Axioms for Information Leakage

Mário S. Alvim* Konstantinos Chatzikokolakis† Annabelle McIver‡
Carroll Morgan§ Catuscia Palamidessi† Geoffrey Smith¶

* Computer Science Department, Universidade Federal de Minas Gerais
† CNRS, Inria, and École Polytechnique
‡ Department of Computing, Macquarie University
§ School of Computer Science & Engineering, University of New South Wales, and Data61
¶ School of Computing & Information Sciences, Florida International University

***Abstract*—Quantitative information flow aims to assess and control the leakage of sensitive information by computer systems. A key insight in this area is that no single leakage measure is appropriate in all operational scenarios; as a result, many leakage measures have been proposed, with many different properties. To clarify this complex situation, this paper studies information leakage axiomatically, showing important dependencies among different axioms. It also establishes a completeness result about the g-leakage family, showing that any leakage measure satisfying certain intuitively-reasonable properties can be expressed as a g-leakage.**

***Index Terms*—information flow, g-vulnerability, information theory, confidentiality.**

I. Introduction

The theory of *quantitative information flow* has seen rapid development over the past decade, motivated by the need for rigorous techniques to *assess* and *control* the leakage of sensitive information by computer systems. The starting point of this theory is the modeling of a *secret* as something whose value is known to the adversary only as a *prior probability distribution* π. This immediately suggests that the "amount" of secrecy might be quantified based on π, where intuitively a uniform π would mean "more" secrecy and a biased π would mean "less" secrecy. But how, precisely, should the quantification be done?

Early work in this area (e.g., [1]) adopted classic information-theoretic measures like *Shannon-entropy* [2] and *guessing-entropy* [3]. But these can be quite misleading in a security context, because they can be arbitrarily high even if π assigns a large probability to one of the secret's possible values, giving the adversary a large chance of guessing that secret correctly in just one try. This led to the introduction of *Bayes vulnerability* [4], which is simply the maximum probability that π assigns to any of the possible values of the secret. Bayes vulnerability indeed measures a basic security threat, but it implicitly assumes an operational scenario where the adversary must guess the secret exactly, in one try. There are of course many other possible scenarios, including those where the adversary benefits by guessing a *part* or a *property* of the secret or by guessing the secret within *three tries*, or where the adversary is *penalized* for making an incorrect guess. This led to the introduction of *g-vulnerability* [5], which uses *gain functions* g to model the operational scenario, enabling specific g-vulnerabilities to be tailored to each of the above scenarios, and many others as well.[1]

This situation may however strike us as a bit of a zoo. We have a multitude of exotic vulnerability measures, but perhaps no clear sense of what a vulnerability measure ought to be. Are all the g-vulnerabilities "reasonable"? Are there "reasonable" vulnerability measures that we are missing?

The situation becomes more complex when we turn our attention to systems. We model systems as information-theoretic *channels*, and the crucial insight, reviewed in Section II-B below, is that each possible output of a channel allows the adversary to update the prior distribution π to a *posterior distribution*, where the posterior distribution itself has a probability that depends on the probability of the output. Hence a channel is a mapping from prior distributions to *distributions on posterior distributions*, called *hyper-distributions* [6].

In assessing *posterior vulnerabilities*, by which we mean the vulnerability after the adversary sees the channel output, we have a number of choices. It is natural to consider the vulnerability of each of the posterior distributions, and take the *average*, weighted by the probabilities of the posterior distributions. Or (if we are pessimistic) we might take the *maximum*. Next we can define the *leakage* caused by the channel by comparing the posterior vulnerability and prior vulnerability, either multiplicatively or additively. These choices, together with the multitude of vulnerability measures, lead us to many different leakage measures, with many different properties. Is there a systematic way to understand them? Can we bring order to the zoo?

Such questions motivate the axiomatic study that we undertake in this paper. We consider a set of axioms that characterize intuitively-reasonable properties that vulnerability measures might satisfy, separately considering axioms for prior vulnerability (Section IV) and axioms for posterior vulnerability and for the *relationship* between prior and posterior vulnerability (Section V). Addressing this relationship is an important novelty of our axiomatization, as compared with

[1] Note that *entropies* measure secrecy from the point of view of the *user* (i.e., more entropy means more secrecy), while *vulnerabilities* measure secrecy from the point of view of the *adversary* (i.e., more vulnerability means less secrecy). The two perspectives are complementary, but to avoid confusion this paper focuses almost always on the vulnerability perspective.

DOI 10.1109/CSF.2016.13

previous axiomatizations of entropy (such as [2], [7], [8]), which considered only prior entropy, or the axiomatization of *utility* by Kifer and Lin [9], which considers posterior utility without investigating its relation to prior utility. As a result, our axiomatization is able to consider properties of *leakage*, usually defined in terms of comparison between the posterior and prior vulnerabilities.[2]

The main contributions of this paper are of two kinds. One kind involves showing interesting *dependencies* among the various axioms. For instance, under axiom *averaging* for posterior vulnerability, we prove in Section V that three other axioms are equivalent: *convexity*, *monotonicity* (i.e., non-negativity of leakage), and the *data-processing inequality*. Convexity is the property that prior vulnerability is a *convex function* from distributions to reals; what is striking here is that it a property that might not be intuitively considered "fundamental", yet our equivalence (assuming averaging) shows that it is. We also show an equivalence under the alternative axiom *maximum* for posterior vulnerability, which then involves *quasi-convexity*.

A second kind of contribution justifies the significance of g-vulnerability. Focusing on the axioms of *convexity* and *continuity* for prior vulnerability, we consider the class of *all* functions from distributions to reals that satisfy them, proving in Section IV that this class *exactly coincides* with the class of g-vulnerabilities. This *soundness* and *completeness* result shows that if we accept averaging, continuity, and convexity (or monotonicity or the data-processing inequality) then prior vulnerabilities are exactly g-vulnerabilities.

The rest of the paper is structured as follows: Section II reviews the basic concepts of quantitative information flow, Section III sets up the framework of our axiomatization, and Sections IV and V discuss axioms for prior and posterior vulnerabilites, respectively. Section VI provides some discussion, Section VII gives an abstract categorical perspective, Section VIII discusses related work, and Section IX concludes.

II. Preliminaries

We now review some basic notions from quantitative information flow. A *secret* is something whose value is known to the adversary only as a *prior probability distribution* π: there are various ways for measuring what we will call its *vulnerability*. A *channel* models systems with observable behavior that changes the adversary's probabilistic knowledge, making the secret more vulnerable and hence causing information *leakage*.

A. Secrets and vulnerability

The starting point of computer security is information that we wish to keep *secret*, such as a user's password, social security number or current location. An adversary typically does not know the value of the secret, but still possesses some *probabilistic information* about it, captured by a probability distribution called the *prior*. We denote by $\mathcal{X}$ the finite set of possible secret values and by $\mathbb{D}\mathcal{X}$ the set of probability distributions over $\mathcal{X}$. A prior $\pi{\in}\mathbb{D}\mathcal{X}$ could either reflect a probabilistic procedure for choosing the secret—e.g., the probability of choosing a certain password—, or it could capture any knowledge the adversary possesses on the population the user comes from—e.g., a young person is likely to be located at a popular bar on Saturday night.

The prior π plays a central role at measuring how *vulnerable* a secret is. For instance, choosing short passwords is not vulnerable because of their length (prefixing passwords with a thousand zeroes does not necessarily render them more secure), but because each password has a high probability of being chosen. To obtain a concrete vulnerability measure one needs to consider an *operational scenario* describing the adversary's capabilities and goals; vulnerability then measures the adversary's expected success in this scenario.

Bayes-vulnerability [4] considers an adversary trying to *guess* the secret in *one try* and measures the threat as the *probability* of the guess being correct. Knowing a prior π, a rational adversary will guess a secret to which it assigns the highest probability: hence Bayes-vulnerability is given by

$$V_b(\pi) \;=\; \max_{x\in\mathcal{X}} \pi_x \;,$$

where we write π_x for the probability π assigns to x. Note that Bayes-vulnerability is called simply "vulnerability" in [4], and is the basic notion behind *min-entropy*, defined as $H_\infty(\pi) = -\lg V_b(\pi)$. It is also the converse of the adversary's probability of error, also called *Bayes-risk* in the area of hypothesis testing [10].

Guessing-entropy [3] considers an adversary trying to guess the secret in an unlimited number of tries, and measures the adversary's uncertainty as the *number of guesses* needed on average. The best strategy is to try secrets in non-increasing order of probability: if x_i is an indexing of $\mathcal{X}$ in such an order, then guessing-entropy is given by

$$G(\pi) \;=\; \textstyle\sum_i i\, \pi_{x_i} \;.$$

Shannon-entropy [2] considers an adversary who tries to infer the secret using Boolean questions (i.e., of the form "does x belong to a certain subset $\mathcal{X}'$ of $\mathcal{X}$?") and measures the adversary's uncertainty as the *number of questions* needed on average. It can be shown that the best strategy is at each step to split the secret space in sets of equal probability (as far as possible). Under this strategy, a secret x will be guessed in $-\lg \pi_x$ steps, hence on average the number of questions needed is

$$H(\pi) \;=\; -\textstyle\sum_{x\in\mathcal{X}} \pi_x \lg \pi_x \;.$$

Note that Bayes-vulnerability measures the *threat* to the secret (the higher the better for the adversary). On the other hand, guessing- and Shannon-entropy measure the adversary's *uncertainty* about the secret (the lower the better for the adversary).

[2]We should however clarify that we do not view axiomatics as a matter of identifying "self-evident" truths. A variety of axioms may appear intuitively reasonable, so while it is sensible to consider intuitive justifications for them, such justifications should not be considered absolute. Rather we see the value of axiomatics as consisting more in understanding the logical dependencies among different properties, so that we might (for instance) identify a minimal set of axioms that is sufficient to imply all the properties that we care about.

Although the operational scenarios described above capture realistic threats for the secret, one could envision a variety of alternative threats we might also be worried about. For instance, an adversary might be interested in guessing only *part* of the secret, an *approximate* value of the secret, a *property* of the secret or guessing the secret in a fixed number of tries. It is for this reason that the more general *g-vulnerability* framework [5] was proposed: it allows one to adapt to many different adversarial models.

Its operational scenario is parametrized by a set $\mathcal{W}$ of *guesses* (possibly infinite) that the adversary can make *about* the secret, and a *gain function* $g: \mathcal{W} \times \mathcal{X} \to \mathbb{R}$. The gain $g(w, x)$ expresses the adversary's benefit for having made the guess w when the actual secret is x. The g-vulnerability function measures the threat as the adversary's expected gain for an optimal choice of guess w:

$$V_g(\pi) \;=\; \sup_{w \in \mathcal{W}} \textstyle\sum_{x \in \mathcal{X}} \pi_x g(w, x)\ . \tag{1}$$

Regarding the set $\mathcal{W}$ of allowable guesses, one might assume that this should just be $\mathcal{X}$, the set of possible values of the secret. This is in fact too restrictive: the adversary's goal might be to guess a *piece* of the secret, or a value *close* to the secret, or some *property* of the secret. As a consequence we allow an arbitrary set of guesses, possibly infinite, and make (almost) no restrictions on the values of g. In particular, a negative value of $g(w, x)$ expresses situations when the adversary is *penalized* for making a particular guess under a particular secret; such values are essential for obtaining the results of Section IV-B. We do however impose one restriction on g, that for each prior π there is at least one guess that gives *non-negative gain*. This essentially forces V_g to be non-negative, although individual guesses (i.e particular w's) can still give negative gain.

Note that, as its name suggests, V_g is a measure of vulnerability, i.e., of the threat to the secret. An equally expressive alternative is to define an "uncertainty" measure similarly, but using a *loss function* l instead of a gain function and assuming that the adversary wants to minimize loss. The uncertainty measure, parametrized by l, can be then defined dually as $U_l(\pi) = \inf_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \pi_x l(w, x)$, and is often called Bayes-risk in the area of decision theory.

Due to the flexibility of gain functions, g-vulnerability is a very expressive framework, one that can capture a great variety of operational scenarios. This raises the natural question of which other vulnerability measures are expressible in this framework. Bayes-vulnerability is a straightforward example, captured by guessing the exact secret, i.e., taking $\mathcal{W}{=}\mathcal{X}$, and using the *identity* gain function defined as $g_{id}(w, x) = 1$ iff $w{=}x$ and 0 otherwise.

Guessing-entropy can be also captured in this framework [11], [12], this time using a loss function since it's an uncertainty measure. The adversary in this case tries to guess a *permutation* of $\mathcal{X}$, i.e., the order in which secrets are chosen in the operational scenario of guessing-entropy. We can naturally define the loss $l(w, x)$ as the index of x in w, i.e. the number of guesses to find x, and using this loss function we get $U_l(\pi) = G(\pi)$.

Similarly, in the case of Shannon-entropy, the adversary tries to guess a strategy for constructing his questions. Strategies can be described as probability distributions: at each step questions split the search space into subsets of as even probability as possible. Hence, guesses are $\mathcal{W}{=}\mathbb{D}\mathcal{X}$, and the loss can be defined as $l(w, x) = -\lg w_x$ (the number of steps needed to find x under the strategy w). Since the best strategy is to take $w{=}\pi$ itself, it can be shown [11] that under this loss function $U_l(\pi) = H(\pi)$.

In Section IV-B we show that g-vulnerability exactly coincides with the generic class of continuous and convex vulnerability functions.

B. Channels, hypers and leakage

So far we have considered secrets for which a probabilistic prior is known, and have discussed different ways for measuring their vulnerability. We now turn our attention to *systems*, which are programs or protocols processing secret information and producing some *observable* behavior. Examples of such systems are password-checkers, implementations of cryptosystems, and anonymity protocols.

A system can be modeled as an *(information theoretic) channel*, a triple $(\mathcal{X}, \mathcal{Y}, C)$, where $\mathcal{X}, \mathcal{Y}$ are finite sets of (secret) input values and (observable) output values respectively and C is a $|\mathcal{X}| \times |\mathcal{Y}|$ channel matrix in which each entry $C_{x,y}$ corresponds to the probability of the channel producing output y when the input is x. Hence each row of C is a probability distribution over $\mathcal{Y}$ (entries are non-negative and sum to 1). A channel is *deterministic* iff each row contains a single 1 identifying the only possible output for that input.

