How to Develop an Intuition for Risk... and Other Invisible Phenomena

Natasha Fernandes
School of Engineering and IT, UNSW Canberra, Australia

Annabelle McIver
Department of Computing, Macquarie University, Sydney, Australia

Carroll Morgan
School of Computer Science and Engineering, UNSW, Sydney, Australia

Abstract

The study of quantitative risk in security systems is often based around complex and subtle mathematical ideas involving probabilities. The notations for these ideas can pose a communication barrier between collaborating researchers even when those researchers are working within a similar framework.

This paper describes the use of geometrical representation and reasoning as a way to share ideas using the minimum of notation so as to build intuition about what kinds of properties might or might not be true. We describe a faithful geometrical setting for the channel model of quantitative information flow (QIF) and demonstrate how it can facilitate "proofs without words" for problems in the QIF setting.

2012 ACM Subject Classification Security and privacy → Formal methods and theory of security

Keywords and phrases Geometry, Quantitative Information Flow, Proof, Explainability, Privacy

Digital Object Identifier 10.4230/LIPIcs.CSL.2022.2

Category Invited Talk

1 Introduction

The analysis of security- and privacy vulnerabilities continues to be a challenging and important problem. Most practitioners acknowledge that online activity will never be absolutely safe, and so rigorous scrutiny on the basis of models and proof has a significant role to play in explaining and evaluating the severity of the threats that remain. But understanding *risk* is generally a fraught process: not only must it contend with opinions about what constitutes risky behaviour, but understanding must necessarily accommodate "invisible" extrinisic influences. For example nothing about the four digit integer 6174 suggests that it would be risky to send it in an email unless we find out that it is someone's PIN and could therefore put their life's savings in jeopardy. [1]

Assuming that evaluating the impact of risk in a given scenario is something that is still a useful thing to do, the following questions arise. Can we provide quantitative measurements that provide some sense of severity of a discovered vulnerability? How can we elaborate qualitative explanations for any numerical measurements of vulnerability that we might compute? How can we advance and share our knowledge of security and privacy defences when those explanations can be highly technical and abstract?

[1] Actually 6174 turns out to have a very curious intrinsic property of being invariant under Kaprekar's operation.

In this paper we describe our experience of addressing those questions using geometrical visualisations of the Quantitative Information Flow (QIF) framework [1]. Our development of this style of "geometrical reasoning" came about through a collaboration between teams who were working on the similar problems in quantification of security risks, using (as it turned out) the same basic mathematical principles, but with very different notations and somewhat differing objectives. We discovered that using visualisations helped enormously in sharing mathematical ideas and building intuition amongst the team members. The visualisations helped confirm facts that we knew in various forms, and enabled us not only to refute conjectures we thought might be true, but also to suggest to us conjectures that we had not discovered for ourselves when we were using purely formal notations. Subsequent rigorous proof was then able to remove the geometrical hunch, providing new theorems in privacy (described below) [8].

Thus a completely unexpected benefit of our "geometry of risk" was its facilitation of "notation-free" communication of fundamental but complex ideas providing summaries of proof steps in the form of geometrical constructions. Equipped with a sense of certainty endowed by the geometry allowed us to formulate and prove formally new theorems thus advancing our understanding of QIF and how it could be used in security and privacy. In this we find we are in agreement with Henri Poincaré's insights on the benefit of pictorial representations to promote intuition and communication of complex mathematical ideas between scientitsts:

> "I have already had occasion to insist on the place intuition should hold in the teaching of the mathematical sciences. Without it young minds could not make a beginning in the understanding of mathematics; they could not learn to love it and would see in it only a vain logomachy; above all, without intuition they would never become capable of applying mathematics. But now I wish before all to speak of the role of intuition in science itself. If it is useful to the student, it is still more so to the creative scientist."
>
> Extract from *Intuition and Logic in mathematics* appearing in *La valeur de la science*, Henri Poincaré, 1905.

1.1 Related work

There are many examples in mathematics of visualisations used to provide explanation and insight for formal reasoning. The earliest instance is of course Euclid's elements for reasoning about spacial relationships; Oliver Byrne's 1847 treatise [3] is a masterful account of how diagrams can be used optimally to convey complex geometrical ideas. Roger Nelsen's *Proofs without words* [11] promotes the use of visualisations to explain mathematical ideas in a range of topics from algebra to calculus, and theorems about sequences and series. And one of the shortest papers ever written consisted essentially of two figures, as the explanation of a mathematical result [6].

The geometry underlying the study of Quantitative Information Flow first appeared in Alvim [1]; this also includes Morgan's "overlapping triangles" demonstration that the at-least-as-secure-as partial order on information-flow channels is not (alas) a lattice as well. Fernandes' application of geometrical ideas to study universally-optimal utility mechanisms for differential privacy appears in [8].

In this paper therefore we emphasise the role played by geometrical ideas in supporting the communication via "proofs without words" for quantifying security and privacy risks using quantitative information flow.

2 Quantitative Information Flow Basics

The informal idea of a secret is that it is something about which there is some uncertainty, and the greater the uncertainty the more difficult it is to discover exactly what the secret is. For example, one's 4-digit PIN should be kept secret, but if the last two digits are discovered to be 7 and 4, then it becomes much easier to guess the rest of it. That is, when *any* information about a secret becomes available to an observer (often referred to as an adversary) the uncertainty is reduced, allowing the property, or even exact value of the secret, to be more accurately inferred. When that happens, we say that information (about the secret) has leaked.

