58)$$

maps the tangent space $T_s S^{|K|-1}$ of the sphere $S^{|K|-1}$ at s into the tangent space $T_{f_A(s)} \partial A$ of the boundary ∂A at point $f_A(s)$. The eigenvalues of the reverse Weingarten map at s are called the *principal radii of curvature* of A at s .

B Proofs of the results in Sect. 3

Using the relations (32) and (34), we express the proofs in the set of signed measures, i.e. in terms of the sets $F_A(p)$ and $N_A(a)$, instead of in the set of probability measures, i.e. in terms of $A^*(p)$ and $\Delta_A^*(a)$.

Confidence set and indifference kernel.

The confidence set $\Delta_A^c(\bar{p})$ of (9) is the nonempty closed and convex set

$$\Delta_A^c(p) = \bigcap_{a \in F_A(p)} N_A(a) \cap \Delta = \bigcap_{a \in A^*(p)} \Delta_A^*(a). \quad (59)$$

The indifference kernel $\Sigma_A^i(\bar{p})$ of (14) is the nonempty vector subspace

$$\Sigma_A^i(p) = [F_A(p) - F_A(p)]^\perp = [A^*(p) - A^*(p)]^\perp = \bigcap_{a \in F_A(p)} N_{F_A(p)}(a). \quad (60)$$

We have the inclusion:

$$\Delta_A^c(p) \subset \Sigma_A^i(p) \cap \Delta. \quad (61)$$

The above inclusion is strict in general. Indeed, consider a case where $F_A(p)$ is a singleton $\{a\}$. Then, on the one hand, $\Delta_A^c(p) = N_A(a) \cap \Delta$. However, on the other hand

$$\Sigma_A^i(p) \cap \Delta = N_{F_A(p)}(a) \cap \Delta = N_{\{a\}}(a) \cap \Delta = \Delta.$$

As soon as $N_A(a) \cap \Delta$ only contains the belief p , we have that $\{p\} = \Delta_A^c(p)$ and $\Sigma_A^i(p) \cap \Delta = \Delta$. As an example, consider the case where the set A is the unit ball:

$$A = B(0, 1), \quad \sigma_A(s) = \|s\|, \quad N_A(a) = \mathbb{R}_+ a, \quad F_A(s) = \left\{ \frac{s}{\|s\|} \right\}, \quad N_{F_A(s)}(a) = \mathbb{R}^2, \quad (62)$$

so that $\Delta_A^c(p) = \{p\}$, $\Sigma_A^i(p) \cap \Delta = \Delta$.

Here is another characterization of the confidence set $\Delta_A^c(\bar{p})$.

Proposition 9 *We have that*

$$\sigma_A(p) - \sigma_A(\bar{p}) - \langle p - \bar{p}, a \rangle = 0, \quad \forall a \in F_A(\bar{p}) \iff p \in \Delta_A^c(\bar{p}). \quad (63)$$

Proof. We have that

$$\sigma_A(p) - \sigma_A(\bar{p}) - \langle p - \bar{p}, a \rangle = 0, \quad \forall a \in F_A(\bar{p}) \iff \sigma_A(p) = \langle p, a \rangle, \quad \forall a \in F_A(\bar{p}) \quad (64a)$$

because $\sigma_A(\bar{p}) = \langle \bar{p}, a \rangle$ for any $a \in F_A(\bar{p})$

$$\iff p \in \bigcap_{a \in F_A(\bar{p})} N_A(a) \quad (64b)$$

by (36)

$$\iff p \in \bigcap_{a \in F_A(p)} N_A(a) \cap \Delta = \Delta_A^c(\bar{p}) \quad (64c)$$

by (59).

■

B.1 Valuable information

Proof. Proof of Proposition 1

Let $a \in F_A(\bar{p})$ and \mathbf{q} be an information structure. We have that

$$\mathbf{Vol}_A(\mathbf{q}) = 0 \iff \mathbb{E} [\sigma_A(\mathbf{q}) - \sigma_A(\bar{p})] = 0 \text{ by (6)} \quad (65)$$

$$\iff \mathbb{E} [\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle] = 0, \text{ as } \mathbb{E} [\mathbf{q} - \bar{p}] = 0 \quad (66)$$

$$\iff \sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle = 0, \text{ } \mathbb{P} - \text{a.s.} \quad (67)$$

because $\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle \geq 0$ by (40) since $a \in F_A(\bar{p})$

$$\iff \sigma_A(\mathbf{q}) = \langle \mathbf{q}, a \rangle, \text{ } \mathbb{P} - \text{a.s.} \quad (68)$$

because $\sigma_A(\bar{p}) = \langle \bar{p}, a \rangle$ since $a \in F_A(\bar{p})$

$$\iff \mathbb{P} \{a \in F_A(\mathbf{q})\} = 1 \quad (69)$$

$$\iff \mathbb{P} \{\langle \mathbf{q}, a' - a \rangle \leq 0, \forall a' \in A\} = 1. \quad (70)$$