It is typically assumed that the adversary knows how the system works, i.e. knows the channel matrix C. Knowing also the prior distribution π, the adversary can compute the joint distribution $p(x, y){=}\pi_x C_{x,y}$ on $\mathcal{X} \times \mathcal{Y}$, producing joint random variables X, Y with marginal probabilities $p(x) = \sum_y p(x, y)$ and $p(y) = \sum_x p(x, y)$, and conditional probabilities $p(y|x) = {}^{p(x,y)}/_{p(x)}$ (if $p(x)$ is non-zero) and $p(x|y) = {}^{p(x,y)}/_{p(y)}$ (if $p(y)$ is non-zero). Note that p_{XY} is the unique joint distribution that recovers π and C, in that $p(x) = \pi_x$ and $p(y \mid x) = C_{x,y}$ (if $p(x)$ is non-zero).[3]

For a given y (s.t. $p(y)$ is non-zero), the conditional probabilities $p(x|y)$ for each $x \in \mathcal{X}$ form the *posterior distribution* $p_{X|y}$, which represents the posterior knowledge the adversary has about input X after observing output y.

Example 1. *Given $\mathcal{X} = \{x_1, x_2, x_3\}$, $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$, and the channel matrix C below, (the uniform) prior $\pi = ({}^1/_3, {}^1/_3, {}^1/_3)$ combined with C leads to joint matrix J:*

C	y_1	y_2	y_3	y_4
x_1	1	0	0	0
x_2	0	$1/2$	$1/4$	$1/4$
x_3	$1/2$	$1/3$	$1/6$	0

$\xrightarrow{\pi}$

J	y_1	y_2	y_3	y_4
x_1	$1/3$	0	0	0
x_2	0	$1/6$	$1/12$	$1/12$
x_3	$1/6$	$1/9$	$1/18$	0

[3]When necessary to avoid ambiguity, we write distributions with subscripts, e.g. p_{XY} or p_Y.

Summing columns of J gives the marginal distributions $p_Y{=}(1/2, 5/18, 5/36, 1/12)$, and normalizing gives the posterior distributions $p_{X|y_1}{=}(2/3, 0, 1/3)$, $p_{X|y_2}{=}(0, 3/5, 2/5)$, $p_{X|y_3}{=}(0, 3/5, 2/5)$, and $p_{X|y_4}{=}(0, 1, 0)$. $\square$

The effect of a channel C is to update the adversary's knowledge from a prior π to a collection of posteriors $p_{X|y}$, each occurring with probability $p(y)$. Hence, following [6], [13], we view a channel as producing a probability distribution over posteriors, called a *hyper-distribution*.[4]

A hyper (for short) on the input space $\mathcal{X}$ is of type $\mathbb{D}^2\mathcal{X}$, which stands for $\mathbb{D}(\mathbb{D}\mathcal{X})$, a distribution on distributions on $\mathcal{X}$. The support of a hyper is the set of possible posteriors that the action of channel C on prior π can produce: we call those posteriors *inners*. The probability assigned by the hyper to a particular inner is the marginal probability of the y that produced that inner. We call those probabilities the *outer* probabilities. We use Δ to denote a hyper, $\lceil\Delta\rceil$ for its *support* (the set of posteriors with non-zero probability), $[\pi]$ to denote the point-hyper assigning probability 1 to π, and $[\pi, C]$ to denote the hyper obtained by the action of C on π. We say that $[\pi, C]$ is the result of *pushing prior π through channel C*.

In Example 1, the hyper $[\pi, C]$ assigns (outer) probabilities $(1/2, 15/36, 1/12)$ to the (inner) posteriors $(2/3, 0, 1/3)$, $(0, 3/5, 2/5)$, and $(0, 1, 0)$, respectively.[5]

Since the outcome of a channel is a hyper, it is natural to extend vulnerability measures from priors to hypers, obtaining a *posterior* vulnerability. For all measures described in Section II-A this has been done in a natural way by taking the vulnerability of each posterior and *averaging* them using the outer. Let $\mathrm{Exp}_\pi\, F{:=}\sum_x \pi_x F(x)$ denote the *expected value* of some random variable $F{:}\,\mathcal{X}{\to}R$ (where R is usually the reals $\mathbb{R}$ but more generally can be a vector space) over a distribution $\pi{:}\,\mathbb{D}\mathcal{X}$. We can then define *posterior Bayes-vulnerability* $\widehat{V_b} : \mathbb{D}^2\mathcal{X} \to \mathbb{R}^+$ as

$$\widehat{V_b}\Delta \quad = \quad \mathrm{Exp}_\Delta\, V_b \ ,$$

and similarly for Shannon-entropy, guessing-entropy and g-vulnerability. For hypers $[\pi, C]$ produced by channels, from the above formula we can get an expression of each posterior vulnerability as a function of π and C, for instance,

$$\begin{array}{rcll} \widehat{V_b}[\pi, C] & = & \sum_y \max_x \pi_x C_{x,y} \ , & \text{and} \\ \widehat{V_g}[\pi, C] & = & \sum_y \sup_w \sum_x \pi_x C_{x,y} g(w, x) \ . & \end{array}$$

Note that, for point-hypers, we have by construction that $\widehat{V_b}[\pi] = V_b(\pi)$, and similarly for the other measures.

Finally, the execution of a system is expected to disclose information about the secret to the adversary, and the *information leakage* of a channel C for a prior π is defined by comparing the vulnerability of the prior π—the adversary's prior knowledge—and that of $[\pi, C]$—the adversary's posterior knowledge. The comparison is typically done either *additively* or *multiplicatively*, giving rise to two versions of leakage:

$$\begin{array}{rcll} \text{additive:}\, \mathcal{L}_b^+(\pi, C) & = & \widehat{V_b}[\pi, C] - V_b(\pi) \ , \quad \text{and} & (2) \\ \text{multiplicative:}\, \mathcal{L}_b^\times(\pi, C) & = & \lg\left(\widehat{V_b}[\pi,C] / V_b(\pi)\right) \ . & (3) \end{array}$$

Note that $\mathcal{L}_b^\times(\pi, C)$ is usually called *min-entropy leakage* [4]. Leakage can be similarly defined for all other measures.

[4]Mappings of priors to hypers are called *abstract* channels in [13].

[5]There might be fewer posteriors in the support of hyper $[\pi, C]$ than there are columns in the joint distribution $p_{X,Y}$ from which it is derived, because if several columns of $p_{X,Y}$ normalize to the same posterior then the hyper will automatically coalesce them [13]. Columns y_2 and y_3 were coalesced in this case.

Object	Type	Typical instance
secret	$\mathcal{X}$	x
prior	$\mathbb{D}\mathcal{X}$	π
hyper-distribution	$\mathbb{D}^2\mathcal{X}$	Δ or $[\pi, C]$
(abstract) channel	$\mathbb{D}\mathcal{X} \to \mathbb{D}^2\mathcal{X}$	C
prior vulnerability	$\mathbb{D}\mathcal{X} \to \mathbb{R}$	$\mathbb{V}$
posterior vulnerability	$\mathbb{D}^2\mathcal{X} \to \mathbb{R}$	$\widehat{\mathbb{V}}$

TABLE I: Notation.

III. Axiomatization

In Section II we discussed vulnerability measures obtained by quantifying the threat to the secret in a specific operational scenario. Channels were then introduced, mapping prior distributions to hypers, and the vulnerability measures were naturally extended to posterior ones by averaging each posterior vulnerability over the hyper.

In this paper we take an alternative approach. Instead of constructing specific vulnerability measures, we consider generic vulnerability functions, that is, functions of type:

$$\begin{array}{rll} \text{prior vulnerability:} & \mathbb{V} : \mathbb{D}\mathcal{X} \to \mathbb{R}^+, & \text{and} \\ \text{posterior vulnerability:} & \widehat{\mathbb{V}} : \mathbb{D}^2\mathcal{X} \to \mathbb{R}^+. & \end{array}$$

We then introduce a variety of properties that "reasonable" vulnerabilities might be expected to have in terms of *axioms*, and study their consequences.

In Section IV we focus on the prior case and give axioms for prior vulnerabilities $\mathbb{V}$ alone. We then show that taking convexity and continuity as our generic properties results in g-vulneraibility exactly. Then, in Section V, we turn our attention to axioms considering either both $\mathbb{V}$ and $\widehat{\mathbb{V}}$, or posterior $\widehat{\mathbb{V}}$ alone. Moreover we study two ways of constructing $\widehat{\mathbb{V}}$ from $\mathbb{V}$ and show that, in each case, several of the axioms become equivalent.

Note that the axioms purely affect the relationship between prior and posterior vulnerabilities, and are orthogonal to the way $\mathbb{V}$ and $\widehat{\mathbb{V}}$ are compared to measure leakage (e.g., multiplicatively or additively). Moreover, although in this paper we consider axioms for vulnerability, dual axioms can be naturally stated for generic uncertainty measures.

Table I summarizes the notation used through the paper, while Table II summarizes the axioms we consider.

IV. Axiomatization of prior vulnerabilities

We now introduce axioms that deal solely with prior vulnerabilities $\mathbb{V}$.

Axioms for prior vulnerabilities	
CNTY	$\forall\pi$: $\mathbb{V}$ is a continuous function of π
CVX	$\forall\sum_i a_i\pi^i$: $\mathbb{V}\left(\sum_i a_i\pi^i\right) \leq \sum_i a_i\mathbb{V}(\pi^i)$
Q-CVX	$\forall\sum_i a_i\pi^i$: $\mathbb{V}\left(\sum_i a_i\pi^i\right) \leq \max_i \mathbb{V}(\pi^i)$
Axioms for posterior vulnerabilities	
NI	$\forall\pi$: $\widehat{\mathbb{V}}[\pi] = \mathbb{V}(\pi)$
DPI	$\forall\pi, C, R$: $\widehat{\mathbb{V}}[\pi, C] \geq \widehat{\mathbb{V}}[\pi, CR]$
MONO	$\forall\pi, C$: $\widehat{\mathbb{V}}[\pi, C] \geq \mathbb{V}(\pi)$
Possible definitions of posterior vulnerabilities	
AVG	$\forall\Delta$: $\widehat{\mathbb{V}}\Delta = \mathrm{Exp}_\Delta\,\mathbb{V}$
MAX	$\forall\Delta$: $\widehat{\mathbb{V}}\Delta = \max_{\lceil\Delta\rceil}\mathbb{V}$

TABLE II: Summary of axioms for pairs of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$.

Continuity (CNTY). A vulnerability $\mathbb{V}$ is a continuous function of π (w.r.t. the standard topology on $\mathbb{D}\mathcal{X}$).

The CNTY axiom imposes that "small" changes on the prior π should have a "small" effect on $\mathbb{V}$. This formalizes the intuition that the adversary should not be infinitely risk-averse. For instance, the non-continuous function $\mathbb{V}(\pi) = (1 \text{ if } \max_x \pi_x \geq \lambda \text{ else } 0)$ would correspond to an adversary who requires the probability of guessing to be above a certain threshold in order to consider an attack effective. But this is an arguably unnatural behavior, the risk of changing the probability to $\lambda-\epsilon$, for an infinitesimal ϵ, should not be arbitrarily large.

A *convex combination* of priors $\pi^1, \ldots, \pi^n$ is a sum $\sum_i a_i\pi^i$ where a_i's are non-negative reals adding up to 1. Since $\mathbb{D}\mathcal{X}$ is a convex set, a convex combination of priors is itself a prior.

Convexity (CVX). A vulnerability $\mathbb{V}$ is a convex function of π, that is for all convex combinations $\sum_i a_i\pi^i$:

$$\mathbb{V}\left(\textstyle\sum_i a_i\pi^i\right) \leq \textstyle\sum_i a_i\mathbb{V}(\pi^i) .$$

This axiom can be interpreted as follows: imagine a "game" in which a secret (say a password) is drawn from two possible distributions π^1 or π^2. The choice of distributions is itself random: we first select $i \in \{1, 2\}$ at random, with $i = 1$ having probability a_1 and $i = 2$ probability $a_2 = 1 - a_1$, and then use π^i to draw the secret.

Now consider the following two scenarios for this game: in the first scenario, the value of i is given to the adversary, so the actual prior the secret was drawn from is known. Using the information in π^i the adversary performs an attack, the expected success of which is measured by $\mathbb{V}(\pi^i)$, so the expected measure of success overall will be $\sum_i a_i\mathbb{V}(\pi^i)$.

In the second scenario, i is not disclosed to the adversary, who only knows that, on average, secrets are drawn from the prior $\sum_i a_i\pi^i$, hence the expected success of an attack will be measured by $\mathbb{V}\left(\sum_i a_i\pi^i\right)$. CVX corresponds to the intuition that, since in the first scenario the adversary has more information, the effectiveness of an attack can only be higher.

Note that, in the definition of CVX, it is sufficient to use convex combinations of *two* priors, i.e., of the form $a\pi^1 + (1-a)\pi^2$; we often use such combinations in proofs. Note also that CVX actually implies continuity everywhere except on the *boundary* of the domain, i.e., on priors having an element with probability exactly 0. CNTY explicitly requires continuity everywhere.

Since the vulnerabilities $\mathbb{V}(\pi^i)$ in the definition of CVX are weighted by the probabilities a_i, we could have cases when the expected vulnerability $\sum_i a_i\mathbb{V}(\pi^i)$ is small although some individual $\mathbb{V}(\pi^i)$ is large. In such cases, one might argue that the bound imposed by CVX is too strict and could be loosened by requiring that $\mathbb{V}\left(\sum_i a_i\pi^i\right)$ is only bounded by the maximum of the individual vulnerabilities. This weaker requirement is called *quasiconvexity*.

Quasiconvexity (Q-CVX). A vulnerability $\mathbb{V}$ is a quasiconvex function of π, that is for all $\sum_i a_i\pi^i$:

$$\mathbb{V}\left(\textstyle\sum_i a_i\pi^i\right) \leq \max_i \mathbb{V}(\pi^i) .$$

In Section V we show that CVX and Q-CVX can be in fact obtained as consequences of fundamental axioms relating prior and posterior vulnerabilities, and specific choices for constructing $\widehat{\mathbb{V}}$.

In the remainder of this section we show that the vulnerability functions satisfying CNTY and CVX are exactly those expressible as V_g for some gain function g. We treat each direction separately; full proofs are given in Appendix A.

A. V_g satisfies CNTY and CVX

We first show that any g-vulnerability satisfies CNTY and CVX. Let $\mathcal{W}$ be a possibly infinite set of guesses and $g\colon \mathcal{W} \times \mathcal{X} \to \mathbb{R}$ be a gain function. We start by expressing V_g as the supremum of a family of functions:

$$V_g(\pi) = \sup_w g_w(\pi), \quad \text{where } g_w(\pi) = \textstyle\sum_x \pi_x g(w, x) .$$

Intuitively, g_w gives the expected gain for the specific guess w, as a function of π. Note that g_w is linear on π, hence both (trivially) convex and continuous.