Quantitative Information Flow (QIF) makes the above intuition mathematically precise. Given a range of possible secret values of (finite) type $\mathcal{X}$, we model a secret as a discrete probability distribution of type $\mathbb{D}\mathcal{X}$, because it ascribes "probabilistic uncertainty" to the secret's exact value. Given some distribution $\pi\colon \mathbb{D}\mathcal{X}$, we write π_x for the probability that π assigns to $x\colon \mathcal{X}$, with the idea that the more likely it is that the real value is some specific x, then the closer π_x will be to 1. Usually the uniform distribution over $\mathcal{X}$ models a secret which could equally well take any one of the possible values drawn from its type and we might say that, beyond the existence of the secret, nothing else is known. There could, of course, be many reasons for using some other distribution: for example if the secret were the height of an individual then a normal distribution might be more realistic. In any case, once we have a secret, we are interested in analysing whether an algorithm, or protocol, that uses it might leak some information about it. To do that we define a measure for uncertainty, and use it to compare the uncertainty of the secret before and after executing the algorithm. If we find that the two measurements are different then we can say that there has been an information leak.

The idea of measuring security risk in terms of quantitative information flow (and that name) was pioneered by Clarke et al. [4]. That and other early QIF analyses of information leaks in computer systems [5, 4] used Shannon entropy [12] to measure uncertainty because it captures the idea that more uncertainty implies "more secrecy" – and indeed the uniform distribution corresponds to maximum Shannon entropy (corresponding to maximum "Shannon uncertainty"). More recent treatments have shown however that Shannon entropy is not *always* the best way to measure uncertainty in security contexts: precisely because of its beautiful generality, it might not model scenarios relevant to the goals of a particular adversary. Indeed there are some circumstances where a Shannon analysis actually gives a more favourable assessment of security than is actually warranted, when the adversary's motivation is taken into account [13].

Alvim et al. [2] proposed a notion of uncertainty, more general than Shannon, based on "gain functions". In this paper we will use the equivalent formulation of loss functions. [2] A *loss function* measures a secret's uncertainty according to how it affects an adversary's actions within his context. We write $\mathcal{W}$ for a (usually finite) set of actions available to an adversary corresponding to an "attack scenario" where the adversary tries to infer something (e.g. some property, but perhaps its actual value) about the secret. For a given secret $x\colon \mathcal{X}$, an adversary's choice of action $w\colon \mathcal{W}$ results in the adversary's losing something beneficial to his objective. That loss can vary depending on the adversary's action (w) and the exact value of the secret (x). The more effective is the adversary's choice in how to act, the more he is able to overcome any uncertainty concerning the secret's value.

[2] Shannon Entropy is a special case: it can be defined using a loss function.

▶ **Definition 1.** *Given a type $\mathcal{X}$ of secrets, a (real valued) loss function $\ell\colon \mathcal{W}\times\mathcal{X}\to\mathbb{R}$ is such that $\ell(w,x)$ determines the loss to an adversary if he chooses w and the secret is x.*

A simple example of a loss function is br (for "Bayes Risk", as explained below), where $\mathcal{W}:=\mathcal{X}$ so that the available actions are simply "guess the value of x", and thus [3]

$$\mathsf{br}(w,x) \;:=\; 0 \;\; \textit{if}\;\; w{=}x \;\; \textit{else}\;\; 1\ . \tag{1}$$

For this scenario, the adversary's goal is to determine the exact value of the secret, so he loses nothing if he correctly guesses the value of a secret (a good outcome for him), and otherwise he loses \$1 (bad).

Elsewhere the great utility and expressivity of loss functions for measuring various attack scenarios relevant to security –far beyond just guessing the secret's value – have been thoroughly explored [1]. Given a loss function we define the *uncertainty* of a secret in $\mathbb{D}\mathcal{X}$ relative to the scenario the loss function describes: it is the minimum *average* loss to an adversary. More explicitly, for each action w, the adversary's average loss relative to some distribution π of the secret is $\sum_{x\in\mathcal{X}}\ell(w,x)\times\pi_x$; thus his minimum average loss is the action that yields that minimal average.

▶ **Definition 2.** *Let $\ell\colon\mathcal{W}\times\mathcal{X}\to\mathbb{R}$ be a loss function, and $\pi\colon\mathbb{D}\mathcal{X}$ be a secret. The* uncertainty *$U_\ell[\pi]$ of the secret wrt. ℓ is given by*

$$U_\ell[\pi] \;:=\; \min_{w\colon\mathcal{W}}\sum_{x\colon\mathcal{X}}\ell(w,x)\times\pi_x\ .$$

For a secret $\pi\colon\mathbb{D}\mathcal{X}$, the uncertainty wrt. the loss function br is $U_{\mathsf{br}}[\pi]{:=}\,1-\max_{x\colon\mathcal{X}}\pi_x$, that is the deficit of the maximum probability assigned by π to possible values of x. The adversary's best strategy for optimising his loss would therefore be to choose the value x that corresponds to the maximum probability under π. This uncertainty U_{br} is called *Bayes' Risk*.

We now define a *mechanism* to be an abstract model of a protocol or algorithm that uses secrets. As the mechanism executes we assume that there are a number of observables, that is outputs it might produce, that can depend on the actual value of the secret it is processing: we write $\mathcal{Y}$ for the type of those observables. The *model* of the mechanism therefore assigns a probability that $y\colon\mathcal{Y}$ might be observed given that the secret is x. Such observables could be sample timings in a timing analysis in cryptography, for example.