Let $F \subset F_A(\bar{p})$ be dense subset of $F_A(\bar{p})$. We immediately get that

$$\mathbf{Vol}_A(\mathbf{q}) = 0 \Rightarrow \mathbb{P} \{\langle \mathbf{q}, a' - a \rangle \leq 0, \forall a' \in A, \forall a \in F\} = 1. \quad (71)$$

As the set $\{a \in F_A(\bar{p}) \mid \langle \mathbf{q}, a' - a \rangle \leq 0, \forall a' \in A\}$ is closed, we deduce that

$$\mathbf{Vol}_A(\mathbf{q}) = 0 \Rightarrow \mathbb{P} \{\langle \mathbf{q}, a' - a \rangle \leq 0, \forall a' \in A, \forall a \in F_A(\bar{p})\} = 1. \quad (72)$$

In other words, we obtain that, by definition (33) of the normal cone $N_A(a)$:

$$\mathbf{Vol}_A(\mathbf{q}) = 0 \Rightarrow \mathbf{q} \in \bigcap_{a \in F_A(\bar{p})} N_A(a), \text{ } \mathbb{P} - \text{a.s.} . \quad (73)$$

Revisiting the proof backward, we easily see that

$$\mathbf{q} \in \bigcap_{a \in F_A(\bar{p})} N_A(a), \text{ } \mathbb{P} - \text{a.s.} \Rightarrow \mathbf{Vol}_A(\mathbf{q}) = 0. \quad (74)$$

■

Proof of Theorem 2. Let $a \in F_A(\bar{p})$ and \mathbf{q} an information structure.

On the one hand, we have that

$$\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) = \mathbf{1}_{\mathbf{q} \in \Delta_A^c(\bar{p})} (\sigma_A(\mathbf{q}) - \sigma_A(\bar{p})) + \mathbf{1}_{\mathbf{q} \notin \Delta_A^c(\bar{p})} (\sigma_A(\mathbf{q}) - \sigma_A(\bar{p})) \quad (75a)$$

$$= \mathbf{1}_{\mathbf{q} \in \Delta_A^c(\bar{p})} \langle \mathbf{q} - \bar{p}, a \rangle + \mathbf{1}_{\mathbf{q} \notin \Delta_A^c(\bar{p})} (\sigma_A(\mathbf{q}) - \sigma_A(\bar{p})) \quad (75b)$$

as $a \in F_A(\mathbf{q})$ since, by (9), $\mathbf{q} \in \Delta_A^c(\bar{p})$ implies that $a \in F_A(\bar{p}) \subset F_A(\mathbf{q})$

$$= \langle \mathbf{q} - \bar{p}, a \rangle + \mathbf{1}_{\mathbf{q} \notin \Delta_A^c(\bar{p})} (\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle) . \quad (75c)$$

By taking the expectation \mathbb{E} , we obtain that

$$\mathbf{VoI}_A(\mathbf{q}) = \mathbb{E} [\sigma_A(\mathbf{q}) - \sigma_A(\bar{p})] = \mathbb{E} [\mathbf{1}_{\mathbf{q} \notin \Delta_A^c(\bar{p})} (\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle)] . \quad (76)$$

Now, on the one hand $(\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle)$ is bounded above by

$$C_A = \sup_{p \in \Delta} (\sigma_A(p) - \sigma_A(\bar{p}) - \langle p - \bar{p}, a \rangle) , \quad (77)$$

because the space Δ of beliefs is compact. Therefore, we have that $C_A \mathbb{P}(\mathbf{q} \notin \Delta_A^c(\bar{p})) \geq \mathbf{VoI}_A(\mathbf{q})$.