The convexity of V_g then follows from the fact that the sup of any family of convex functions is itself a convex function. On the other hand, showing continuity is more challenging, since the supremum of continuous functions is not necessarily continuous itself.

To show that V_g is continuous, we employ the concept of semi-continuity. Informally speaking, a function is upper (resp. lower) semi-continuous at x_0 if, for values close to x_0, the function is either close to $f(x_0)$ or less than $f(x_0)$ (resp. greater than $f(x_0)$).

Lower semi-continuity is obtained from the following proposition:

Proposition 2. *If f is the supremum of a family of continuous functions then it is lower semi-continuous.*

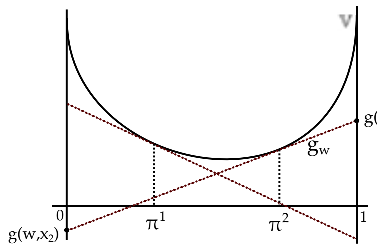


Fig. 1: Supporting hyperplanes on different priors

On the other hand, upper semi-continuity follows from the structure of the probability simplex and the Gale-Klee-Rockafellar theorem:

Theorem 3 (Gale-Klee-Rockafellar, [14])**.** *If f is convex and its domain is a polyhedron then it is upper semi-continuous.*

Hence, V_g is both lower semi-continuous (as the supremum of continuous functions) and upper semi-continuous (it is convex and $\mathbb{D}\mathcal{X}$ is a polyhedron), and any function satisfying both semi-continuities is necessarily continuous.

Corollary 4. *Any g-vulnerability V_g satisfies* `CNTY`*,* `CVX`*.*

B. `CNTY` *and* `CVX` *exactly characterize V_g*

Gain functions and g-vulnerability were introduced in [5], [6] in order to capture a variety of operational scenarios. Besides naturally retrieving Bayes-vulnerability as a special case, the flexibility of g-vulnerability allows us to retrieve other well-known entropy measures, such as Shannon- and guessing-entropy, using properly constructed gain functions [11], [12]. This suggests the question of how expressive g-vulnerabilities are in general.

Remarkably, it turns out that g-vulnerabilities are expressive enough to capture any vulnerability function $\mathbb{V}$ satisfying `CNTY` and `CVX`, although in the general case a countably infinite set $\mathcal{W}$ of guesses might be needed.

Theorem 5. *Let $\mathbb{V} : \mathbb{D}\mathcal{X} \to \mathbb{R}^+$ be a vulnerability function satisfying* `CNTY` *and* `CVX`*. Then there exists a gain function g with a countable number of guesses such that $\mathbb{V} = V_g$.*

The full proof is given in Appendix A; in the remainder of this section we try to convey the main arguments. A geometric view of gain functions is very helpful. Recall that $g_w(\pi)$, expressing the expected gain of a fixed guess w, is a linear function of π. A crucial observation is that the *graph* of g_w, that is the set of vectors $\{(\pi, g_w(\pi)) \mid \pi \in \mathbb{D}\mathcal{X}\}$, forms a hyperplane. Moreover, it can be shown that any such hyperplane[6] can be obtained as the graph of g_w by properly choosing the gains $g(w, x)$.

The correspondence between g_w and hyperplanes allows us to employ the *supporting hyperplane theorem*, which states that for any point s at the boundary of a convex set S, there is a hyperplane passing through s and leaving the whole set S on the same half space. Since $\mathbb{V}$ is a convex function, its *epigraph* $\mathrm{epi}\mathbb{V} = \{(\pi, y) \mid y \geq \mathbb{V}(\pi)\}$ is a convex set. Given any prior π^*, the point $(\pi^*, \mathbb{V}(\pi^*))$ lies on the boundary of $\mathrm{epi}\mathbb{V}$ hence there is a hyperplane passing from this point such that $\mathbb{V}$ lies above the hyperplane. Supporting hyperplanes on different priors are illustrated in Figure 1.

[6]To be precise, only the hyperplanes not orthogonal to that of probability distributions: see the full proof for details.

Since such a hyperplane can be constructed for each prior, we are going to use priors as guesses, making $\mathcal{W} = \mathbb{D}\mathcal{X}$. For a guess $w \in \mathbb{D}\mathcal{X}$ we choose the gains $g(w, x)$ such that the graph of g_w is exactly the supporting hyperplane passing through $(w, \mathbb{V}(w))$. Since $\mathbb{V}$ lies above the hyperplane, we get:

$$
\begin{aligned}
g_w(\pi) &= \mathbb{V}(\pi) && \text{for } w = \pi \text{ , and} \\
g_w(\pi) &\leq \mathbb{V}(\pi) && \text{for all } \pi \in \mathbb{D}\mathcal{X} \text{ .}
\end{aligned}
$$

Finally, from the definition of V_g we have that

$$
V_g(\pi) \;=\; \sup_{w \in \mathbb{D}\mathcal{X}} g_w(\pi) \;=\; \mathbb{V}(\pi) \ .
$$

The restriction to a countable set of guesses can be obtained by limiting w to priors with rational elements, and using the continuity of V_g. The details can be found in Appendix A.

V. Axiomatization of posterior vulnerabilities

In this section we consider axioms for posterior vulnerabilities and axioms that relate posterior and prior vulnerabilities. We investigate how different definitions of posterior vulnerabilities shape the interrelation among these postulates. We consider the following three axioms.

Non-interference (`NI`). The vulnerability of a point-hyper equals the vulnerability of the unique inner of this hyper:

$$
\forall \pi: \qquad \widehat{\mathbb{V}}[\pi] \;=\; \mathbb{V}(\pi) \ .
$$

This axiom means that an adversary who has learned with certainty that the secret follows distribution π has the same amount of information $\mathbb{V}(\pi)$ one would have had from π itself.

This postulate can also be interpreted in terms of non-interference. A channel C_{NI} is *non-interfering* if the result of pushing any prior π through C_{NI} is the point-hyper $[\pi]$, meaning that the adversary's state of knowledge is never changed by the observation of the output of the channel. It is well known that a channel C_{NI} is non-interfering iff all its rows are the same (see, for instance, [15]), so the simplest non-interfering channel is the null-channel, denoted here by $\bar{0}$, with only one column (i.e., every secret yields the same output). It can be easily verified that every non-interfering channel C_{NI} is equivalent to $\bar{0}$, since $[\pi, C_{NI}] = [\pi, \bar{0}] = [\pi]$. The `NI` axiom, then, is equivalent to stating that an adversary observing the output of a non-interfering channel does not gain or lose any information about the secret:

$$
\forall \pi: \qquad \widehat{\mathbb{V}}[\pi, \bar{0}] \;=\; \mathbb{V}(\pi) \ .
$$

Data-processing inequality (`DPI`). Post-processing does not increase vulnerability:[7]

$$
\forall \pi, C, R: \qquad \widehat{\mathbb{V}}[\pi, C] \;\geq\; \widehat{\mathbb{V}}[\pi, CR] \ ,
$$

[7]The data-processing inequality is a well well-known property of Shannon mutual information [16]: if $X \to Y \to Z$ forms a Markov chain, then $I(X;Y) \geq I(X;Z)$.

where the number of columns in matrix C is the same as the number of rows in matrix R.

This axiom can be interpreted as follows. Consider that a secret is fed into a channel C, and the produced output is, then, post-processed by being fed into another channel R (naturally the input domain of R must be the same as the output domain of C). Now consider two adversaries A and A' such that A can only observe the output of channel C, and A' can only observe the output of the cascade $C' = CR$. For any given prior π on secret values, A's posterior knowledge about the secret is given by the hyper $[\pi, C]$, whereas that of A''s is given by $[\pi, C']$. Note, however, that from A's knowledge it is always possible to reconstruct A''s, but the converse is not necessarily true.[8] Given this asymmetry, DPI formalizes that a vulnerability $\widehat{\mathbb{V}}$ should not evaluate A's information as any less than A''s.

Monotonicity (MONO). Pushing a prior through a channel does not decrease vulnerability:

$$\forall \pi, C: \qquad \widehat{\mathbb{V}}[\pi, C] \;\geq\; \mathbb{V}(\pi)\ .$$

One interpretation for this axiom is that by observing the output of a channel an adversary cannot lose information about the secret; in the worst case, the output can be ignored if it is not useful.[9] A direct consequence of this axiom is that, since posterior vulnerabilities are always greater than the corresponding prior vulnerabilities, additive and multiplicative versions of leakage as defined in Equations (2) and (3) are always non-negative.

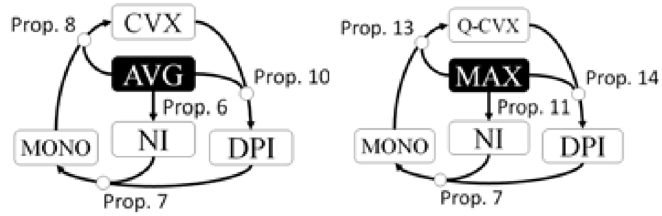
Having presented the three axioms of NI, DPI and MONO, we discuss next how posterior vulnerabilities can be defined so to respect them. Differently from the case of prior vulnerabilities, in which the axioms considered (CVX and CNTY) were sufficient to determine g-vulnerabilities as the unique family of prior-measures that satisfy them, our axioms for posterior vulnerabilities do not determine explicitly a unique family of posterior vulnerabilities $\widehat{\mathbb{V}}$. In the following sections we consider alternative definitions of posterior vulnerabilities and discuss the interrelation of the axioms they induce.

A. Posterior vulnerability as expectation

As seen in Section II, the posterior versions of Shannon-, guessing-, and min-entropy, as well as of g-vulnerability, are all defined as the expectation of the corresponding prior measures applied to each posterior distribution, weighted by the probability of each posterior's being realized. We will now consider the consequences of taking as an axiom the definition of posterior vulnerability as expectation.

[8] A can use π and C to compute $[\pi, CR']$ for any R', including the particular R used by A'. On the other hand, A' only knows π and C', and in general the decomposition of C' into a cascade of two channels is not unique (i.e., there may be several pairs C_i, R_i of matrices satisfying $C' = C_i R_i$), so it is not always possible for A' to uniquely recover C from C' and compute $[\pi, C]$.

[9] This axiom is a generalization of Shannon-entropy's *"information can't hurt"* property [16]: $H(X \mid Y) \leq H(X)$, for all random variables X, Y.



(a) Defining $\widehat{\mathbb{V}}$ as AVG (b) Defining $\widehat{\mathbb{V}}$ as MAX

Fig. 2: Equivalence of axioms. The merging arrows indicate joint implication: for example, on the left-hand side we have that MONO + AVG imply CVX.

Averaging (AVG). The vulnerability of a hyper is the expected value, with respect to the outer distribution, of the vulnerabilities of its inners:

$$\forall \Delta: \qquad \widehat{\mathbb{V}}\Delta \;=\; \mathrm{Exp}_{\Delta}\, \mathbb{V}\ ,$$

where the hyper $\Delta{:}\,\mathbb{D}^2\mathcal{X}$ might result from $\Delta{=}[\pi, C]$ for some π, C.

The main results of this section consist in demonstrating that by imposing AVG on a prior/posterior pair $(\mathbb{V}, \widehat{\mathbb{V}})$ of vulnerabilities, NI too is necessarily satisfied for this pair, and, furthermore, the axioms of CVX, DPI and MONO become equivalent to each other. Figure 2a illustrates these results.

Proposition 6 (AVG $\Rightarrow$ NI)**.** *If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies AVG, then it also satisfies NI.*

Proof: If AVG is assumed, for any prior π it is the case that $\widehat{\mathbb{V}}[\pi] = \mathrm{Exp}_{[\pi]}\,\mathbb{V} = \mathbb{V}(\pi)$, since $[\pi]$ is a point-hyper. □

Proposition 7 (NI + DPI $\Rightarrow$ MONO)**.** *If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies NI and DPI, then it also satisfies MONO.*

Proof: For any π, C, let $\bar{0}$ denote the non-interfering channel with only one column and as many rows as the columns of C. Then

$$\begin{array}{lll} & \widehat{\mathbb{V}}[\pi, C] & \\ \geq & \widehat{\mathbb{V}}[\pi, C\bar{0}] & \text{(by DPI)} \\ = & \widehat{\mathbb{V}}[\pi, \bar{0}] & (C\bar{0} = \bar{0}) \\ = & \widehat{\mathbb{V}}[\pi] & \text{(since } \bar{0} \text{ has only one column)} \\ = & \mathbb{V}(\pi) & \text{(by NI)} \end{array}$$

□

Proposition 8 (AVG + MONO $\Rightarrow$ CVX)**.** *If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies AVG and MONO, then it also satisfies CVX.*

Proof: Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be a finite set, and let π^1 and π^2 be distributions over $\mathcal{X}$. Let $0{\leq}a{\leq}1$, so that also $\pi^3 = a\pi^1 + (1{-}a)\pi^2$ is a distribution on $\mathcal{X}$. Define C^* to be the channel matrix

$$C^* \;=\; \begin{bmatrix} a\pi^1_1/\pi^3_1 & (1-a)\pi^2_1/\pi^3_1 \\ \vdots & \vdots \\ a\pi^1_i/\pi^3_i & (1-a)\pi^2_i/\pi^3_i \\ \vdots & \vdots \\ a\pi^1_n/\pi^3_n & (1-a)\pi^2_n/\pi^3_n \end{bmatrix} . \tag{4}$$

By pushing π^3 through C^* we obtain the hyper $[\pi^3, C^*]$ with outer distribution $(a, 1{-}a)$, and associated inners π^1 and π^2. Since AVG is assumed, we have

$$\widehat{\mathbb{V}}[\pi^3, C^*] \quad = \quad a\mathbb{V}(\pi^1) + (1{-}a)\mathbb{V}(\pi^2)\ . \tag{5}$$

But note that by MONO, we also have

$$\widehat{\mathbb{V}}[\pi^3, C^*] \;\geq\; \mathbb{V}(\pi^3) \;=\; \mathbb{V}(a\pi^1 + (1{-}a)\pi^2)\ . \tag{6}$$

Taking (5) and (6) together, we obtain CVX. □

For our next result, we will need the following lemma.