▶ **Definition 3.** *A* mechanism *is a stochastic channel*[4] *$C\colon\mathcal{X}\times\mathcal{Y}\to[0,1]$. The value C_{xy} is the probability that y is observed given that the secret is x.*

Given a (prior) secret $\pi\colon\mathbb{D}\mathcal{X}$ and mechanism C we write $\pi{\triangleright}C$ for the joint distribution *in $\mathbb{D}(\mathcal{X}\times\mathcal{Y})$ defined*

$$(\pi{\triangleright}C)_{xy} \;:=\; \pi_x\times C_{xy}\ .$$

For each $y\colon\mathcal{Y}$, the marginal probability *that y is observed is $p_y := \sum_{x\colon\mathcal{X}}(\pi{\triangleright}C)_{xy}$.* [5]

For each observable y, the corresponding posterior probability *of the secret is the conditional $\pi|^y$ in $\mathbb{D}\mathcal{X}$ defined $(\pi|^y)_x := (\pi{\triangleright}C)_{xy}/p_y$.* [6]

---

[3] We write $:=$ for "is defined to be".

[4] "Stochastic" means that the rows sum to 1, i.e. that $\sum_{y\colon\mathcal{Y}}C_{xy}=1$ for each x.

[5] Equivalently that is $\sum_x\pi_xC_{xy}$.

[6] We assume for convenience that when we write p_y the terms C, π and y are understood from the context. Notation suited for formal calculation would need to incorporate C and π explicitly.

Given a prior secret π and mechanism C, it's clear that the entry $\pi_x \times C_{xy}$ of the joint distribution $\pi \triangleright C$ is the probability that the actual secret value is x and the observation is y. That joint distribution contains two important pieces of information, which we single out: the probability p_y of observing y and the corresponding posterior $\pi|^y$ which represents the adversary's updated view about the uncertainty of the secret's value. If the uncertainty of the posterior increases, then information about the secret has leaked and a rational adversary *might* use it to decrease his loss by changing how he acts, i.e. by altering his choice of action w. The adversary's average overall loss, taking the observations into account, is defined to be the average posterior uncertainty (i.e. the average loss of each posterior distribution, weighted according to their respective marginals):

$$U_\ell[\pi \triangleright C] := \sum_{y \in \mathcal{Y}} p_y \times U_\ell[\pi|^y] \ , \qquad \text{where } p_y \text{ and } \pi|^y \text{ are defined at Def. 3.} \tag{2}$$

The "rationality" of the adversary is inside the U_ℓ, expressed by the $\min_{w:\mathcal{W}}$ there (Def. 2).

2.1 Hyper-distributions summarise the risk

In the above model the key structure used to compute the posterior loss is $[\pi \triangleright C]$ – which can be represented (more abstractly) as a *hyper-distribution*, that is a distribution of type $\mathbb{D}^2\mathcal{X}$ where the outer probability is p^y, the marginal probability of an observation and the inner distribution is the posterior corresponding to that y.

An advantage of that abstraction is that there is then a partial order on $\mathbb{D}^2\mathcal{X}$ which allows the robust comparison of channels wrt. their information flow properties. It is the relation ($\sqsubseteq$), defined

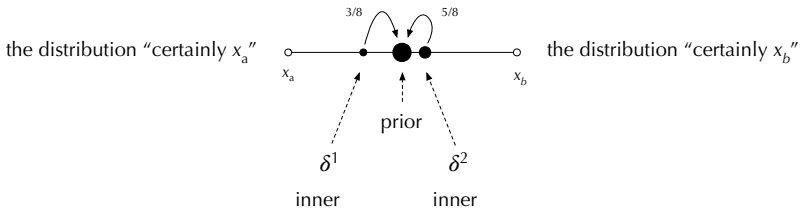
▶ **Definition 4.** *Let C, D be channels. We say that $C \sqsubseteq D$ if for all loss functions ℓ and prior distributions π we have $U_\ell[\pi \triangleright C] \leq U_\ell[\pi \triangleright D]$.*

That is, if D refines C then we can be sure that the adversary always loses no less with D than with C in any scenario that can be defined by some π and ℓ of the correct type: that is, for a defender D is at least as secure as C. As we noted above, Shannon Entropy is a special case of an ℓ-uncertainty – but with infinitely many actions– and we call its loss function se, so that Shannon entropy is U_{se}.

2.2 Reasoning geometrically

The basic model of QIF – set out mathematically above – turns out to have an appealing geometrical interpretation, one that we can use to visualise the relationships between channels and also loss functions. The first step is to visualise hyper-distributions using a *barycentric* representation of $\mathbb{D}\mathcal{X}$ [1][chapter 12]. Recall from §2.1 above that a hyper-distribution is of type $\mathbb{D}^2\mathcal{X}$, equivalently $\mathbb{D}(\mathbb{D}\mathcal{X})$ or a "distribution of distributions". The "inner $\mathbb{D}$" defines what we call "inners", distribution on $\mathcal{X}$ directly, that correspond to posteriors associated with the observations – and when $\mathcal{X}$ is finite the inners are (non-negative) 1-summing vectors in $\mathbb{R}^{|\mathcal{X}|}$. Thus a hyper-distribution corresponds to a convex sum of 1-summing vectors. The "outer $\mathbb{D}$" is the "outer" that gives the marginal probabilities associated with each inner, the weights in that convex sum.