On the other hand, let \mathcal{Q} be an open subset of Δ that contains the confidence set $\Delta_A^c(p)$, that is, $\Delta_A^c(\bar{p}) \subset \mathcal{Q}$. We have that

$$\mathbb{E} [\mathbf{1}_{\mathbf{q} \notin \Delta_A^c(\bar{p})} (\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle)] \geq \mathbb{E} [\mathbf{1}_{\mathbf{q} \notin \mathcal{Q}} (\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle)] . \quad (78)$$

Now, on the set $\mathbf{q} \notin \mathcal{Q}$, $(\sigma_A(\mathbf{q}) - \sigma_A(\bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle)$ is bounded below by the positive number

$$c_A = \inf_{p \notin \mathcal{Q}} (\sigma_A(p) - \sigma_A(\bar{p}) - \langle p - \bar{p}, a \rangle) > 0 . \quad (79)$$

Indeed, $\sigma_A(p) - \sigma_A(\bar{p}) - \langle p - \bar{p}, a \rangle \geq 0$ by (40) since $a \in F_A(\bar{p})$. As $\Delta_A^c(\bar{p}) \subset \mathcal{Q}$, we deduce from Proposition 9 that $\sigma_A(p) - \sigma_A(\bar{p}) - \langle p - \bar{p}, a \rangle > 0$. Since the complementary set \mathcal{Q}^c is closed and bounded, hence compact, we obtain that $c_A > 0$. Therefore, we have that

$$\mathbf{VoI}_A(\mathbf{q}) \geq c_A \mathbb{P}(\mathbf{q} \notin \mathcal{Q}) . \quad (80)$$

■

B.2 Undecided

Proof of Proposition 2.

We prove that the face $F_A(\bar{p})$ of A in the direction $\bar{p} \in \Delta$ is a singleton if and only if the value function V_A in (4) is differentiable at \bar{p} .

- Suppose that the face $F_A(\bar{p})$ of A in the direction $\bar{p} \in \Delta$ is a singleton.

As the face $F_A(\bar{p})$ is the subdifferential at \bar{p} of the support function σ_A (Hiriart-Uruty and Lemaréchal, 1993, p. 258), we deduce that σ_A is differentiable at \bar{p} (Hiriart-Uruty and Lemaréchal, 1993, p. 251). Therefore, the value function V_A in (4) is differentiable at \bar{p} , since $V_A : \Delta \rightarrow \mathbb{R}$ is the restriction of σ_A to probability distributions Δ , as in (29).

- Suppose the value function V_A in (4) is differentiable at \bar{p} .

We consider the *extended value function* defined by

$$\tilde{V}_A : \mathbb{R}_+^* \Delta \rightarrow \mathbb{R}, \quad s \mapsto \|s\| V_A\left(\frac{s}{\|s\|}\right). \quad (81)$$

Since the support function σ_A is positively homogeneous, we have that $\tilde{V}_A : \mathbb{R}_+^* \Delta \rightarrow \mathbb{R}$ is the restriction of σ_A to the cone $\mathbb{R}_+^* \Delta$:

$$\tilde{V}_A(s) = \sigma_A(s), \quad \forall s \in \mathbb{R}_+^* \Delta. \quad (82)$$

As the prior \bar{p} has full support, the extended value function \tilde{V}_A in (81) is well defined on a neighborhood of \bar{p} and is differentiable at \bar{p} , since so is V_A . Since $\tilde{V}_A : \mathbb{R}_+^* \Delta \rightarrow \mathbb{R}$ is the restriction of σ_A to the cone $\mathbb{R}_+^* \Delta$, we deduce that the support function σ_A is differentiable at \bar{p} .

Since, on the one hand, a convex function with domain \mathbb{R}^K is differentiable at \bar{p} if and only if the subdifferential at \bar{p} is a singleton (Hiriart-Uruty and Lemaréchal, 1993, p. 251), and, on the other hand, the face $F_A(\bar{p})$ is the subdifferential at \bar{p} of the support function σ_A (Hiriart-Uruty and Lemaréchal, 1993, p. 258), we conclude that the face $F_A(\bar{p})$ of A in the direction $\bar{p} \in \Delta$ is a singleton.

■

Proof of Theorem 3. We prove the three inequalities in (16). By definition (28) of a support function, we have that $\sigma_A(\cdot) \leq \|A\| \times \|\cdot\|$, where $\|A\| = \sup\{\|a\|, a \in A\} < +\infty$. Thus $C_A = \|A\|$ in the left hand side inequality in (16). We now prove the middle inequality.