Lemma 9. *Let $X{\rightarrow}Y{\rightarrow}Z$ form a Markov chain with triply joint distribution $p(x,y,z) = p(x)p(y|x)p(z|y)$ for all $(x,y,z) \in \mathcal{X}{\times}\mathcal{Y}{\times}\mathcal{Z}$. Then $\sum_y p(y|z)p(x|y) = p(x|z)$ for all x, y, z.*

Proof: First we note that the probability of z depends only on the probability of y, and not x, so $p(z|x,y) = p(z|y)$ for all x, y, z. Then we can use the fact that

$$p(y,z)p(x,y) \;=\; p(x,y,z)p(y) \tag{7}$$

to derive:

$$\begin{array}{lll} & \sum_y p(y|z)p(x|y) & \\ = & \sum_y \frac{p(y,z)p(x,y)}{p(z)p(y)} & \\ = & \sum_y \frac{p(x,y,z)p(y)}{p(z)p(y)} & \text{(by Equation (7))} \\ = & \sum_y p(x,y|z) & \\ = & p(x|z) & \end{array}$$

□

Proposition 10 (AVG + CVX $\Rightarrow$ DPI)**.** *If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies AVG and CVX, then it also satisfies DPI.*

Proof: Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be sets of values. Let be π be a prior on $\mathcal{X}$, C be a channel from $\mathcal{X}$ to $\mathcal{Y}$, and R be a channel from $\mathcal{Y}$ to $\mathcal{Z}$. Note that the cascading CR of channels C and R is a channel from $\mathcal{X}$ to $\mathcal{Z}$.

Let $p(x,y,z)$ be the triply joint distribution defined as $p(x,y,z) = \pi_x C_{x,y} R_{y,z}$ for all $(x,y,z) \in \mathcal{X}{\times}\mathcal{Y}{\times}\mathcal{Z}$. By construction, this distribution has the special Markov property that the probability of z depends only on the probability of y, and not x. Thus $p(z \mid x,y) = p(z \mid y)$.

Note that, by pushing prior π through channel C, we obtain hyper $[\pi, C]$, in which the outer distribution on y is $p(y)$, and the inners are $p_{X|y}$. Thus we can derive:

$$\begin{array}{lll} & \widehat{\mathbb{V}}[\pi, C] & \\ = & \sum_y p(y)\mathbb{V}(p_{X|y}) & \text{(by AVG)} \\ = & \sum_y \left(\sum_z p(z)p(y \mid z)\right)\mathbb{V}(p_{X|y}) & \\ = & \sum_z p(z) \sum_y p(y \mid z)\mathbb{V}(p_{X|y}) & \\ \geq & \sum_z p(z)\mathbb{V}\left(\sum_y p(y \mid z)p_{X|y}\right) & \text{(by CVX)} \\ = & \sum_z p(z)\mathbb{V}(p_{X|z}) & \text{(by Lemma 9)} \\ = & \widehat{\mathbb{V}}[\pi, CR] & \text{(by AVG)} \end{array}$$

□

Appendix B provides a concrete illustration of Proposition 10.

B. Posterior vulnerability as maximum

An important consequence of AVG is that an observable happening with very small probability will have a negligible effect on $\widehat{\mathbb{V}}$, even if it completely reveals the secret. If such a scenario is not acceptable, an alternative approach is to consider the maximum information that may be obtained from any single output of the channel—produced with non-zero probability—no matter how small this probability is. This conservative approach is employed, for instance, in the original definition of *differential-privacy* [17].

We shall now consider the consequences of taking the following definition of $\widehat{\mathbb{V}}$ as an axiom.

Maximum (MAX). The vulnerability of a hyper is the maximum of the vulnerabilities of the inners in its support:

$$\forall \Delta: \qquad \widehat{\mathbb{V}}\Delta \;=\; \max_{\lceil \Delta \rceil} \mathbb{V}\ ,$$

where the hyper $\Delta{:}\,\mathbb{D}^2\mathcal{X}$ might result from $\Delta{=}[\pi, C]$ for some π, C.

The first result below shows that by imposing MAX on a prior/posterior pair $(\mathbb{V}, \widehat{\mathbb{V}})$ of vulnerabilities, NI is too satisfied for this pair.

Proposition 11. *[MAX $\Rightarrow$ NI] If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies MAX, then it also satisfies NI.*

Proof: If MAX is assumed, for any prior π we will have $\widehat{\mathbb{V}}[\pi] = \max_{\lceil [\pi] \rceil} \mathbb{V} = \mathbb{V}(\pi)$, since $[\pi]$ is a point-hyper. □

However, in contrast to the case of AVG, the symmetry among CVX, MONO and DPI is broken under MAX: although the axioms of MONO and DPI are still equivalent (shown later in this section, see Figure 2b), they are weaker than the axiom of CVX. This is demonstrated by the following example, showing a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfying MAX, MONO and DPI but not CVX.

Example 12 (MAX+MONO+DPI $\not\Rightarrow$ CVX)**.** *Consider the pair $(\mathbb{V}_1, \widehat{\mathbb{V}}_1)$ such that for every prior π and channel C:*

$$\mathbb{V}_1(\pi) = 1 - \left(\min_x \pi_x\right)^2, \qquad \textit{and}$$

$$\widehat{\mathbb{V}}_1[\pi, C] = \max_{\lceil [\pi, C] \rceil} \mathbb{V}_1.$$

To see that $\mathbb{V}_1$ does not satisfy CVX, consider distributions $\pi^1 = (0,1)$ and $\pi^2(1/2, 1/2)$, and its convex combination $\pi^3 = 1/2\,\pi^1 + 1/2\,\pi^2 = (1/4, 3/4)$. We calculate $\mathbb{V}_1(\pi^1) = 1 - 0^2 = 1$, $\mathbb{V}_1(\pi^2) = 1 - (1/2)^2 = 3/4$, $\mathbb{V}_1(\pi^3) = 1 - (1/4)^2 = 15/16$, and $1/2\,\mathbb{V}_1(\pi^1) + 1/2\,\mathbb{V}(\pi^2) = 7/8$ to conclude that $\mathbb{V}_1(\pi^3) > 1/2\,\mathbb{V}_1(\pi^1) + 1/2\,\mathbb{V}_1(\pi^2)$ and, hence, CVX is not satisfied.

The pair $(\mathbb{V}_1, \widehat{\mathbb{V}}_1)$ *satisfies* `MAX` *by construction. To show that it satisfies* `MONO` *and* `DPI`*, we first notice that* $\mathbb{V}_1$ *is quasiconvex. Using results from Figure 2b (proved later in this section), we conclude that* `MONO` *and* `DPI` *are also satisfied.*

The vulnerability function used in the counter-example above is quasiconvex. It turns out that this is not a coincidence: by replacing `CVX` with `Q-CVX` (a weaker property), the symmetry between the axioms can be restored. The remaining of this section establishes the equivalence of `Q-CVX`, `MONO` and `DPI` under `MAX`, as illustrated in Figure 2b.

Proposition 13. *[*`MAX` *+* `MONO` $\Rightarrow$ `Q-CVX`*] If a pair of prior/posterior vulnerabilities* $(\mathbb{V}, \widehat{\mathbb{V}})$ *satisfies* `MAX` *and* `MONO`*, then it also satisfies* `Q-CVX`*.*

Proof: By contradiction, let us assume that $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfy `MAX` and `MONO`, but does not satisfy `Q-CVX`.

Since `Q-CVX` is not satisfied, there must exist a value $0 \leq a \leq 1$ and three distributions π^1, π^2, π^3, such that $\pi^3 = a\pi^1 + (1-a)\pi^2$ and

$$\mathbb{V}(\pi^3) \;>\; \max\left(\mathbb{V}(\pi^1), \mathbb{V}(\pi^2)\right) . \tag{8}$$

Consider the channel C^* defined as in Equation (4). Then the hyper-distribution $[\pi^3, C^*]$ has outer distribution $(a, 1-a)$, and corresponding inner distributions π^1 and π^2. Since `MAX` is assumed, we have that

$$\widehat{\mathbb{V}}[\pi^3, C^*] \;=\; \max\left(\mathbb{V}(\pi^1), \mathbb{V}(\pi^2)\right) , \tag{9}$$

and because we assumed `MONO`, we also have that

$$\widehat{\mathbb{V}}[\pi^3, C^*] \;\geq\; \mathbb{V}(\pi^3) . \tag{10}$$

Equations (9) and (10) give $\mathbb{V}(\pi^3) \leq \max\left(\mathbb{V}(\pi^1), \mathbb{V}(\pi^2)\right)$, which contradicts our assumption in Equation (8). $\square$

Proposition 14. *[*`MAX` *+* `Q-CVX` $\Rightarrow$ `DPI`*] If a pair of prior/posterior vulnerabilities* $(\mathbb{V}, \widehat{\mathbb{V}})$ *satisfies* `MAX` *and* `Q-CVX`*, then it also satisfies* `DPI`*.*

Proof: Let π be a prior on $\mathcal{X}$, and C, R be channels from $\mathcal{X}$ to $\mathcal{Y}$ and from $\mathcal{Y}$ to $\mathcal{Z}$, respectively, with joint distribution $p(x, y, z)$ defined in the same way as in the proof of Proposition 10.

Note that, by pushing prior π through channel CR, we obtain hyper $[\pi, CR]$ in which the outer distribution on z is $p(z)$, and the inner are $p_{X|z}$. Thus we can derive:

$$\begin{array}{cll}
& \widehat{\mathbb{V}}[\pi, CR] & \\
= & \max_z \mathbb{V}(p_{X|z}) & \text{(by MAX)} \\
= & \max_z \mathbb{V}\left(\sum_y p(y \mid z) p_{X|y}\right) & \text{(by Lemma 9)} \\
\leq & \max_z \left(\max_y \mathbb{V}(p_{X|y})\right) & \text{(by Q-CVX)} \\
= & \max_y \mathbb{V}(p_{X|y}) & \\
= & \widehat{\mathbb{V}}[\pi, C] & \text{(by MAX)}
\end{array}$$

$\square$

Finally, note that, although `Q-CVX` is needed to recover the full equivalence of the axioms, `CVX` is strictly stronger than `Q-CVX`; hence, using a convex vulnerability measure (such as any V_g), `MONO` and `DPI` are still guaranteed under `MAX`.

Corollary 15. *[*`MAX` *+* `CVX` $\Rightarrow$ `MONO` *+* `DPI`*] If a pair* $(\mathbb{V}, \widehat{\mathbb{V}})$ *satisfies* `MAX` *and* `CVX`*, then it also satisfies* `MONO` *and* `DPI`*.*

Proof: Using the results of Figure 2b and the fact that `CVX` $\Rightarrow$ `Q-CVX`. $\square$

C. Other definitions of posterior vulnerabilities

In this section we explore the consequences of constraining posterior vulnerability more loosely than explicitly defining it as `AVG` or `MAX`. We require only that the posterior vulnerability cannot be greater than the vulnerability resulting from the most-informative channel output, nor less than the vulnerability resulting from the least-informative channel output.

Bounds (`BNDS`). The vulnerability of a hyper lies between the minimum and the maximum of the vulnerabilities of the inners in its support:

$$\forall \Delta: \qquad \min_{\lceil\Delta\rceil} \mathbb{V} \;\leq\; \widehat{\mathbb{V}}\Delta \;\leq\; \max_{\lceil\Delta\rceil} \mathbb{V} ,$$

where the hyper $\Delta{:}\,\mathbb{D}^2\mathcal{X}$ might result from $\Delta{=}[\pi, C]$ for some π, C.

The next results show that, whereas `BNDS` is strong enough to ensure `NI` (Proposition 16), by replacing `MAX` with `BNDS`, the equivalence among `Q-CVX`, `DPI` and `MONO` no longer holds (Example 12).

Proposition 16 (`BNDS` $\Rightarrow$ `NI`)**.** *If a pair of prior/posterior vulnerabilities* $(\mathbb{V}, \widehat{\mathbb{V}})$ *satisfies* `BNDS`*, then it also satisfies* `NI`*.*

Proof: If $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies `BNDS`, then $\min_{\lceil\Delta\rceil} \mathbb{V} \leq \widehat{\mathbb{V}}\Delta \leq \max_{\lceil\Delta\rceil} \mathbb{V}$ for every hyper Δ. Consider, then, the particular case when $\Delta = [\pi]$. Since $[\pi]$ is a point-hyper with inner π, we have that $\min_{\lceil[\pi]\rceil} \mathbb{V} = \max_{\lceil[\pi]\rceil} \mathbb{V} = \mathbb{V}(\pi)$. This in turn implies that $\mathbb{V}(\pi) \leq \widehat{\mathbb{V}}[\pi] \leq \mathbb{V}(\pi)$, which is `NI`. $\square$

The next example shows that under `BNDS`, not even `CVX`—which is stronger than `Q-CVX`—is sufficient to ensure `MONO` or `DPI`.