Suppose first that $\mathcal{X} := \{x_a, x_b\}$ and that $\mathsf{u}: \mathbb{D}\mathcal{X}$ is the uniform prior. As explained above u corresponds to a 1-summing vector, so that as expected in this case $\mathsf{u} := (1/2, 1/2)$. Here the first component is the probability that the secret is x_a and the second component is the



The barycentric representation of $\mathbb{D}\{x_a, x_b\}$ is the horizontal line between x_a and x_b (corresponding to the 1-summing vectors $(1,0)$ and $(0,1)$). The two inners are labelled δ^1 and δ^2 with their positions relative to the $x_a(x_b)$ labels corresponding to the probabilities $\delta^1_{x_a}(\delta^1_{x_b})$ and $\delta^2_{x_a}(\delta^2_{x_b})$. The outers are indicated as the relative distance between the δ's and their weighted average to obtain the original prior u.

Figure 1 Barycentric representation of $[\mathsf{u} \triangleright C]$.

probability that the secret is x_b – since u is uniform, those probabilities are the same. Next let the observations $\mathcal{Y}{:=}\{y_1, y_2\}$ be and consider the channel $C \in \mathcal{X}{\times}\mathcal{Y}$, and its corresponding hyper-distribution $[\mathsf{u}{\triangleright}C]$ when the prior is u.

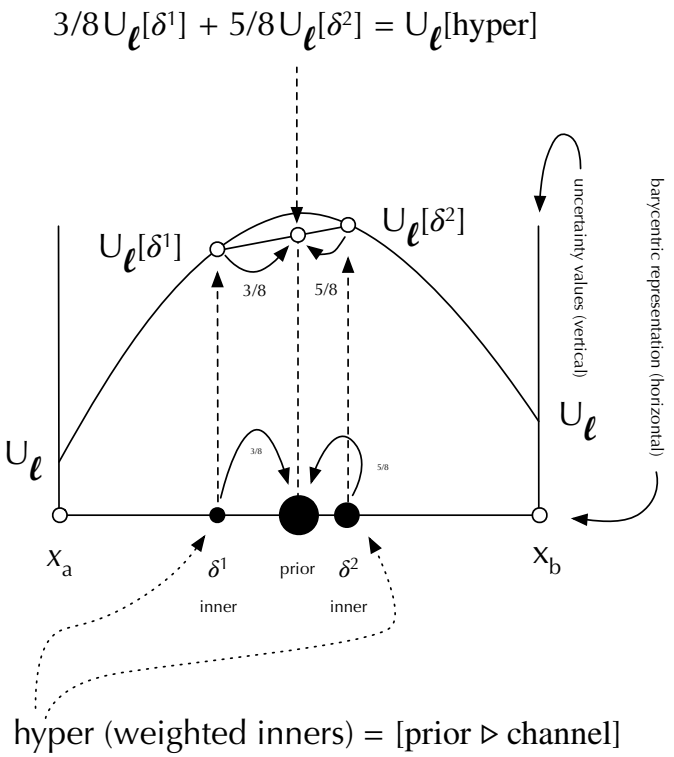
$$C \;:=\; \begin{array}{c} \\ x_a \\ x_b \end{array} \begin{array}{c} \begin{array}{cc} y_1 & y_2 \end{array} \\ \left(\begin{array}{cc} 1/2 & 1/2 \\ 1/4 & 3/4 \end{array}\right) \end{array} \qquad [\mathsf{u}{\triangleright}C] \;:=\; \begin{array}{c|cc} & 3/8 & 5/8 \\ \hline x_a & 2/3 & 2/5 \\ x_b & 1/3 & 3/5 \end{array}$$

As we can see, the inners (i.e. posteriors) are $\delta^1{:=}(2/3, 1/3)$ and $\delta^2{:=}(2/5, 3/5)$, and their corresponding "outers" the marginal probabilities $3/8, 5/8$.

Since $\mathcal{X}$ has only two elements, the barycentric representation of $\mathbb{D}\mathcal{X}$ is one-dimensional, running from distribution "certainly x_a" at left to "certainly x_b" at right, as in Fig. 1. A point on that line represents a linear combination of those two extremes, and the larger the probability the distribution assigns to x_a, say, the closer its representing point is to the left-hand side. This barycentric representation therefore locates inners $\delta^{1,2}$ on that horizontal line, with for example δ^1 lying closer to the left-hand side because it assigns greater probability to x_a.

Fig. 1 also shows the prior (as a point in the middle of the line, because it's uniform), and the points representing $\delta^{1,2}$ are given sizes corresponding to the *outers* associated with them. If linearly combined with those sizes as coefficients, they will give the prior at the position shown (and with "size" 1). That is a property of the construction $[\pi{\triangleright}C]$ for any prior π (uniform or not) and channel C.

Next, given a loss function ℓ we can plot its associated uncertainty $U_\ell(\pi)$ on the vertical axis above the barycentric representation, and in fact (for all ℓ) it will determine a concave and continuous curve there. Fig. 2 also shows an uncertainty U_ℓ as a function from inners to reals. Here a particular loss $\ell(w, \cdot)$ becomes a tangent to the curve U_ℓ, and therefore the curve shown is the envelope of all those w-determined tangents. Locating δ^1, δ^2 on the (horizontal) barycentric axis, we can easily read off $U_\ell[\delta^1]$ and $U_\ell[\delta^2]$ as the height of the curve U_ℓ above them.



The barycentric axis is the horizontal labelled x_a, x_b, and the inners (posteriors) of $[\mathsf{u}\triangleright C]$ are indicated as δ^1, δ^2. The circles' sizes represent their respective outer (i.e. marginal) probabilities. The curve denotes U_ℓ as a function of points (distributions) on the horizontal barycentric axis. Thus $U_\ell[\delta^1])$ is the height of the vertical line from δ^1 up to the curve U_ℓ.