For all $s \in \Sigma$, we have that

$$\sigma_A(s) - \sigma_A(\bar{p}) \geq \sigma_{F_A(\bar{p})}(s - \bar{p}) \quad (83a)$$

by (40) since $F_A(\bar{p}) \neq \emptyset$

$$= \langle s - \bar{p}, a \rangle, \quad \forall a \in F_A(\bar{p}) \quad (83b)$$

by definition of $\sigma_{F_A(\bar{p})}$

$$= \sigma_{F_A(\bar{p})}(s) - \sigma_{F_A(\bar{p})}(\bar{p}) \quad (83c)$$

by definition of $\sigma_{F_A(\bar{p})}$.

By taking the expectation \mathbb{E} , we obtain that

$$\mathbf{Vol}_A(\mathbf{q}) = \mathbb{E} [\sigma_A(\mathbf{q}) - \sigma_A(\bar{p})] \text{ by (6) and (29)} \quad (84a)$$

$$\geq \mathbb{E} [\sigma_{F_A(\bar{p})}(\mathbf{q} - \bar{p})] \text{ by (83a)} \quad (84b)$$

$$= \mathbb{E} [\sigma_{F_A(\bar{p})}(\mathbf{q}) - \sigma_{F_A(\bar{p})}(\bar{p})] \text{ by (83c)} \quad (84c)$$

$$= \mathbf{Vol}_{F_A(\bar{p})}(\mathbf{q}) \text{ by (6) and (29)}.$$

This ends the proof of the middle inequality.

Now, there remains to prove the right hand side inequality in (16). For this purpose, we recall that the *affine hull* $\text{aff}(S)$ of a subset S of \mathbb{R}^K is the intersection of all affine manifolds containing S . Let n be the dimension of the affine hull $\text{aff}(F_A(\bar{p}))$ of $F_A(\bar{p})$, and let a_1, \dots, a_n be n actions in $F_A(\bar{p})$ that generate $\text{aff}(F_A(\bar{p}))$. We put

$$T = \{a_1, \dots, a_n\} \subset F_A(\bar{p}) \text{ so that } \text{aff}(F_A(\bar{p})) = \text{aff}\{a_1, \dots, a_n\} = \text{aff}(T). \quad (85)$$

We will now show that

$$\|\cdot\|_{\Sigma_A^i(\bar{p})} = \frac{1}{n} \sigma_{T-T}(\cdot) \quad (86)$$

is a semi-norm with kernel $(F_A(\bar{p}) - F_A(\bar{p}))^\perp$ that satisfies the right hand side inequality in (16) with $c_A = 1$.

First, the support function σ_{T-T} is a semi-norm with kernel $(T - T)^\perp$ by item 5 in Proposition 7. Now, we can easily see that, for any subset $S \subset \mathbb{R}^K$, one has

$$(S - S)^\perp = (\text{aff}(S - S))^\perp = (\text{aff}(S) - \text{aff}(S))^\perp. \quad (87)$$

Using these equalities with $S = T$ and $S = F_A(\bar{p})$, we deduce that $(T - T)^\perp = (F_A(\bar{p}) - F_A(\bar{p}))^\perp$, since $\text{aff}(T) = \text{aff}(F_A(\bar{p}))$ by (85).

Second, we show that the right hand side inequality in (16) is satisfied with $c_A = 1$. We have that

$$\mathbf{Vol}_A(\mathbf{q}) \geq \mathbb{E} [\sigma_{F_A(\bar{p})}(\mathbf{q} - \bar{p})] \text{ by (84b)} \quad (88a)$$

$$\geq \mathbb{E} [\sigma_T(\mathbf{q} - \bar{p})] \quad (88b)$$

because $T \subset F_A(\bar{p})$

$$= \mathbb{E} [\sigma_T(\mathbf{q} - \bar{p}) - \langle \mathbf{q} - \bar{p}, a \rangle], \quad \forall a \in A \quad (88c)$$

because $\mathbb{E} [\langle \mathbf{q} - \bar{p}, a \rangle] = 0$.

$$= \mathbb{E} [\sigma_{T-a}(\mathbf{q} - \bar{p})], \quad \forall a \in A \quad (88d)$$

because $\sigma_{T-a} = \sigma_{T+\{-a\}} = \sigma_T + \sigma_{\{-a\}}$.