Example 17 (`BNDS` + `CVX` $\not\Rightarrow$ `MONO` or `DPI`)**.** *Consider the pair* $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ *such that for every prior* π *and hyper* Δ*:*

$$\mathbb{V}_2(\pi) = \max_x \pi_x, \qquad \textit{and}$$
$$\widehat{\mathbb{V}}_2\Delta = (\max_{\lceil\Delta\rceil} \mathbb{V}_2 + \min_{\lceil\Delta\rceil} \mathbb{V}_2)/2.$$

The pair $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ *satisfies* `BNDS`*, since* $\widehat{\mathbb{V}}_2$ *is the simple arithmetic average of maximum and minimum vulnerabilities of the inners. The pair* $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ *also satisfies* `CVX`*, since* $\mathbb{V}_2(\pi)$ *is just the Bayes vulnerability of* π*.*

To see that the pair $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ *does not satisfy* `MONO`*, consider prior* $\pi^2 = (9/10, 1/10)$ *and channel*

$$C_2 \;=\; \begin{array}{|c|c|} \hline 8/9 & 1/9 \\ \hline 0 & 1 \\ \hline \end{array} \; .$$

We can calculate that $\mathbb{V}_2(\pi^2) = 9/10$*, and that* $[\pi^2, C_2]$ *has outer distribution* $(4/5, 1/5)$*, and inner distributions* $(1, 0)$ *and* $(1/2, 1/2)$*. Hence*

$$\widehat{\mathbb{V}}_2[\pi^2, C_2] \;=\; (1+1/2)/2 \;=\; 3/4 ,$$

which violates `MONO` *because* $\widehat{\mathbb{V}}_2[\pi^2, C_2] < \mathbb{V}_2(\pi^2)$*.*

Now to see that the pair $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ does not satisfy `DPI`*, consider the prior $\pi^3 = (3/7, 4/7)$ and the channels*

$$C_3 \;=\; \begin{array}{|c|c|}\hline 1/3 & 2/3 \\ \hline 1/4 & 3/4 \\ \hline \end{array} \quad \textit{and} \quad R_3 \;=\; \begin{array}{|c|c|}\hline 1/4 & 3/4 \\ \hline 3/4 & 1/4 \\ \hline \end{array} \;.$$

We can calculate that $[\pi^3, C_3]$ has outer distribution $(2/7, 5/7)$, and inners $(1/2, 1/2)$ and $(2/5, 3/5)$. Hence

$$\widehat{\mathbb{V}}_2[\pi^3, C_3] \;=\; {}^{(1/2+3/5)}/_2 \;=\; 11/20 \;=\; 0.55 \;.$$

On the other hand, the cascade C_3R_3 yields the channel

$$C_3R_3 \;=\; \begin{array}{|c|c|}\hline 7/12 & 5/12 \\ \hline 5/8 & 3/8 \\ \hline \end{array} \;,$$

and we can calculate $[\pi^3, C_3R_3]$ to have outer distribution $(17/28, 11/28)$, and inners $(7/17, 10/17)$ and $(5/11, 6/11)$. Hence

$$\widehat{\mathbb{V}}[\pi^3, C_3R_3] \;=\; {}^{(10/17+6/11)}/_2 \;=\; 106/187 \;\approx\; 0.567 \;,$$

which makes $\widehat{\mathbb{V}}[\pi^3, C_3R_3] > \widehat{\mathbb{V}}[\pi^3, C_3]$, violating `DPI`*.*

VI. Discussion

In this section, we briefly discuss two applications of the results in Sections IV and V, showing how they can help to clarify the multitude of possible leakage measures.

One application concerns Rényi entropy [8], a family of entropy measures defined by

$$H_\alpha(\pi) = \frac{1}{1-\alpha} \lg \left(\textstyle\sum_{x \in \mathcal{X}} \pi_x^\alpha \right)$$

for $0 \leq \alpha \leq \infty$ (taking limits in the cases of $\alpha = 1$, which gives Shannon entropy, and $\alpha = \infty$, which gives min-entropy). It would be natural to use Rényi entropy to define a family of leakage measures by defining posterior Rényi entropy $\widehat{H}_\alpha$ using `AVG` and defining Rényi leakage by

$$\mathcal{L}_\alpha(\pi, C) = H_\alpha(\pi) - \widehat{H}_\alpha[\pi, C] \;.$$

However, it turns out that H_α is not concave for $\alpha > 2$. Therefore, by the dual version of Proposition 8, we find that Rényi leakage $\mathcal{L}_\alpha$ for $\alpha > 2$ would sometimes be *negative*. As an illustration, Figure 3 shows how the nonconcavity of min-entropy H_∞ can cause posterior min-entropy to be greater than prior min-entropy, giving negative min-entropy leakage.[10]

A second application concerns the robustness of the *composition refinement relation* $\sqsubseteq_\circ$ studied in [5], [6], [13]. Given channels C and D, both taking input X, C is composition refined by D, written $C \sqsubseteq_\circ D$, if $D = CR$ for some "refining" channel R. As proved in [5], [13], composition refinement is *sound* and *complete* for the *strong g-leakage ordering*: we have $C \sqsubseteq_\circ D$ iff the g-leakage of D never exceeds that of C, regardless of the prior π or gain function g. Still, we might worry that composition refinement implies a leakage ordering *only* with respect to g-leakage, leaving open the possibility that the leakage ordering might conceivably fail for some yet-to-be defined leakage measure. But our Propositions 8 and 10 show that if the hypothetical new leakage measure is defined using `AVG`, and never gives negative leakage, then it also satisfies the data-processing inequality `DPI`. And hence composition refinement is also sound for the new leakage measure.

[10]Note that min-entropy leakage, as defined in [4], does not in fact define posterior min-entropy using `AVG` but instead by $\widehat{H}_\infty[\pi, C] = -\lg \widehat{V}_b[\pi, C]$.

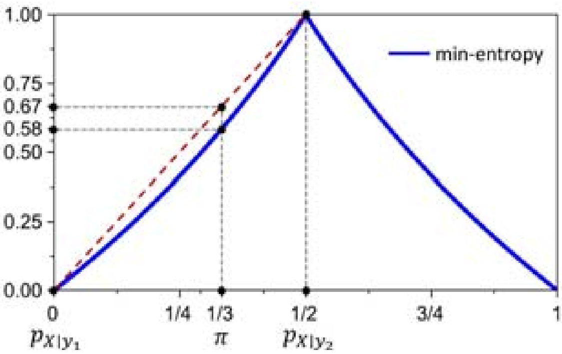


Fig. 3: A picture showing how posterior min-entropy can be greater than prior min-entropy. Pushing prior $\pi = (1/3, 2/3)$ through channel C gives hyper $[\pi, C]$ with outer $(1/3, 2/3)$ and inners $p_{X|y_1} = (0, 1)$ and $p_{X|y_2} = (1/2, 1/2)$. So $H_\infty(\pi) = -\lg 2/3 \approx 0.58$ and $\widehat{H}_\infty[\pi, C] = 1/3 \cdot 0 + 2/3 \cdot 1 \approx 0.67$.

VII. A categorical perspective

The axioms relating to Averaging are in fact instances of the monad laws proved by Giry for probabilistic computation [18], and in this section we give details. The benefit of this more general view is that it provides immediate access to well developed mathematical theories extending these results to infinite states and proper measures [19]. And this categorical perspective gives a direct connection to higher-order reasoning tools that dramatically simplify proofs, thereby leading directly to practical frameworks for calculating leakage [20].

The operator $\mathbb{D}$ that takes a sample space $\mathcal{X}$ to (discrete) distributions $\mathbb{D}\mathcal{X}$ on that space is widely recognised as the "probability monad", that is in effect a type constructor that obeys a small collection of laws shared by other, similar constructors like the powerset operator $\mathbb{P}$ [18]. Each monad has two polymorphic functions η, for "unit", and μ, for "multiply", that interact with each other in elegant ways. For example in (the) $\mathbb{P}$ (monad), unit has type $\mathcal{X} \to \mathbb{P}\mathcal{X}$ and ηx is $\{x\}$, the singleton set containing just x; in $\mathbb{D}$ we have type $\mathcal{X} \to \mathbb{D}\mathcal{X}$ and ηx is $[x]$, the point-distribution on x. In $\mathbb{P}$, multiply μ is distributed union that takes a set of sets to the one set that is the union of them all, having thus the type $\mathbb{P}^2\mathcal{X} \to \mathbb{P}\mathcal{X}$; and in $\mathbb{D}$ we have $(\mu\Delta)_x = \sum_{\delta \colon \lceil \Delta \rceil} \Delta_\delta \delta_x$, with μ thus of type $\mathbb{D}^2\mathcal{X} \to \mathbb{D}\mathcal{X}$ and taking the outer-weighted average of all the inner distributions δ in the support of hyper Δ: it is the "weighted average" of the hyper. This means, for example, that $\mu[\pi, C] = \pi$, i.e. that if you average the hyper produced by prior π and channel C you get the prior back again.

Furthermore the monadic type-constructors are *functors*, meaning they can be applied to functions as well as to objects: thus for f in $\mathcal{X} \to \mathcal{Y}$ the function $\mathbb{P}f$ of type $\mathbb{P}\mathcal{X} \to \mathbb{P}\mathcal{Y}$ is such that for X in $\mathbb{P}\mathcal{X}$ we have $f(X) = \{f(x) \mid x \in X\}$ in $\mathbb{P}\mathcal{Y}$. In

$\mathbb{D}$ instead we get the *push forward* of f, so that for π in $\mathbb{D}\mathcal{X}$ we have $(\mathbb{D}f)(\pi)_y = \sum_{f(x)=y} \pi_x$.

With these tools, some of our axioms can be expressed in a very general way, for example

(1) AVG becomes $\widehat{\mathbb{V}}{=}\mu{\circ}\mathbb{D}\mathbb{V}$.
(2) NI becomes $\widehat{\mathbb{V}}{\circ}\eta{=}\mathbb{V}$. Assuming (1), that follows from the general monad laws $\mu{\circ}\eta{=}1$ and $\mathbb{D}V{\circ}\eta{=}\eta{\circ}V$.
(3) CVX becomes $\mathbb{V}{\circ}\mu \leq \mu{\circ}\mathbb{D}\mathbb{V}$.

A consequence of accepting averaging (1) is that $\widehat{\mathbb{V}}(\Delta^1{}_p{+}\Delta^2) = \widehat{\mathbb{V}}(\Delta^1)\,{}_p{+}\,\widehat{\mathbb{V}}(\Delta^2)$, i.e. linearity of $\widehat{\mathbb{V}}$, where ${}_p{+}$ takes the p-weighted sum of its operands: on the left we sum over hypers; on the right we sum over scalars. This is more generally $\widehat{\mathbb{V}}(\mu\underline{\Delta}) = \mu((\mathbb{D}\widehat{\mathbb{V}})\underline{\Delta})$ where $\underline{\Delta}$ is in $\mathbb{D}^3\mathcal{X}$, a distribution of hypers, another monad law when $\widehat{\mathbb{V}}{=}\mu{\circ}\mathbb{D}\mathbb{V}$.

The space $\mathbb{D}^3\mathcal{X}$ also gives a hyper-formulated definition of the secrecy order $\sqsubseteq$ over hypers, i.e., that $\Delta^1{\sqsubseteq}\Delta^2$ just when $\widehat{\mathbb{V}}\Delta^1{\geq}\widehat{\mathbb{V}}\Delta^2$ for all $\widehat{\mathbb{V}}$ satisfying the axioms: it is that $\Delta^1{\sqsubseteq}\Delta^2$ just when there is a $\underline{\Delta}$ such that $\Delta^1{=}\mu\underline{\Delta}$ and $\Delta^2{=}(\mathbb{D}\mu)\underline{\Delta}$ [19], [21]. This formulation allows soundness of $\sqsubseteq$, i.e. that it can only decrease g-vulnerability, to be shown even for infinite state-spaces $\mathcal{X}$ and general measures. (See Appendix C.)

Finally, the monadic structure coupled with the Kantorovich metric gives us continuity criteria not only for $\mathbb{V}$ but also for $\widehat{\mathbb{V}}$ [18], [22]. If we give the underlying $\mathcal{X}$ the discrete metric, that $dist(x_1, x_2) = (0 \text{ if } x_1{=}x_2 \text{ else } 1)$, then the Kantorovich-induced distance on $\mathbb{D}\mathcal{X}$ is equivalent to the Manhattan metric, the notion used in the axiom CNTY. But the great generality of the monadic construction gives us that AVG, i.e., that $\widehat{\mathbb{V}} = \mu{\circ}\mathbb{D}\mathbb{V}$, makes $\widehat{\mathbb{V}}$ continuous as well, this time with respect to the Kantorovich metric on hypers. That in turn allows higher-order calculations that limit information flow in a very robust way [11].

VIII. Related work

In [23] Csiszár surveys the most commonly used postulates for a function f of the uncertainty contained in a finite probability distribution $(p_1, \ldots, p_n)$ for $n{>}0$. They are:

(P1) *Positivity*: $f(p_1, \ldots, p_n) \geq 0$;
(P2) *Expansibility*: $f(p_1, \ldots, p_n, 0) = f(p_1, \ldots, p_n)$;
(P3) *Symmetry*: $f(p_1, \ldots, p_n)$ is invariant under permutations of $(p_1, p_2, \ldots, p_n)$;
(P4) *Continuity*: $f(p_1, \ldots, p_n)$ is a continuous function of $(p_1, \ldots, p_n)$, for fixed n;
(P5) *Additivity*: $f(P{\times}Q) = f(P) + f(Q)$, where $P{\times}Q$ is the product-distribution of P and Q (i.e., the distribution in which events have probability $p_i q_j$ for each $p_i \in P$ and $q_j \in Q$);
(P6) *Subadditivity*: $f(A, B) \leq f(A) + f(B)$, where A and B are discrete random variables;
(P7) *Strong additivity*: $f(A, B) = f(A) + f(B|A)$;
(P8) *Recursivity*: $f(p_1, p_2, \ldots, p_n) = f(p_1{+}p_2, p_3, \ldots, p_n){+}(p_1{+}p_2)f\,(p_1/(p_1{+}p_2), p_2/(p_1{+}p_2))$;
(P9) *Sum-property*: $f(p_1, \ldots, p_n) = \sum_{k=1}^{n} g(p_k)$ for some function g.

Shannon-entropy is the only uncertainty measure to satisfy all axioms (P1)-(P9) listed by Csiszár; but in fact different subsets of these axioms are sufficient to fully characterize Shannon-entropy. In particular, Shannon himself showed that continuity, strong additivity, and the property that the uncertainty of a uniform distribution should not decrease as the number of elements in the distribution increases, are sufficient to determine entropy up to a constant factor [2]. Khinchin proved a similar result using strong additivity, expansibility, and the property that the maximum uncertainty should be realized in a uniform distribution [7].

In [8] Rényi explored ways to relax the axiomatization of Shannon-entropy to derive more general uncertainty measures. He showed that Shannon entropy could be characterized by five postulates: (R1) symmetry; (R2) continuity; (R3) $f(1/2, 1/2) = 1$; (R4) additivity; and (R5) the entropy of the union of two incomplete distributions is the arithmetic weighted average of each individual distribution. By replacing the weighted average in postulate (R5) with the (more relaxed) exponential mean, Rényi uniquely determined the family of Rényi entropies for full probability distributions $H_\alpha(p_1, p_2, \ldots, p_n) = 1/(1{-}\alpha)\lg(\sum_{k=1}^{n} p_k^\alpha)$, where $0{<}\alpha{\neq}1$ is a parameter. In the limit of α tending to 1, H_α coincides with Shannon-entropy, and in the limit of α tending to infinity, H_α is min-entropy, a measure that turned out to be highly relevant in the field quantitative information flow (QIF) [4].

Following Denning's seminal work [24], Shannon-entropy has been widely used in the field of QIF for the leakage of confidential information [1], [25]–[30]. But as the field of QIF evolved, new measures of uncertainty and of information have been proposed. Contrary to Rényi's motivation, however, most measures were not derived from mathematical principles, but instead were motivated by specific operational scenarios. Some examples are guessing-entropy [3], min-vulnerability [4], [6], [31], and g-vulnerability [5], to cite a few. Although many "healthiness properties" have been proved for these measures (e.g., non-negativity, non-decrease of uncertainty by post-processing, etc.), there has not always been a derivation of such measures from basic principles, or attempts to verify whether they can be unified in a more general framework.