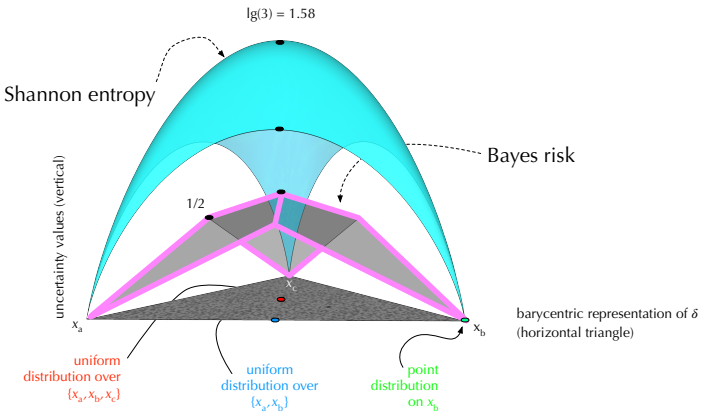
The average of the inners δ^1 and δ^2 weighted according to their respective outer probabilities in $[\mathsf{u}\triangleright C]$ yields the prior u again; but the (same) average of the expected losses is $U_\ell[\mathsf{u}\triangleright C]$, whose value is the height of the vertical line from the prior u (in this case) to the line joining the two points on the U_ℓ curve.

Figure 2 Barycentric representation of uncertainties and loss functions.

With all that done, it is easy to compute $U_\ell[\mathsf{u}\triangleright C]$ as that same weighted average of the heights of U_ℓ above them: and that is done geometrically by connecting those two points on $U_\ell[\mathsf{u}\triangleright C]$ with a straight line, and noting its height *above the prior*. That is, we simply take the weighted average $3/8\times U_\ell[\delta^1] + 5/8\times U_\ell[\delta^2]$, which is depicted on the figure as the point at which the vertical line from u meets the line joining the uncertainties at $U_\ell[\delta^{1,2}]$.

Because U_ℓ is concave, we now have a "proof without words" [11] of the well known Jensen's inequality. [7] [8]

[7] For more than two inners, the same "proof" generalises; but the combinations have to be done two-by-two.
[8] As has been noted by many authors, a "proof without words" is not a *proof*. We use it here to mean that it is a suggestion for a proof strategy – of course the actual proof must be demonstrated with appropriate words.



The barycentric representation of distributions over three values $\{x_a, x_b, x_c\}$ is a (two-dimensional) triangle, rather than a straight line as it was in Fig. 2 above when $\mathcal{X}$ had just two values: it's always one dimension smaller than the size of $\mathcal{X}$, because the constraint that the probabilities sum to 1 loses one degree of freedom.

Figure 3 Barycentric representations on three points.

In the following sections we use pictures and constructions given above as a way to build and share intuition about the behaviour of channels. Our demonstrations can be thought of as providing hints to explain a complex argument; here we suppress the accompanying formal arguments leaving the constructions as "proofs with very few words".

3 Some QIF proofs without *many* words

3.1 Refinement seen geometrically

In Def. 4 the *refinement* partial order is defined between hyper-distributions, that the loss with respect to *any* uncertainty must increase. But the *Coriaceous Theorem* [10] gives an equivalent geometric definition. One hyper – an (outer) distribution over (its inner) distributions – is a refinement of another just when the more refined's inners can be realised as a weighted merge of the less refined's inners. In Fig. 2 for example, the hyper represented by the two smaller dots can be refined to another – to many others – by "carving off" pieces of the inners and merging them according to their respective weights. [9] And the proof of the Coriaceous Theorem is *itself* inspired geometrically, because it relies on the Separating-Hyperplane Lemma, where the convex region represents the more refined hyper, and the separating plane's normal gives the coefficients of the loss function that satisfies the original Def. 4.

The geometric view tells us immediately that a more-refined hyper's inners must lie (non-strictly) within the convex hull of the less-refined hyper, and helps us (in §3.2) to see – again geometrically – whether the refinement partial order is a lattice. The following stronger fact (requiring a full proof) allows us to make stronger geometric arguments [1].

[9] In the extreme case, they can be refined to the singleton hyper whose sole inner is the original prior, as illustrated by the arrows in that figure.

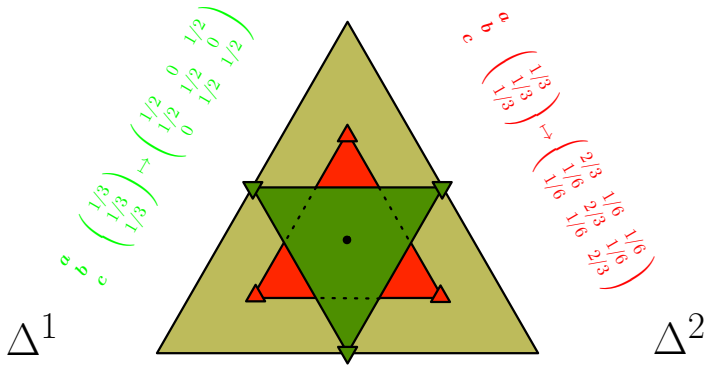


Figure 4 Two triangular hypers $\Delta^{1,2}$ on $\mathcal{X} = \{x_a, x_b, x_c\}$.

▶ **Lemma 5.** *Let the (finite) state space $\mathcal{X}$ have N elements, and let some hyper Δ have N linearly independent inners. Then any other hyper Δ' all of whose inners lie within the convex hull of Δ (and that is derived from the same prior) is a refinement of Δ, no matter how many inners Δ' might have.*

The original "qualitative" *Lattice of Information* [9] is (by its very title) a lattice; but in the quantitative case (here), it turns out that it is not a lattice.