Therefore,

$$\mathbf{Vol}_A(\mathbf{q}) \geq \frac{1}{n} \sum_{i=1}^n \mathbb{E} [\sigma_{T-a_i}(\mathbf{q} - \bar{p})] = \frac{1}{n} \mathbb{E} [\sigma_{\sum_{i=1}^n (T-a_i)}(\mathbf{q} - \bar{p})] \quad (89)$$

because the support function transforms a Minkowski sum of sets into a sum of support functions (Hiriart-Uruty and Lemaréchal, 1993, p.226). Now, as $T = \{a_1, \dots, a_n\}$, it is easy to see that the sum $\sum_{i=1}^n (T - a_i)$ contains any element of the form $a_k - a_l$:

$$a_k - a_l = (a_1 - a_1) + \dots + (a_{l-1} - a_{l-1}) + (a_k - a_l) + (a_{l+1} - a_{l+1}) + \dots + (a_n - a_n) \in \sum_{i=1}^n (T - a_i). \quad (90)$$

As support functions are monotone with respect to set inclusion, we deduce that

$$\sigma_{\sum_{i=1}^n (T-a_i)} \geq \sigma_{\{a_k - a_l, k, l=1, \dots, n\}} = \sigma_{T-T}, \quad (91)$$

and that

$$\mathbf{Vol}_A(\mathbf{q}) \geq \frac{1}{n} \mathbb{E} [\sigma_{\{a_k - a_l, k, l=1, \dots, n\}}(\mathbf{q} - \bar{p})] = \frac{1}{n} \mathbb{E} [\sigma_{T-T}(\mathbf{q} - \bar{p})] = c_A \mathbb{E} \|\mathbf{q} - \bar{p}\|_{\Sigma_A^i(\bar{p})}, \quad (92)$$

with $c_A = 1$. ■

B.3 Flexible decisions

Proof of Proposition 3. All the recalls on geometric convex analysis in §A.2 were done with outer normal vectors belonging to the unit sphere of signed measures. Now, as we work with beliefs — positive measures of mass 1 — we are going to adapt the notions.

We will consider the diffeomorphism

$$\nu : S^{|K|-1} \cap \mathbb{R}_+^K \rightarrow \Delta, \quad s \mapsto \frac{s}{\langle s, 1 \rangle}, \quad (93)$$

that maps unit positive measures into probability measures, with inverse

$$\nu^{-1} : \Delta \rightarrow S^{|K|-1} \cap \mathbb{R}_+^K, \quad p \mapsto \frac{p}{\|p\|}. \quad (94)$$

Since, by assumption, the action set A has boundary ∂A which is a C^2 submanifold of \mathbb{R}^K , we know by Proposition 8 that the spherical image map $n_A : \partial A \rightarrow S^{|K|-1}$ in (51) is well defined, is of class C^1 , and has for right inverse the reverse spherical image map $f_A : \text{regn}(A) \rightarrow \partial A$ in (54), that is, $n_A \circ f_A = \text{Id}_{\text{regn}(A)}$.

The *set of relevant regular points* is the subset of the set $\text{reg}(A)$ of regular points defined by

$$a \in \text{reg}^+(A) \iff \exists p \in \Delta, \quad N_A(a) = \mathbb{R}_+ p. \quad (95)$$

For a regular action $a \in \text{reg}^+(A)$, there is only one probability $p \in \Delta$ such that $N_A(a) = \mathbb{R}_+ p$, and it is $p = \nu(n_A(a))$. We have

$$a \in \text{reg}^+(A) \Rightarrow N_A(a) = \mathbb{R}_+ \nu(n_A(a)) \quad \text{where } \nu(n_A(a)) \in \Delta. \quad (96)$$

The *set of regular probabilities* is

$$\text{regn}^+(A) = \left(\mathbb{R}_+^* \text{regn}(A) \right) \cap \Delta. \quad (97)$$

For a regular probability $p \in \text{regn}^+(A)$, there is only one action $a \in \partial A$ such that $F_A(p) = \{a\}$, and it is $a = f_A(\nu^{-1}(p))$. Indeed, by definition (30) of the (exposed) face, we have that

$$F_A(\lambda s) = F_A(s), \quad \forall \lambda \in \mathbb{R}_+^*, \quad \forall s \in \Sigma, \quad s \neq 0. \quad (98)$$