Naturally, since measures other than Shannon-entropy cannot satisfy all postulates (P1)-(P9), the axioms for vulnerability considered in this paper differ from those listed by Csiszár. Some differences are unimportant: they are just adaptations of axioms of uncertainty to axioms of vulnerability (e.g., conditioning of random variables reduces uncertainty, but increases vulnerability, so some inequalities must be reversed).

Other differences are, however, more fundamental, as they reflect our departure from Shannon's obliviousness to the meaning of different secret values. The axiom of symmetry (P3), for instance, assumes that all secret values are equally informative, which is false in many scenarios: for instance, not everyone's bank account is as worth breaking in to as everyone else's—so evidently a permutation on the probabilities of every particular account being broken into does not amount to the same vulnerability. The axioms of additivity (P5), subadditivity (P6) and strong additivity (P7) assume that the uncertainty of a pair of joint random variables is a function only of the correlation of the random variables, which is also not a valid assumption in many security scenarios: the information of the combination of two secrets may exceed the information contents of each separate secret: for instance, the benefit of knowing someone's PIN-code and bank-account number at the same time greatly surpasses the sum of the benefits of knowing each separately. Recursivity (P8) and sum-property (P9) assume that the probability of each secret value contributes on equal terms to the overall uncertainty of the probability distribution, which also is a false assumption for many relevant measures. Bayes vulnerability, for instance, satisfies neither recursivity nor the sum-property, as the information of a probability distribution is a function of the maximum probability only.

Relation with Kifer and Lin's work. Kifer and Lin's work is the one most closely related to ours. In a series of papers [9], [32]–[34], these authors proposed an axiomatic characterization of "good" properties that sanitization mechanisms should provide, focusing in particular on *privacy* and *utility measures*. They considered utility as *information preservation*, which captures how "faithful" the output of the mechanism is to its input,[11] and as such is closely related to our notion of vulnerability. This notion derives from the more general concept of utility used in decision theory. Kifer and Lin argued that utility has not been studied systematically in the context of privacy, and that some proposals have led to inconsistencies and paradoxes.

In the following we summarize the connection between our paper and their work. We start by briefly recalling their basic concepts and notation. A sanitization mechanism $\mathcal{M}$ is a randomized algorithm from inputs to outputs,[12] whose behavior is described by conditional probabilities $P_{\mathcal{M}}(o|i)$ of observing output o when input is i. Such privacy mechanisms correspond exactly to our channels. Given two mechanisms $\mathcal{M}_1$ and $\mathcal{M}_2$ and $p \in [0,1]$, $\mathcal{M}_1 {\oplus_p} \mathcal{M}_2$ denotes the mechanism that, on input D, returns $\mathcal{M}_1(D)$ with probability p and $\mathcal{M}_2(D)$ with probability $1{-}p$, and also reveals whether the output was created using $\mathcal{M}_1$ or $\mathcal{M}_2$.

A measure of information preservation is a function μ mapping a mechanism $\mathcal{M}$ to a real value. Lin and Kifer [9] describe five axioms that such measures should satisfy:

(1) *Sufficiency*: $\mu(\mathcal{M}) \geq \mu(\mathcal{A} {\circ} \mathcal{M})$ for any randomized algorithm $\mathcal{A}$. Here $\circ$ represents functional composition.
(2) *Continuity*: μ is continuous in the components of $\mathcal{M}$ (viewed as a matrix).
(3) *Branching*: Given a mechanism $\mathcal{M}$ with output space $\{o_1, \dots o_n\}$ there is a function G such that $\mu(\mathcal{M}) = G(P_{\mathcal{M}}(o_1|\cdot), P_{\mathcal{M}}(o_2|\cdot)) {+} \mu(\mathcal{M}')$, where $\mathcal{M}'$ is obtained from $\mathcal{M}$ by adding together the columns $P_{\mathcal{M}}(o_1|\cdot), P_{\mathcal{M}}(o_2|\cdot)$ and leaving the others unchanged.
(4) *Quasi-convexity*: $\mu(\mathcal{M}_1 {\oplus_p} \mathcal{M}_2) {\leq} \max(\mu(\mathcal{M}_1), \mu(\mathcal{M}_2))$.
(5) *Quasi-concavity*: $\mu(\mathcal{M}_1 {\oplus_p} \mathcal{M}_2) {\geq} \min(\mu(\mathcal{M}_1), \mu(\mathcal{M}_2))$.

Lin and Kifer analyzed in [9] many popular measures of utility from the literature of privacy, and showed that almost all of them fail to satisfy the above axioms. One exception is the notion of g-vulnerability, as we will see in a moment.

By observing that our notion of vulnerability is essentially the utility of the adversary, we can make several connections between Kifer and Lin's principles and our own. First, their sufficiency axiom is clearly related to our data-processing inequality (`DPI`), since $\mathcal{A} {\circ} \mathcal{M}$ represents the post-processing of $\mathcal{M}$ by $\mathcal{A}$. Furthermore, they showed in [9] that Axioms (1)–(3) characterize a measure based on posterior g-vulnerability. More formally:[13]

Theorem 18 (Lin and Kifer [9], Theorem 6.2)**.**

- $\forall g \, \forall \pi \, \exists \mu$ *satisf. (1)-(5)* $: \ \forall \mathcal{M} \, \widehat{V}_g[\pi, \mathcal{M}] = \mu(\mathcal{M})$
- $\forall \mu$ *satisf. (1)-(3)* $\exists \pi \, \exists g : \ \forall \mathcal{M} \, \widehat{V}_g[\pi, \mathcal{M}] = \mu(\mathcal{M})$

From previous sections, we know that any function satisfying continuity (`CNTY`),[14] convexity (`CVX`), and averaging (`AVG`) corresponds to a posterior g-vulnerability for some g. Together with the above result, this suggests a strong relation between information preservation and the notion of average-based posterior vulnerability explored in this paper.

However, there are important differences. First of all, the type of μ and that of posterior vulnerability are different: posterior vulnerability applies to a hyper-distribution, typically derived from a channel $\mathcal{M}$ and a prior π. On the other hand, μ applies only to a channel $\mathcal{M}$. This means that the prior π is *implicitly encoded* into μ, and that the utility $\mu(\mathcal{M})$ is the utility of $\mathcal{M}$ under the fixed prior π. A second (related) difference is that, while we can express the prior vulnerability as a particular case of posterior vulnerability, this is not the case for μ. In fact, we can express the utility of the distribution π associated to μ as $\mu(\bar{0})$, but we cannot express the utility of a generic distribution via the same μ. Indeed, because of Axiom (1), for any $\mathcal{M}$, $\mu(\mathcal{M})$ has an utility greater than or equal to that of $\mu(\bar{0})$, thus it cannot represent the utility of any π' that has less utility than π. As a consequence, it seems that the relation between prior and posterior measures, which is a major contribution of our paper, cannot be expressed in Kifer and Lin's framework. At least, not by using μ alone: one would need to introduce and axiomatize a new function. In particular, the averaging axiom (`AVG`) cannot be formulated by using μ alone. Similarly, the maximum (`MAX`) and the bounds (`BNDS`) axioms cannot be formulated, despite the resemblance of the latter with the axioms (4) and (5) above.

[11] This is in contrast with utility as *usability*, which expresses how easily the output can be used. An example of the difference is provided by an encryption mechanism, which perfectly preserves information, but whose output is not usable except by users who know the decryption key.

[12] In Kifer and Lin's work, the inputs of a mechanism are assumed to be datasets, and denoted by D. However, the discussion of this section apply to inputs and outputs of any kind.

[13] Theorem 18 was actually formulated for the converse functions: the *information loss* and the *expected error of a Bayesian decision maker*, which are converse of the information preservation and of the posterior g-vulnerability, respectively.

[14] Note that (`CNTY`) and (2) refer to different type of arguments.

In summary, a main novelty with respect to the work of Kifer and Lin is that we investigate the relation between prior and posterior vulnerabilities. Another novel contribution is the study of the relationships between alternative sets of axioms. In general, indeed, our focus is different from that of Kifer and Lin: they focused on finding a collection of axioms for analyzing utility specifically, and used them to review the current practices in the field of privacy. In contrast, our main motivation is to establish the scientific principles which can help in the development or adaptation of new measures in response to novel situations. Thus, we explored different sets of possible axioms, thereby clarifying the implications between the principles themselves.

Relation with Boreale and Pampaloni's work. In [35], [36], Boreale and Pampaloni have conducted one of the first studies of adaptive adversaries in the context of quantitative information flow. They did not consider explicitly an axiomatic framework, but, in order for their results to be as general as possible, they adopted a generic notion of entropy, specified by a few properties which turn out to be our axioms of concavity, continuity, and averaging. Furthermore, in [36] they pointed out a known theorem in decision theory, which states that a function $H : \mathbb{D}\mathcal{X} \to \mathbb{R}^+$, satisfies concavity and continuity iff it is of the form $H(\pi) = \sum_x \pi_x S(x, \pi)$, where $S : \mathcal{X} \times \mathbb{D}\mathcal{X} \to \mathbb{R}^+$, is any function which satisfies the condition that $\sum_x \pi_x S(x, \pi')$ is minimal when $\pi' = \pi$. Such function S, called *Proper Scoring Rule* in decision theory, is similar to the (converse of) *gain functions* used in g-vulnerability, and therefore the above definition is related to that of prior g-entropy. Thus this result is similar to that of the completeness of g-vulnerability with respect to our axiomatization of the prior vulnerability (Theorem 5).

IX. Conclusion and Future Work

We have presented axioms that might be satisfied by intuitively reasonable measures of the prior- and posterior vulnerability of a secret as it is being processed by a system: this allowed us to derive properties of leakage. Our first main contribution was (1) the equivalence of the axioms of convexity, monotonicity (i.e. non-negativity of leakage), and data-processing inequality (DPI) when posterior vulnerability is defined as the average vulnerability of the posteriors, and (2) the equivalence of quasiconvexity, monotonicity and DPI when posterior vulnerability is defined as the worst-case vulnerability of posterior distributions. A deep implication of these results is that convexity (and quasiconvexity) of information measures do not need to be taken as fundamental properties, but are derivable from more intuitive principles, such as averaging (or worst-case analysis) and DPI.

The second main contribution was the demonstration of the soundness and completeness of g-vulnerabilities with respect to the axioms of convexity and continuity. Moreover, because of the equivalences we established, it follows that g-vulnerability exactly captures all average-based information measures that respect DPI or monotonicity.

We now want to further investigate the full family of vulnerabilities under quasiconvexity and continuity, characterizing all worst-case based vulnerabilities that respect DPI or monotonicity.

Acknowledgment

Mário S. Alvim was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). Geoffrey Smith was partially supported by the National Science Foundation under grant CNS-1116318. Morgan is supported by the Australian government via Data61 and, with McIver, via the ARC grant DP140101119. Also, the authors are grateful for support from Digiteo and the INRIA équipe associée Princess.

References

[1] D. Clark, S. Hunt, and P. Malacaria, “Quantitative information flow, relations and polymorphic types,” *Journal of Logic and Computation*, vol. 18, no. 2, pp. 181–199, 2005.

[2] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[3] J. L. Massey, “Guessing and entropy,” in *Proc. 1994 IEEE International Symposium on Information Theory*, 1994, p. 204.

[4] G. Smith, “On the foundations of quantitative information flow,” in *Proc. 12th International Conference on Foundations of Software Science and Computational Structures (FoSSaCS ’09)*, ser. Lecture Notes in Computer Science, L. de Alfaro, Ed., vol. 5504, 2009, pp. 288–302.

[5] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, “Measuring information leakage using generalized gain functions,” in *Proc. 25th IEEE Computer Security Foundations Symposium (CSF 2012)*, Jun. 2012, pp. 265–279.

[6] A. McIver, L. Meinicke, and C. Morgan, “Compositional closure for Bayes risk in probabilistic noninterference,” in *Proc. ICALP’10*, 2010, pp. 223–235.

[7] A. Khinchin, *Mathematical Foundations of Information Theory*, ser. Dover Books on Mathematics. Dover Publications, 1957.

[8] A. Rényi, “On measures of entropy and information,” in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley, Calif.: University of California Press, 1961, pp. 547–561.

[9] B. Lin and D. Kifer, “Information measures in statistical privacy and data processing applications,” *TKDD*, vol. 9, no. 4, p. 28, 2015.

[10] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden, “On the Bayes risk in information-hiding protocols,” *J. of Comp. Security*, vol. 16, no. 5, pp. 531–571, 2008.

[11] M. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, “Additive and multiplicative notions of leakage, and their capacities,” in *Computer Security Foundations Symposium (CSF), 2014 IEEE 27th*, July 2014, pp. 308–322.

[12] M. S. Alvim, A. Scedrov, and F. B. Schneider, “When not all bits are equal: Worth-based information flow,” in *Proc. 3rd Conference on Principles of Security and Trust (POST 2014)*, 2014, pp. 120–139.

[13] A. McIver, C. Morgan, G. Smith, B. Espinoza, and L. Meinicke, “Abstract channels and their robust information-leakage ordering,” in *Proc. 3rd Conference on Principles of Security and Trust (POST 2014)*, 2014, pp. 83–102.

[14] D. Gale, V. Klee, and R. Rockafellar, “Convex functions on convex polytopes,” *Proceedings of the American Mathematical Society*, vol. 19, no. 4, pp. 867–873, 1968.

[15] M. Bhargava and C. Palamidessi, “Probabilistic anonymity,” in *Proceedings of the 16th International Conference on Concurrency Theory (CONCUR 2005)*, ser. Lecture Notes in Computer Science, M. Abadi and L. de Alfaro, Eds., vol. 3653. Springer, 2005, pp. 171–185.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[17] C. Dwork, “Differential privacy,” in *Proc. of ICALP*, ser. LNCS, vol. 4052. Springer, 2006, pp. 1–12.

[18] M. Giry, "A categorical approach to probability theory," in *Categorical Aspects of Topology and Analysis*, ser. Lecture Notes in Mathematics. Springer Verlag, 1981, vol. 915, pp. 68–85.

[19] A. McIver, L. Meinicke, and C. Morgan, "A Kantorovich-monadic powerdomain for information hiding, with probability and nondeterminism," in *Proc. LiCS 2012*, 2012.