3.2 The refinement order is not a lattice

In Fig. 3 we show how our barycentric representation of distributions appears when $\mathcal{X}$ is $\{x_a, x_b, x_c\}$, no longer a line but now an equilateral triangle, shown in grey at the bottom. We will reason about hypers in that triangle.

Fig. 4 shows two hypers $\Delta^{1,2}$ (green and red resp.) over a state space $\mathcal{X} = \{x_a, x_b, x_c\}$. They are both generated from the uniform prior $(1/3, 1/3, 1/3)$, shown as a black dot at the centre of the barycentric triangle, and each of the hypers is a (smaller) equilateral triangle itself, having three inners (each) all three with outer probability $1/3$. All refinements of Δ^1 must lie within the green triangle; and all refinements of Δ^2 must lie within the red triangle; and so all refinements of *both* must lie within the yellow hexagon.

And so from Lem. 5 we know that *any* hyper (with the same prior) in the yellow hexagon must refine both $\Delta^{1,2}$, because they have only three inners (each).

Now consider three more hypers, each with only two inners (each with outer $1/2$), named $\Delta^{3,4,5}$ as shown in Fig. 5. Each of $\Delta^{3,4,5}$ is a refinement of each of $\Delta^{1,2}$, as noted just above, which fact we write compactly as $\Delta^{1,2} \sqsubseteq \Delta^{3,4,5}$. We show (after the following lemma) that there is no hyper Δ satisfying

$$\Delta^{1,2} \quad \sqsubseteq \quad \Delta \quad \sqsubseteq \quad \Delta^{3,4,5} \quad . \tag{3}$$

▶ **Lemma 6.** *If hypers Δ', Δ'' (with the same prior) lie (non-strictly) within the yellow hexagon of Fig. 5, and $\Delta' \sqsubseteq \Delta''$, then the outer probability of any of the (six) hexagon vertices in Δ'' cannot exceed the outer probability of that same vertex in Δ'.* [10]

That is because refinement interpolates inners, and interpolation of any *inners of Δ' (which, remember, are non-strictly within the hexagon) cannot increase the outer of one of the hexagon's vertices, because the hexagon is convex and its vertices are its extreme points.*

[10] If some vertex of the hexagon is not "actually" one of the inners of the hyper considered, we just consider its outer to be zero.

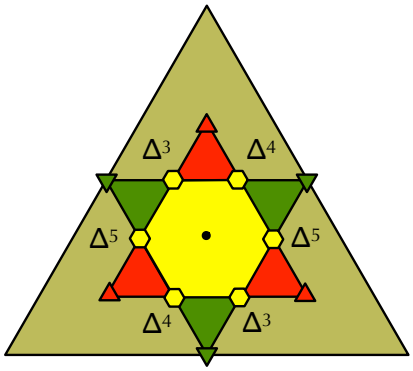


Figure 5 Three more two-inner hypers $\Delta^{3,4,5}$ on $\mathcal{X} = \{a, b, c\}$.

Now from Lem. 6 we have immediately that there is *no* Δ satisfying (3), because any Δ that refines both $\Delta^{1,2}$ must lie within the hexagon; and by Lem. 6 if Δ additionally is refined by all of $\Delta_{3,4,5}$ then the outer of Δ at all six vertices be at least $1/2$, impossible because those outers must sum to 1.

And so $(\sqsubseteq)$ on $\mathcal{X} = \{x_a, x_b, x_c\}$ is not a lattice: for both the join $\Delta^1 \sqcup \Delta^2$ and the meet $\Delta^3 \sqcap \Delta^4 \sqcap \Delta^5$, if they existed, would as Δ satisfy (3) – which Lem. 6 showed was impossible. And so neither exists.

3.3 Channel composition is not idempotent. . . unless it is deterministic

Channel composition (or parallel) composition is defined to be the channel obtained by independent executions of two channels, taking both their observations into account.

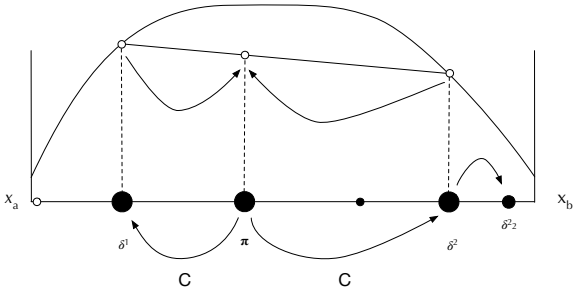
▶ **Definition 7.** *Let* $C{:}\mathcal{X}{\times}\mathcal{Y} \to [0,1]$, $D{:}\mathcal{X}{\times}\mathcal{Z} \to [0,1]$. *The parallel composition* $C\|D{:}\mathcal{X}{\times}(\mathcal{Y}{\times}\mathcal{Z}) \to [0,1]$ *of* C, D *is defined as the product space of observations which are now drawn from* $\mathcal{Y}{\times}\mathcal{Z}$*:*

$$(C\|D)_{x(yz)} := C_{xy} \times D_{xz} \ .$$

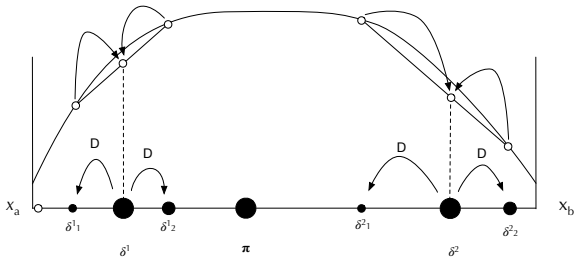
Landauer [9] showed that parallel composition of deterministic channels are idempotent; this suggests the question of whether parallel composition more generally is *idempotent*. Can it be the case that $C\|C = C$ when C is not deterministic?