Therefore, we have that

$$p \in \text{regn}^+(A) \Rightarrow F_A(p) = \{f_A(\nu^{-1}(p))\}. \quad (99)$$

The following mappings are well defined

$$\nu \circ n_A : \text{reg}^+(A) \rightarrow \Delta \quad \text{and} \quad f_A \circ \nu^{-1} : \text{regn}^+(A) \rightarrow \partial A, \quad (100)$$

and we have that

$$(\nu \circ n_A) \circ (f_A \circ \nu^{-1}) = \text{Id}_{\text{regn}^+(A)}. \quad (101)$$

- Item 2 \Rightarrow item 1.

Suppose that the face $F_A(\bar{p})$ is a singleton $\{a^\sharp\}$ and the curvature of the boundary ∂A of payoffs at a^\sharp is positive.

Since, by assumption, the action set A has boundary ∂A which is a C^2 submanifold of \mathbb{R}^K , we know that the spherical image map n_A in (51) is defined over the whole boundary ∂A and is of class C^1 , with differential the Weingarten map.

As the curvature of the boundary ∂A of payoffs at a^\sharp is positive, the Weingarten map $T_{a^\sharp} n_A$ is of maximal rank at a^\sharp (Schneider, 2014, p. 115). Therefore, by the inverse function theorem, there exists an open neighborhood \mathcal{A} of a^\sharp in A such that $n_A(\mathcal{A})$ is an open neighborhood of $n_A(a^\sharp)$ in $S^{|\mathcal{K}|-1}$, and such that the restriction $n_A : \mathcal{A} \rightarrow n_A(\mathcal{A})$ of the spherical image map in (51) is a diffeomorphism. By (56), we have that $n_A(a^\sharp) = \frac{\bar{p}}{\|\bar{p}\|}$ and the local inverse coincides with the restriction $f_A : n_A(\mathcal{A}) \rightarrow \mathcal{A}$ of the reverse spherical image map in (54).

As $n_A(\mathcal{A})$ is an open neighborhood of $\frac{\bar{p}}{\|\bar{p}\|}$ in $S^{|\mathcal{K}|-1}$, and as the prior \bar{p} has full support, we deduce that $\nu(n_A(\mathcal{A}))$ is an open neighborhood of \bar{p} in Δ , where the diffeomorphism ν is defined in (93).

We easily deduce that $f_A \circ \nu^{-1} : \nu(n_A(\mathcal{A})) \rightarrow \mathcal{A}$ is a diffeomorphism. By (99), we conclude that $f_A \circ \nu^{-1}$ is the restriction of the set-valued mapping $F_A : \Delta \rightrightarrows A$, $p \mapsto F_A(p)$ in (17).

- Item 1 \Rightarrow item 3.

Suppose that the set-valued mapping $F_A : \Delta \rightrightarrows A$, $p \mapsto F_A(p)$ in (17) is a local diffeomorphism at \bar{p} .

By definition (53) of the set of regular unit signed measures, there exists an open neighborhood Π of \bar{p} in Δ such that $\Pi \subset \text{regn}^+(A)$, where the set of relevant regular points is defined in (95). In addition, the mapping $f_A \circ \nu^{-1} : \Pi \rightarrow f_A(\nu^{-1}(\Pi))$ is a diffeomorphism.

As $F_A(p) = \{f_A(\nu^{-1}(p))\}$, for all beliefs $p \in \Pi$, we know that the support function σ_A is differentiable and that its derivative is $\nabla_p \sigma_A = f_A(\nu^{-1}(p))$. As $f_A \circ \nu^{-1}$ is a local diffeomorphism at \bar{p} , and as the mapping ν in (93) is a diffeomorphism, we deduce that the support function σ_A is twice differentiable with Hessian having full rank. As the value function V_A is the restriction of σ_A to Δ , we conclude that V_A is twice differentiable at \bar{p} and the Hessian is definite positive.

- Item 3 \Rightarrow item 2.

Suppose that the value function V_A is twice differentiable at \bar{p} and the Hessian is definite positive.