[20] T. Schrijvers. (2015) A monadic model for computations that leak secrets. [Online]. Available: http://www.cse.unsw.edu.au/~carrollm/LiCS15-TS.pdf

[21] A. McIver, L. Meinicke, and C. Morgan, "Hidden-markov program algebra with iteration," *Mathematical Structures in Computer Science*, vol. 25, no. 2, pp. 320–360, 2015.

[22] F. van Breugel, "The metric monad for probabilistic nondeterminism," 2005, draft available at http://www.cse.yorku.ca/~franck/research/drafts/monad.pdf.

[23] I. Csiszár, "Axiomatic characterizations of information measures," *Entropy*, vol. 10, no. 3, p. 261, 2008.

[24] D. Denning, *Cryptography and Data Security*. Addison-Wesley, 1983.

[25] P. Malacaria, "Assessing security threats of looping constructs," in *Proc. 34th Symposium on Principles of Programming Languages (POPL '07)*, 2007, pp. 225–235.

[26] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden, "Anonymity protocols as noisy channels," *Information and Computation*, vol. 206, pp. 378–401, 2008.

[27] P. Malacaria and H. Chen, "Lagrange multipliers and maximum information leakage in different observational models," in *Proc. of the 2008 Workshop on Programming Languages and Analysis for Security (PLAS 2008)*. ACM, June 2008, pp. 135–146.

[28] I. S. Moskowitz, R. E. Newman, and P. F. Syverson, "Quasi-anonymous channels," in *Proc. of CNIS*. IASTED, 2003, pp. 126–131.

[29] I. S. Moskowitz, R. E. Newman, D. P. Crepeau, and A. R. Miller, "Covert channels and anonymizing networks," in *Proceedings of the 2003 ACM Workshop on Privacy in the Electronic Society*, ser. WPES '03. New York, NY, USA: ACM, 2003, pp. 79–88.

[30] M. S. Alvim, M. E. Andrés, and C. Palamidessi, "Quantitative information flow in interactive systems," *Journal of Computer Security*, vol. 20, no. 1, pp. 3–50, 2012.

[31] C. Braun, K. Chatzikokolakis, and C. Palamidessi, "Quantitative notions of leakage for one-try attacks," in *Proc. 25th Conference on Mathematical Foundations of Programming Semantics (MFPS 2009)*, ser. ENTCS, vol. 249, 2009, pp. 75–91.

[32] D. Kifer and B.-R. Lin, "Towards an axiomatization of statistical privacy and utility," in *Proceedings of the Twenty-ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, ser. PODS '10. New York, NY, USA: ACM, 2010, pp. 147–158.

[33] B. Lin and D. Kifer, "Reasoning about privacy using axioms," in *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers, ACSCC 2012, Pacific Grove, CA, USA, November 4-7, 2012*, M. B. Matthews, Ed. IEEE, 2012, pp. 975–979.

[34] D. Kifer and B.-R. Lin, "An axiomatic view of statistical privacy and utility," *Journal of Privacy and Confidentiality*, vol. 4, no. 1, pp. 5–49, 2012.

[35] M. Boreale and F. Pampaloni, "Quantitative information flow under generic leakage functions and adaptive adversaries," in *Formal Techniques for Distributed Objects, Components, and Systems - 34th IFIP WG 6.1 International Conference, FORTE 2014, Held as Part of the 9th International Federated Conference on Distributed Computing Techniques, DisCoTec 2014, Berlin, Germany, June 3-5, 2014. Proceedings*, ser. Lecture Notes in Computer Science, E. Ábrahám and C. Palamidessi, Eds., vol. 8461. Springer, 2014, pp. 166–181.

[36] ——, "Quantitative information flow under generic leakage functions and adaptive adversaries," *Logical Methods in Computer Science*, vol. 11, no. 4, 2015.

Appendix A
Proofs of technical results.

A. Proofs of Section IV-A

The interior $\mathrm{int}(S)$ of a set $S \subseteq \mathbb{R}^n$ is the set of points $x \in S$ such that their is some ball centered at x which is contained in S. The boundary of S is $\mathrm{bd}(S) = S \setminus \mathrm{int}(S)$. A set is called open if $S = \mathrm{int}(S)$ and closed if its complement is open.

A function f is upper (resp. lower) semi-continuous, if its value around x_0 is close to $f(x_0)$ or less than (resp. greater than) $f(x_0)$. This can be formulated in terms of limits, but a simpler equivalent definition is the following: f is upper semi-continuous if the set

$$\{x \mid f(x) < \alpha\}$$

is open for all $\alpha \in \mathbb{R}$, and lower semi-continuous if the set $\{x \mid f(x) > \alpha\}$ is open for all $\alpha \in \mathbb{R}$.

Proposition 2. *If f is the supremum of a family of continuous functions then it is lower semi-continuous.*

Proof: Let $\mathcal{F}$ be a set of continuous functions and let $f(x) = \sup_{f' \in \mathcal{F}} f'(x)$. We show that f is lower semi-continuous.

Fix some $\alpha \in \mathbb{R}$. We need to show that $A = \{x \mid f(x) > \alpha\}$ is open. Let $x_0 \in A$; we are going to show that there exists a ball around x_0 contained in A. Since $\alpha < \sup_{f' \in \mathcal{F}} f'(x_0)$, there exists some $f' \in \mathcal{F}$ such that $f'(x_0) > \alpha$. Since f' is continuous, there exists some ball $B_\epsilon(x_0)$ such that $f'(x) > \alpha$ for all $x \in B_\epsilon(x_0)$. Hence $f(x) \geq f'(x) > \alpha$ for all $x \in B_\epsilon(x_0)$ which means that $B_\epsilon(x_0) \subseteq A$. $\square$

B. Proofs of Section IV-B

In this section we develop in full detail the line of reasoning of Section IV-B, leading to the proof of Theorem 5.

We start with a geometric view of gain functions. A hyperplane is the set of vectors x satisfying

$$a \cdot x = b$$

for some $a \in \mathbb{R}^n, a \neq \mathbf{0}$ (the normal) and $b \in \mathbb{R}$. The hyperplane splits $\mathbb{R}^n$ into two closed half-spaces

$$a \cdot x \leq b \qquad \text{and} \qquad a \cdot x \geq b$$

A hyperplane supports a set S if S lies within of the two closed half-spaces and at least one point lies on the hyperplane. The supporting hyperplane theorem states that if S is convex and $x \in \mathrm{bd}(S)$ then there exists a supporting hyperplane that contains x.

Geometrically, a guess w can be thought of as a vector in $\mathbb{R}^n$, for $n = |\mathcal{X}|$[15] containing the loss for each $x \in \mathcal{X}$. In this case g_w (giving the expected gain of a specific guess w) can be simply expressed as the dot product:

$$g_w(\pi) = \pi \cdot w$$

For fixed w, the graph of $g_w(\pi)$ is a hyperplane on $\mathbb{D}\mathcal{X} \times \mathbb{R}$ with parameters $a = (-w, 1)$ and $b = 0$, since

$$(-w, 1) \cdot (\pi, y) = 0 \quad \Leftrightarrow \quad y = \pi \cdot w$$

Conversely, any hyperplane on $\mathbb{D}\mathcal{X} \times \mathbb{R}$ of the form

$$(a, 1) \cdot (\pi, y) = b \qquad a \in \mathbb{R}^n, b \in \mathbb{R}$$

[15]Note that we represent priors and guesses as $|\mathcal{X}|$-dimensional vectors. However, it is sometimes convenient to drop the last coordinate: in the case $|\mathcal{X}| = 2$ a prior $(x, 1-x)$ can be represented by a single point x (Figure 1).

(which means that the last coordinate of the normal should not be 0; any other value can be scaled to 1) is the graph of $g_w(\pi)$ for a guess vector $w = b\mathbf{1} - a$.

In the case of two dimensions ($|\mathcal{X}| = 2$), priors can be represented by a single real $x = \pi_{x_1} \in [0, 1]$, and the graph of g_w is a line, as illustrated in Figure 1. Note that, the individial gains $g(w, x_1), g(w, x_2)$ correspond to the points where the hyperplane hits the lines $x = 0$ and $x = 1$.

The idea for creating a g-vulnerability V_g that coincides with an arbitrary $\mathbb{V}$ is to create one guess for each prior $\pi \in \mathbb{D}\mathcal{X}$ and obtain a supporting hyperplane passing through $(\pi, \mathbb{V}(\pi))$, as shown in Figure 1. The hyperplane is the graph of $g_w(\pi)$, for a suitably constructed vector w. Since all hyperplanes are below $\mathbb{V}$, and at least one is touching $\mathbb{V}$ at each π, then V_g coincides with $\mathbb{V}$. Extra care is needed to avoid hyperplanes *orthogonal* to the *probability hyperplane*, since those cannot be expressed as $g_w(\pi)$.

Theorem 5. *Let $\mathbb{V} : \mathbb{D}\mathcal{X} \to \mathbb{R}^+$ be a vulnerability function satisfying* `CNTY` *and* `CVX`*. Then there exists a gain function g with a countable number of guesses such that $\mathbb{V} = V_g$.*

Proof: Let A be the elements of the *(relative) interior* of $\mathbb{D}\mathcal{X}$ (i.e., $\pi[i] > 0, \forall i$ and $\sum_i \pi[i] < 1$) having *rational* coordinates. We are going to create one guess w_π for each such $\pi \in A$. Since epi$\mathbb{V}$ is convex, and $(\pi, \mathbb{V}(\pi)) \in \mathrm{bd}(\mathrm{epi}\mathbb{V})$, from the supporting hyperplane theorem there exists a hyperplane

$$(a, c) \cdot (\pi', y) = b \qquad a \in \mathbb{R}^n, a \neq \mathbf{0}, b, c \in \mathbb{R}$$

containing $(\pi, \mathbb{V}(\pi))$, and such that epi$\mathbb{V}$ lies above the hyperplane, i.e., $(a, c) \cdot (\pi', \mathbb{V}(\pi')) \geq b$ for all $\pi' \in \mathbb{D}\mathcal{X}$.

In general, the hyperplane might have $c = 0$, which happens iff it is orthogonal to the hyperplane $(\mathbf{0}, 1) \cdot (\pi', y) = 0$ of probability distributions (i.e., the hyperplane containing only vectors of the form $(\pi', 0)$). We now show that this can only happen at the (relative) boundary of $\mathbb{D}\mathcal{X}$, that is, since $\pi \in \mathrm{int}(\mathbb{D}\mathcal{X})$, we must have $c \neq 0$. Assuming $c = 0$, we have that $\pi \cdot a = b$. Since π is an interior point, there exists a ball $B_\epsilon(\pi) \subseteq \mathbb{D}\mathcal{X}$. The hyperplane passes through the center of the ball, so there exist points in the ball on both sides of the hyperplane. Thus take $\pi' \in B_\epsilon(\pi)$ such that $\pi' \cdot a < b$; hence $(a, c) \cdot (\pi', \mathbb{V}(\pi')) < b$ which is a contradiction.

Now since $c \neq 0$, the hyperplane is the graph of g_w for $w = c^{-1}(b\mathbf{1} - a)$. Hence we have that

$$\begin{array}{rcll} w \cdot \pi & = & \mathbb{V}(\pi) & \quad \text{for } w = \pi \text{ , and} \\ w \cdot \pi & \leq & \mathbb{V}(\pi) & \quad \text{for all } \pi \in \mathbb{D}\mathcal{X} \text{ .} \end{array}$$

Creating one such guess for each element of A, we have

$$V_g(\pi) = \sup_{w \in A} w \cdot \pi = \mathbb{V}(\pi)$$

for all $\pi \in A$, i.e., V_g and $\mathbb{V}$ coincide on A.

Finally, since all irrationals are the limit of a sequence of rationals, and boundary points are the limit of a sequence of interior points, from continuity we conclude that V_g and $\mathbb{V}$ coincide everywhere. □

Appendix B
Concrete illustration of Proposition 10

Example 19. *Let channels C and R be as follows:*

C	y_1	y_2	y_3
x_1	$1/2$	$1/4$	$1/4$
x_2	$1/4$	$3/4$	0

R	z_1	z_2	z_3
y_1	$1/2$	$1/2$	0
y_2	$1/3$	$1/3$	$1/3$
y_3	1	0	0

With prior $\pi = (3/4, 1/4)$, we get the two hypers:

$[\pi, C]$	$7/16$	$3/8$	$3/16$
x_1	$6/7$	$1/2$	1
x_2	$1/7$	$1/2$	0

$[\pi, CR]$	$17/32$	$11/32$	$1/8$
x_1	$14/17$	$8/11$	$1/2$
x_2	$3/17$	$3/11$	$1/2$

Now we show the steps that establish $\widehat{\mathbb{V}}[\pi, C] \geq \widehat{\mathbb{V}}[\pi, CR]$, assuming the axioms of `AVG` *and* `CVX`*:*

$$\begin{array}{cll}
& \widehat{\mathbb{V}}[\pi, C] & \\
= & \sum_y p(y)\mathbb{V}(p_{X|y}) & \textit{(by AVG)} \\
= & \frac{7}{16}\mathbb{V}\begin{bmatrix}\frac{6}{7}\\ \frac{1}{7}\end{bmatrix} + \frac{3}{8}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} + \frac{3}{16}\mathbb{V}\begin{bmatrix}1\\ 0\end{bmatrix} & \\
= & \left(\frac{7}{32} + \frac{7}{32}\right)\mathbb{V}\begin{bmatrix}\frac{6}{7}\\ \frac{1}{7}\end{bmatrix} + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right)\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} + \frac{3}{16}\mathbb{V}\begin{bmatrix}1\\ 0\end{bmatrix} & \\
= & \left(\frac{7}{32}\mathbb{V}\begin{bmatrix}\frac{6}{7}\\ \frac{1}{7}\end{bmatrix} + \frac{1}{8}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} + \frac{3}{16}\mathbb{V}\begin{bmatrix}1\\ 0\end{bmatrix}\right) & \\
& \quad + \left(\frac{7}{32}\mathbb{V}\begin{bmatrix}\frac{6}{7}\\ \frac{1}{7}\end{bmatrix} + \frac{1}{8}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}\right) + \frac{1}{8}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} & \\
= & \frac{17}{32}\left(\frac{7}{17}\mathbb{V}\begin{bmatrix}\frac{6}{7}\\ \frac{1}{7}\end{bmatrix} + \frac{4}{17}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} + \frac{6}{17}\mathbb{V}\begin{bmatrix}1\\ 0\end{bmatrix}\right) & \\
& \quad + \frac{11}{32}\left(\frac{7}{11}\mathbb{V}\begin{bmatrix}\frac{6}{7}\\ \frac{1}{7}\end{bmatrix} + \frac{4}{11}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}\right) + \frac{1}{8}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} & \\
\geq & \frac{17}{32}\mathbb{V}\left(\frac{7}{17}\begin{bmatrix}\frac{6}{7}\\ \frac{1}{7}\end{bmatrix} + \frac{4}{17}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} + \frac{6}{17}\begin{bmatrix}1\\ 0\end{bmatrix}\right) & \textit{(by CVX)} \\
& \quad + \frac{11}{32}\mathbb{V}\left(\frac{7}{11}\begin{bmatrix}\frac{6}{7}\\ \frac{1}{7}\end{bmatrix} + \frac{4}{11}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}\right) + \frac{1}{8}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} & \\
= & \frac{17}{32}\mathbb{V}\begin{bmatrix}\frac{14}{17}\\ \frac{3}{17}\end{bmatrix} + \frac{11}{32}\mathbb{V}\begin{bmatrix}\frac{8}{11}\\ \frac{3}{11}\end{bmatrix} + \frac{1}{8}\mathbb{V}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} & \\
= & \sum_z p(z)\mathbb{V}(p_{X|z}) & \\
= & \widehat{\mathbb{V}}[\pi, CR] &
\end{array}$$

Appendix C
Elementary examples supporting Section VII

Recall that we take our base space to be $\mathcal{X}$, distributions on that to be $\mathbb{D}\mathcal{X}$, and hypers to be $\mathbb{D}^2\mathcal{X}$. Typical elements of $\mathbb{D}\mathcal{X}$ are lower-case Greek letters, possibly superscripted. Thus π_x is the probability π assigns to x and π^1_x is the probability π^1 assigns to x and $\Delta^2_{\delta^3}$ is the probability that hyper $\Delta^2{:}\,\mathbb{D}^2\mathcal{X}$ assigns to distribution $\delta^3{:}\,\mathbb{D}\mathcal{X}$. In Δ_δ usually δ will be in the support $\lceil\Delta\rceil$ of Δ; if not, then of course the assigned probability is zero.