We show that $C\|C \neq C$ for properly probabilistic channels by the geometric constructions shown in Fig. 2 and Fig. 6. First Fig. 2 shows that whatever the prior π, when a properly probabilistic channel with two observations is applied to it, there will be two inners, averaging to the original channel. But now Fig. 6 shows how to apply that construction twice: first C is applied tp the original prior, yielding two inners $\delta^{1,2}$ and then, because of the independence, D is now applied to each of $\delta^{1,2}$, where these new corresponding inners average to the δ's.

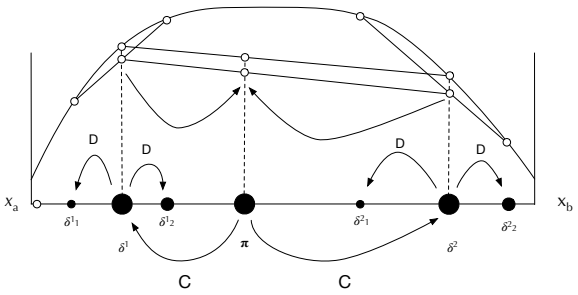
Using now the uncertainty U_ℓ we can visualise how the information flow must be different between C and $C \parallel D$: we apply the construction in Fig. 2 twice, each time averaging: first to compute $U_\ell[\delta^{1,2} \triangleright D]$ must be strictly less than each of $U_\ell[\delta^{1,2}]$; and then again to show that $U_\ell[\pi \triangleright (C\|D)]$. Our question is now answered by setting D to C.



Channel C makes inners $\delta^{1,2}$ from prior $\boldsymbol{\pi}$



Those inners $\delta^{1,2}$ become priors for D



The expected loss for $[\pi \triangleright C \| D]$ is determined by applying the construction from Fig. 2 twice. First $[\pi \triangleright C]$ is represented by its two inners $\delta^{1,2}$ which average to π; then the same construction is applied to $[\delta^1 \triangleright D]$ and $[\delta^2 \triangleright D]$ as shown. The average $U_\ell[\delta^1 \triangleright D]$ and $U_\ell[\delta^2 \triangleright D]$ are then weighted averages as described in Fig. 2. The expected loss $U_\ell[\pi \triangleright C \| D]$ is then computed by averaging those averages.

Figure 6 Construction for $C \| D$ in 2 dimensions using Shannon Entropy.

We note finally that this geometric construction assumes that the inners in $[\pi \triangleright C]$ are spread either side of π. This does not happen when C is deterministic i.e. does not have any values except 1's and 0's. For deterministic channels, the secrets are separated into equivalence classes and the information flow yields exactly which class a secret is in. Geometrically this means that when the support of an inner δ lies entirely within an equivalent class the inners $[\delta \triangleright C]$ are actually δ.

3.4 Which is better, the Laplace- or Geometric mechanism for implementing differential privacy?

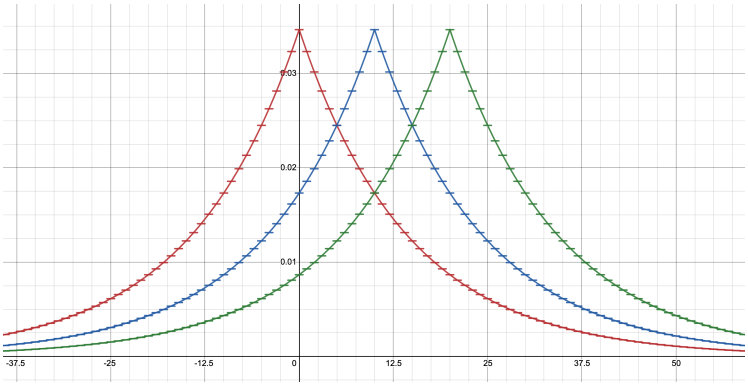


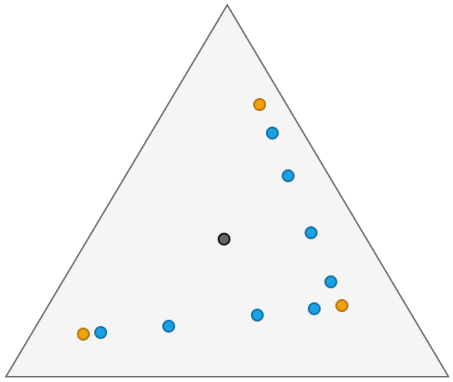
Figure 7 The Laplace (continuous pdf) and Geometric (discrete lines) noise mechanisms.

Differential privacy [7] is a technique for for providing individuals' data some measure of privacy when that data is shared through e.g. a query (to the database containing it). The idea is that rather than reporting the result of a raw query, instead some random noise (chosen according to a parameter ϵ) is added to the result and the noisy answer is then reported. Different methods of adding random noise have different properties of course – and those that are in keeping with the spirit of differential privacy can guarantee to make similar query results "indistinguishable in output" so that in practice an observer cannot tell apart the outputs of inputs that are already similar (in the raw), even when those raw results are distinguishable enough to risk a privacy breach.