On the one hand, as the prior \bar{p} has full support, there exists an open neighborhood Π of \bar{p} in Δ such that V_A is differentiable on Π . On the other hand, as the support function σ_A is positively homogeneous, and by (29), we have that

$$\sigma_A(s) = \langle s, 1 \rangle V_A \circ \nu(s), \quad \forall s \in S^{|K|-1} \cap \mathbb{R}_+^K. \quad (102)$$

Therefore, as the mapping ν in (93) is a diffeomorphism, the support function σ_A is differentiable on the open neighborhood $\nu^{-1}(\Pi)$ of $\nu^{-1}(\bar{p}) = \frac{\bar{p}}{\|\bar{p}\|}$ in $S^{|K|-1} \cap \mathbb{R}_+^K$.

Since, on the one hand, a convex function with domain \mathbb{R}^K is differentiable at s if and only if the subdifferential at s is a singleton (Hiriart-Uruty and Lemaréchal, 1993, p. 251), and, on the other hand, the face $F_A(s)$ is the subdifferential at s of the support function σ_A (Hiriart-Uruty and Lemaréchal, 1993, p. 258), we conclude that the face $F_A(s)$ of A in the direction $s \in \nu^{-1}(\Pi)$ is a singleton.

Therefore, by definition (53) of the set of regular unit signed measures, we have that $\nu^{-1}(\Pi) \subset \text{regn}(A)$. In addition, the restriction $f_A : \nu^{-1}(\Pi) \rightarrow f_A(\nu^{-1}(\Pi))$ of the reverse spherical image map in (54) is well defined, and we have that

$$\nabla_s \sigma_A = f_A(s), \quad \forall s \in \nu^{-1}(\Pi). \quad (103)$$

Therefore, the mapping $f_A : \nu^{-1}(\Pi) \rightarrow f_A(\nu^{-1}(\Pi))$ is differentiable at $\nu^{-1}(\bar{p}) = \frac{\bar{p}}{\|\bar{p}\|}$, and has full rank. Indeed, σ_A is twice differentiable at $\nu^{-1}(\bar{p}) = \frac{\bar{p}}{\|\bar{p}\|}$ and the Hessian

is definite positive. This comes from (102), where the mapping ν in (93) is a C^∞ diffeomorphism and the value function V_A is twice differentiable at \bar{p} with definite positive Hessian.

As f_A is differentiable at $\frac{\bar{p}}{\|\bar{p}\|}$ and has full rank, the reverse Weingarten map $T_s f_A$ in (58) is well defined and has full rank. Therefore, the principal radii of curvature of A at $\frac{\bar{p}}{\|\bar{p}\|}$ are positive. Letting $a^\sharp = f_A\left(\frac{\bar{p}}{\|\bar{p}\|}\right)$, we conclude that $F_A(\bar{p}) = \{a^\sharp\}$ and that the curvature of the boundary ∂A of payoffs at a^\sharp is positive.

This ends the proof. ■

Proof of Theorem 4. We suppose that the value function V_A is twice differentiable at \bar{p} and the Hessian is definite positive. We also denote $F_A(\bar{p}) = \{a^\sharp\}$.

First, we show that the function

$$g(p) = \frac{V_A(p) - V_A(\bar{p}) - \langle p - \bar{p}, a^\sharp \rangle}{\|p - \bar{p}\|^2} \quad (104)$$

is continuous and positive on Δ . Indeed, g is continuous on $\Delta \setminus \{\bar{p}\}$, and also at \bar{p} since the value function V_A is twice differentiable at \bar{p} . In addition, $g(\bar{p}) > 0$ since the Hessian of V_A at \bar{p} is definite positive. We have $g \geq 0$ on $\Delta \setminus \{\bar{p}\}$, because $F_A(\bar{p}) = \{a^\sharp\}$ is the subdifferential at \bar{p} of the support function σ_A , and by (29). We now prove by contradiction that $g > 0$. If there existed a belief $p \neq \bar{p}$ such that $g(p) = 0$, we would have $V_A(p) - V_A(\bar{p}) - \langle p - \bar{p}, a^\sharp \rangle = 0$; this equality would then hold true over the whole segment $[p, \bar{p}]$, and we would conclude that the second derivative of V_A at \bar{p} along the direction $p - \bar{p}$ would be zero; this would contradict the assumption that the Hessian of V_A at \bar{p} is definite positive. Therefore, we conclude that $g > 0$.

Second, letting $C_A > 0$ and $c_A > 0$ be the maximum and the minimum of the function $g > 0$ on the compact set Δ , we easily deduce (18) from (6).

This ends the proof. ■