We (continue to) write the point, or "singleton" distribution on x as $[x]$, so that $[x]_{x'} = (1 \text{ if } x{=}x' \text{ else } 0)$ — it is the same as ηx if we are using a monad. A "doubleton" distribution say $\delta{=}x_1{}_p{\oplus}x_2$ is such that $\delta_{x_1}{=}p$ and $\delta_{x_2}{=}1{-}p$. The p-weighted sum of two values is defined $x_1{}_p{+}x_2 = px_1{+}(1{-}p)x_2$, thus not the same thing as $x_1{}_p{\oplus}x_2$: for example if $\mathcal{X}$ were the reals $\mathbb{R}$, then $x_1{}_p{\oplus}x_2$ would also be a real, but $x_1{}_p{\oplus}x_2$ would be a (doubleton) distribution in $\mathbb{D}\mathbb{R}$. Indeed we have $x_1{}_p{\oplus}x_2 = [x_1]_p{+}[x_2]$. In both cases the p-factor applies on the left.

In this section we use μ more generally than multiply of a monad, as introduced in Section VII above: here μ will as well simply average *any* distribution taken over a vector space. Thus in particular we have $\mu(\delta^1{}_p{\oplus}\delta^2) = \delta^1{}_p{+}\delta^2$ because $\delta^1{}_p{\oplus}\,\delta^2$, a hyper with just two inners, is in $\mathbb{D}^2\mathcal{X} = \mathbb{D}(\mathbb{D}\mathcal{X})$ and $\mathbb{D}\mathcal{X}$ is a vector space.

We return first return to the higher-order formulation $\widehat{\mathbb{V}}{=}\mu{\circ}\mathbb{D}\mathbb{V}$ of $\texttt{AVG}$. With a doubleton hyper for illustration, say $\Delta = \pi^1{}_p{\oplus}\pi^2$, that gives

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}}\Delta = (\mu{\circ}\mathbb{D}\mathbb{V})\Delta & (\text{apply } \texttt{AVG} \text{ to } \Delta) \\
\Rightarrow & \widehat{\mathbb{V}}(\pi^1{}_p{\oplus}\pi^2) = (\mu{\circ}\mathbb{D}\mathbb{V})(\pi^1{}_p{\oplus}\pi^2) & (\Delta = \pi^1{}_p{\oplus}\pi^2) \\
\text{iff} & \widehat{\mathbb{V}}(\pi^1{}_p{\oplus}\pi^2) = \mu(\mathbb{D}\mathbb{V}(\pi^1{}_p{\oplus}\pi^2)) & (\text{composition}) \\
\text{iff} & \widehat{\mathbb{V}}(\pi^1{}_p{\oplus}\pi^2) = \mu(\mathbb{V}\pi^1{}_p{\oplus}\mathbb{V}\pi^2) & (\text{definition functor } \mathbb{D}) \\
\text{iff} & \widehat{\mathbb{V}}(\pi^1{}_p{\oplus}\pi^2) = \mathbb{V}\pi^1{}_p{+}\mathbb{V}\pi^2\ , & (\text{property of } \mu)
\end{array}
$$

showing that $\widehat{\mathbb{V}}$ indeed takes the weighted sum of $\mathbb{V}$ applied to the (two, in this case) posteriors in Δ. As an illustration of more general reasoning, we give the proof promised in Section VII, i.e., that applies even when the weighted average is not necessarily over just two hypers. We have

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}} \circ \mu & \\
= & \mu \circ \mathbb{D}\mathbb{V} \circ \mu & (\text{assumption } \texttt{AVG}) \\
= & \mu \circ \mu \circ \mathbb{D}^2\mathbb{V} & (\mu \text{ is natural transformation } \mathbb{D}^2{\to}\mathbb{D}) \\
= & \mu \circ \mathbb{D}\mu \circ \mathbb{D}^2\mathbb{V} & (\text{monad coherence condition on } \mu) \\
= & \mu \circ \mathbb{D}(\underline{\mu \circ \mathbb{D}\mathbb{V}}) & (\mathbb{D} \text{ functor}) \\
= & \mu \circ \mathbb{D}\widehat{\mathbb{V}}\ , & (\text{assumption } \texttt{AVG})
\end{array}
$$

which overall equality says intuitively that applying $\widehat{\mathbb{V}}$ to the weighted sum of some hypers, i.e., $\widehat{\mathbb{V}}(\mu\underline{\Delta})$, is the same as applying $\widehat{\mathbb{V}}$ to the hypers separately and then taking the weighted sum of the results, i.e., $\mu(\mathbb{D}\widehat{\mathbb{V}}\underline{\Delta})$.

The higher-order $\texttt{NI}$ captures its traditional definition via $\widehat{\mathbb{V}}[\pi]{=}\widehat{\mathbb{V}}(\eta\pi){=}(\widehat{\mathbb{V}}{\circ}\eta)\pi{=}\mathbb{V}\pi$. Here is how the higher-order version of $\texttt{NI}$ follows from $\texttt{AVG}$ and the monad laws, as we claimed in Section VII:

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}}{\circ}\eta & \\
= & \mu \circ \mathbb{D}\mathbb{V} \circ \eta & (\text{assumption } \texttt{AVG}) \\
= & \mu \circ \eta \circ \mathbb{V} & (\eta \text{ is natural transformation } 1{\to}\mathbb{D}) \\
= & \mathbb{V}\ . & (\text{monad coherence condition } \mu{\circ}\eta = 1)
\end{array}
$$

Applying that to arbitrary π gives $\widehat{\mathbb{V}}(\eta\pi) = \mathbb{V}\pi$.

For $\texttt{CVX}$ again we use a doubleton hyper $\Delta = \pi^1{}_p{\oplus}\pi^2$ as an example, so that the traditional formulation of $\texttt{CVX}$ is found in the middle of the following string of equalities:

$$
\begin{array}{lll}
 & (\mathbb{V}{\circ}\mu)\Delta & \\
= & (\mathbb{V}{\circ}\mu)(\pi^1{}_p{\oplus}\pi^2) & (\text{definition } \Delta) \\
= & \mathbb{V}(\mu(\pi^1{}_p{\oplus}\pi^2)) & (\text{composition}) \\
= & \mathbb{V}(\pi^1{}_p{+}\pi^2) & (\text{property of } \mu) \\
\le & \mathbb{V}\pi^1{}_p{+}\,\mathbb{V}\pi^2 & (\text{traditional formulation of } \texttt{CVX}) \\
= & \mu(\mathbb{V}\pi^1{}_p{\oplus}\,\mathbb{V}\pi^2) & (\text{property of } \mu) \\
= & \mu(\mathbb{D}\mathbb{V}(\pi^1{}_p{\oplus}\,\pi^2)) & (\text{definition functor } \mathbb{D}) \\
= & (\mu{\circ}\mathbb{D}\mathbb{V})(\pi^1{}_p{\oplus}\,\pi^2) & (\text{composition}) \\
= & (\mu{\circ}\mathbb{D}\mathbb{V})\Delta\ , & (\text{defintion } \Delta \text{ again})
\end{array}
$$

showing how $\texttt{CVX}$ for this particular Δ agrees with the higher-order formuation $\mathbb{V}{\circ}\mu \le \mu{\circ}\mathbb{D}\mathbb{V}$.

Now we return to the formulation of the partial order $\sqsubseteq$ on hypers in terms of the surprising "hyper-hyper" $\underline{\Delta}$ in $\mathbb{D}^3\mathcal{X}$. As we did above (at †), we will assist the intuition by taking a simple case $\underline{\Delta} = \Delta^1{}_p{\oplus}\Delta^2$, thus a doubleton hyper-hyper over two hypers Δ^1 with probability p and Δ^2 with probability $1{-}p$. We show that the higher-order definition of $\sqsubseteq$ implies the $\mathbb{V}$-based definition, in this case, provided we assume $\texttt{CVX}$. (The reverse direction is harder, related to the *Coriaceous Conjecture* described in [5] and proved in [6], [11].)

We start by setting $\Delta^{+}{=}\mu\underline{\Delta}$ and $\Delta^{-}{=}(\mathbb{D}\mu)\underline{\Delta}$, as in the higher-order formulation of $\sqsubseteq$ from which we would expect to be able to prove that $\widehat{\mathbb{V}}\Delta^{+} \ge \widehat{\mathbb{V}}\Delta^{-}$. Then we have

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}}\Delta^{+} & \\
= & \widehat{\mathbb{V}}(\mu\underline{\Delta}) & (\text{definition } \Delta^{+}) \\
= & \widehat{\mathbb{V}}(\mu(\Delta^1{}_p{\oplus}\Delta^2)) & (\text{definition } \underline{\Delta}) \\
= & \widehat{\mathbb{V}}(\Delta^1{}_p{+}\Delta^2) & (\text{property of } \mu) \\
= & \widehat{\mathbb{V}}\Delta^1{}_p{+}\,\widehat{\mathbb{V}}\Delta^2 & (\text{linearity of } \widehat{\mathbb{V}}, \text{ implied by } \texttt{AVG}) \\
= & (\mu{\circ}\mathbb{D}\mathbb{V})\Delta^1{}_p{+}\,(\mu{\circ}\mathbb{D}\mathbb{V})\Delta^2 & (\text{assume } \texttt{AVG}) \\
\ge & (\mathbb{V}{\circ}\mu)\Delta^1{}_p{+}\,(\mathbb{V}{\circ}\mu)\Delta^2 & (\text{assume } \texttt{CVX}) \\
= & \mu((\mathbb{V}{\circ}\mu)\Delta^1{}_p{\oplus}\,(\mathbb{V}{\circ}\mu)\Delta^2) & (\text{property } \mu) \\
= & \mu(\mathbb{D}(\mathbb{V}{\circ}\mu)(\Delta^1{}_p{\oplus}\,\Delta^2)) & (\text{functor } \mathbb{D}) \\
= & (\underline{\mu{\circ}\mathbb{D}\mathbb{V}}{\circ}\mathbb{D}\mu)\underline{\Delta} & (\text{composition; functor } \mathbb{D}\text{; defintion } \underline{\Delta}) \\
= & \widehat{\mathbb{V}}(\mathbb{D}\mu\underline{\Delta}) & (\text{assume } \texttt{AVG}\text{; composition}) \\
= & \widehat{\mathbb{V}}\Delta^{-}\ . & (\text{definition } \Delta^{-})
\end{array}
$$

The hard-core higher-order proof, for general $\underline{\Delta}$, is in effect a soundness proof for $\sqsubseteq$, that it can only decrease vulnerability (given $\texttt{CVX}$ and $\texttt{AVG}$); and because of the great generality of the monad framework [18] it applies even for infinite state spaces $\mathcal{X}$ and measures. Although less intuitive (at first), it is much shorter:

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}} \circ \mu & \\
= & \mu \circ \mathbb{D}\widehat{\mathbb{V}} & (\text{linearity of } \widehat{\mathbb{V}}, \text{ proved earlier at †}) \\
\ge & \mu \circ \mathbb{D}(\mathbb{V} \circ \mu) & (\texttt{AVG} \text{ and } \texttt{CVX}\text{; see ‡ below}) \\
= & \underline{\mu \circ \mathbb{D}\mathbb{V}} \circ \mathbb{D}\mu & (\mathbb{D} \text{ functor}) \\
= & \widehat{\mathbb{V}} \circ \mathbb{D}\mu\ , & (\texttt{AVG} \text{ with "the other } \mu\text{"})
\end{array}
$$

so that the longer proof just above is recovered by applying each line to $\underline{\Delta} = \Delta^1{}_p{\oplus}\Delta^2$.

The "see below" appeals to the elementary general fact that if two functions $f, f'{:}\,\mathcal{S}{\to}\mathbb{R}$ satisfy $f(s){\ge}f'(s)$ for all $s{:}\,\mathcal{S}$, then also $(\mu{\circ}\mathbb{D}f)(\delta) \ge (\mu{\circ}\mathbb{D}f')(\delta)$ for all δ in $\mathbb{D}\mathcal{S}$, in words that if two random variables over the same distribution satisfy $\ge$ everwhere, then so do their expected values. Above we used $f{=}\widehat{\mathbb{V}}$ and $f'{=}\mathbb{V}{\circ}\mu$ and $\mathcal{S}{=}\mathbb{D}^2\mathcal{X}$, appealing to $\texttt{AVG}$ and $\texttt{CVX}$ for the inequality.