Two popular methods of randomisation are based on the Geometric and Laplace probability distributions leading to the definition of the corresponding Geometric and Laplace mechanisms. Given the output of a query is some number d (consisting of e.g. the count of data entries satisfying a condition, or some average value) instead of outputting the raw d, the Geometric mechanism would output $d+c$ where c is distributed according to a geometric distribution; similarly the Laplace mechanism would output $d+e$ where e is distributed according to the Laplace distribution. The two different methods of randomisation are depicted in Fig. 7.

Interestingly, although they have broadly similar shapes (Fig. 7) albeit the Laplace gives a continuous"probability density function" and the Geometric a discrete number of outputs, there is no obvious way to compare the properties of these mechanisms in terms of how their privacy properties work. Perhaps they leak the same, or entirely different amounts of information when used as randomisers. It turns out that when viewed in terms of QIF channels we find that, for the same ϵ parameter we can say definitively that the Geometric mechanism leaks more information than does the Laplace mechanism, and thus the Laplace mechanism is strictly more private.

Our proof with very few words in illustrated in Fig. 8 which depicts the barycentric visualisation of the hyper-distributions corresponding to the Geometric- and Laplace mechanisms when applied to three secret values (i.e. potential raw query results) and represented as QIF channels. As explained above, the inners of the corresponding hyper-distributions can be located as points on the plane in three dimensions. Curiously we see that the Geometric mechanism produces three inners (orange dots in Fig. 8) which are linearly independent because they do not lie on a line. Even more curiously the inners from the Laplace mechanism (blue dots in Fig. 8) lie within the convex hull of the Geometric's inners, and therefore by Lem. 5 the Laplace perforce refines the Geometric, which from Def. 4 means that the Geometric mechanism always leaks more information about the secret than does the Laplace mechanism. This observation, discovered purely by this visualisation led to the fully formal proof of universal optimality of the Laplace mechanism for continuous inputs [8].



The Geometric mechanism on 3 secrets parametrised by ϵ consists of 3 linearly independent posteriors (orange). The Laplace mechanism parametrised by the same ϵ (and on the same domain) consists of the blue posteriors sitting on the edges – ie. in the convex hull – defined by the Geometric posteriors. It follows that the Laplace mechanism is a refinement of the Geometric mechanism.

Figure 8 Construction of Geometric (orange) and Laplace (blue) hypers in 3 dimensions.

4 Conclusions

In this paper we have demonstrated how to use geometrical ideas to explain complex ideas within the framework of quantitative information flow. Although they do not represent full formal proofs of these results they have proved to be useful for sharing ideas between different groups of collaborators and therefore in developing the field.

References

1 M. S. Alvim, K. Chatzikokolakis, A.K. McIver, C.C. Morgan, C. Palamidessi, and G. Smith. *The Science of Quantitative Information Flow*. Information Security and Cryptography. Springer International Publishing, 2020.

2 Mário S. Alvim, Kostas Chatzikokolakis, Catuscia Palamidessi, and Geoffrey Smith. Measuring information leakage using generalized gain functions. In *Proc. 25th IEEE Computer Security Foundations Symposium (CSF 2012)*, pages 265–279, June 2012.

3 Oliver Byrne, Bruce Rogers, and Euclid. *The first six books of the elements of Euclid: in which coloured diagrams and symbols are used instead of letters for the greater ease of learners / by Oliver Byrne*. William Pickering London, 1847.

4 David Clark, Sebastian Hunt, and Pasquale Malacaria. Quantitative analysis of the leakage of confidential data. *Electr. Notes Theor. Comput. Sci.*, 59(3):238–251, 2001.

5 Michael R. Clarkson, Andrew C. Myers, and Fred B. Schneider. Belief in information flow. In *18th IEEE Computer Security Foundations Workshop, (CSFW-18 2005), 20-22 June 2005, Aix-en-Provence, France*, pages 31–45, 2005.

6 J. Conway and A. Soifer. Can $n^2 + 1$ unit equilateral trianges cover an equilateral triangle of side $> n$, say $n + \epsilon$? *The American Mathematical Monthly*, 18(143), 2005.

7 Cynthia Dwork. Differential privacy. In *Proc. 33rd International Colloquium on Automata, Languages, and Programming (ICALP 2006)*, pages 1–12, 2006.

8 Natasha Fernandes, Annabelle McIver, and Carroll Morgan. The laplace mechanism has optimal utility for differential privacy over continuous queries. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*, pages 1–12. IEEE, 2021.

9 Jaisook Landauer and Timothy Redmond. A lattice of information. In *Proc. 6th IEEE Computer Security Foundations Workshop (CSFW'93)*, pages 65–70, June 1993.

10 Annabelle McIver, Carroll Morgan, Geoffrey Smith, Barbara Espinoza, and Larissa Meinicke. Abstract channels and their robust information-leakage ordering. In Martín Abadi and Steve Kremer, editors, *Principles of Security and Trust - Third International Conference, POST 2014, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2014, Grenoble, France, April 5-13, 2014, Proceedings*, volume 8414 of *Lecture Notes in Computer Science*, pages 83–102. Springer, 2014. doi:10.1007/978-3-642-54792-8_5.

11 Roger Nelsen. *Proofs Without Words: Exercises in Visual Thinking*, volume 1. MAA Press, 1993.

12 C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

13 Geoffrey Smith. On the foundations of quantitative information flow. In Luca de Alfaro, editor, *Proc. 12th International Conference on Foundations of Software Science and Computational Structures (FoSSaCS '09)*, volume 5504 of *Lecture Notes in Computer Science*, pages 288–302, 2009.