To appear in IEEE Proc LiCS 2012

A Kantorovich-Monadic Powerdomain for Information Hiding, with Probability and Nondeterminism

Annabelle McIver
Dept. Computer Science
Macquarie University
Sydney, Australia
Email: annabelle.mciver@mq.edu.au

Larissa Meinicke
School of Inf. Tech. and Elect. Eng.
University of Queensland
Brisbane, Australia
Email: l.meinicke@uq.edu.au

Carroll Morgan
School of Comp. Sci. and Eng.
Univ. New South Wales
Sydney, Australia
Email: carrollm@cse.unsw.edu.au

***Abstract*—We propose a novel domain-theoretic model for nondeterminism, probability and hidden state, with relations on it that compare information flow. One relation is Smyth-like, based on a structural, refinement-like order between semantic elements; the other is a testing order that generalises several extant entropy-based techniques. Our principal theorem is that the two orders are equivalent.**

The model is based on the Giry/Kantorovich monads, and it abstracts *Partially Observable Markov Decision Processes* by discarding observables' actual values but retaining the effect they had on an observer's knowledge.

We illustrate the model, and its orders, on some small examples, where we find that our formalism provides the apparatus for comparing systems in terms of the information they leak.

***Index Terms*—Semantics, probabilistic domains, refinement orders, probabilistic monads, quantitative information flow.**

I. Introduction

The probabilistic power domain [12] sets out a general construction for promoting computational behaviour to a probability space. As well as our own earlier work [11], [25], [17], [18], generalised by Tix, Kiemel and Plotkin [31], also Mislove [24] and Varacca and Winskel [35] study denotational models for probabilistic and nondeterministic behaviour. Whilst these treatments provide a domain theoretic presentation for the well-known Markov Decision Process model, much less is known about power-domain models combining hidden state, nondeterminism and probability.

Here we propose a novel power domain construction based on a Giry/Kantorovich structure, and motivated by Partially Observable Markov Decision Processes (*POMDP*'s), together with two information-leakage orders ultimately to provide denotations for programs containing all three features. Our principal theorem is the equivalence of those two orders: that a "structural" refinement is equivalent to a form of "disorder testing" that generalises several other entropy-based tests including Shannon Entropy, Bayes Risk and Guessing Entropy.

In earlier work [19] we considered a specific language of probabilistic sequential programs with hidden state and probability, *but no nondeterminism*. There we presented a-posteriori probability distributions, of information flowing from hidden to observable behaviour, as "hyper-distributions," i.e. observable distributions over hidden distributions over secure program state, and the space of "hypers" was shown to have a partial order equivalent to a testing order based on Bayes Risk [30], [5].

Later, we generalised that and related it to Hidden Markov Models with its refinement order consistent with preservation of "information flow" [21].

The innovation here is to add nondeterminism via *sets* of hyperdistributions, and to identify the mathematical properties that support the definition of a complete partial order corresponding to disorder refinement in this richer context of hidden state, probability *and nondeterminism*.

Our novel construction is based on the Kantorovich monad, of subprobability measures over the category of 1-bounded compact metric spaces and non-expansive functions [33]. The move to measures simplifies the overall presentation as well as giving access to concepts of compactness and closure provided by the (Kantorovich) metric; that leads to hyper-*measures* for modelling information flows, and to sets of them when we consider nondeterminism as well.

We use "refinement" below in the sense of "releasing no more information than." Our main theorem will be to establish a duality between "positive" judgements that one system A is refined by another system B, based on a witness that it is so, and "negative" judgements that A is not refined by B, again based on a witness. [1]

One application of our model is to security, where judgements are popularly made in terms of entropy. An example is Bayes Risk, the conditional probability that a worst-case attacker will not successfully guess the hidden state: a higher risk implies less information release. [2] However this can be

[1] Compare bisimulation, a positive judgement based on a specific relation, and its duality with testing equivalence, where a negative judgement can be supported by a single test.

[2] The "risk" here is the probability that information will be lost when the leak is viewed as a channel, which is why the terminology is reversed [30].

Let P be the set $\{0,1,2\}$ of possible positions for the prize.

$p{:\in}\, P$	$\leftarrow$	*Host puts prize behind curtain.*
$c{:\in}\, P$	$\leftarrow$	*Contestant guesses which one.*
$h{:\in}\, P{-}\{c,p\}$	$\leftarrow$	*Host opens one of the curtains.*

How much does the contestant now know about p?

Curtains p, c, h are chosen, by the host, the contestant, and again the host respectively. The contestant's decision of whether or not to switch will depend on how much he has learned by the host's "leaking" information via his choice h.

Fig. 1. The Monty Hall Puzzle as a information-hiding problem

generalised: our disorder test is a certain kind of continuous function over hidden distributions, and we show that Shannon Entropy, Bayes Risk and Guessing Entropy are all specialisations of it (§IV-C).

II. The Monty Hall puzzle

We begin with an infamous example of a *POMDP*, which we use here to illustrate our approach:

· A prize is hidden behind a curtain $p{\in}\{0,1,2\}$; a contestant C wants the prize, and guesses it's behind curtain c. The host H then opens a curtain h that is neither p nor c.

· Seeing that $h{\neq}p$, but knowing no more about p than that, the contestant C is allowed to change his guess c to improve his chances.

· Should C indeed change his guess c?

This puzzle exposes the difficulties of Bayesian reasoning; and indeed the challenge is to build a coherent model that makes the correct answer clear. The puzzle is small, but not obvious, and it has the features of our proposed semantic space: hidden state, probability, nondeterminism and Bayesian inference. It is set out informally in Fig. 1.

We vary the usual scenario by assuming the placement of p is not uniformly random: rather it is curtain $0, 1, 2$ with probabilities 0.1, 0.3 and 0.6 resp., a distribution we call δ_P. But we do assume that H's subsequent choice $h{:\in}\, P{-}\{c,p\}$ is uniform. Both of these policies are known to C (he has watched the TV show many times); but on any occasion of course he does not know the actual p until the game is over.

We now consider three possible strategies for C's initial guess of c. For clarity we refer to C's subsequent decision of whether to change his guess as his *policy*.

— C guesses $c{:=}\,0$ Knowing δ_P, we reason that C guessed correctly wprob. 0.1, and so p is wprob. 0.9 in $\{1,2\}$, but C also knows $p{\neq}h$. And so indeed he should switch to $\{1,2\}{-}\{h\}$. But what does C actually know about p more precisely?

That depends on h, whose outcome is observed to be distributed between values $\{1,2\}$ with probabilities calculated as follows. The chance that $h{=}1$ is $0.1{\times}\frac{1}{2} + 0.6 = 0.65$ (i.e. the chance $p{=}0$ times the chance it sets h to 1, plus the chance $p{=}2$, where H *must* choose $h{=}1$). And for $h{=}2$ it is $1{-}0.65 = 0.35$. We write this distribution (on h) as $1_{0.65}{\oplus}2$, and analyse next how much C learns from it.

In the first case $h{=}1$, the Bayesian *a posteriori* distribution of p is that $p{=}0$ wprob. $\Pr(p{=}0|h{=}1) = \Pr(h{=}1|p{=}0){\times}\Pr(p{=}0)/\Pr(h{=}1) = 0.5{\times}0.1/0.65 = \frac{1}{13}$, and $1{-}\frac{1}{13} = \frac{12}{13}$ is the probability that $p{=}2$. We write that distribution as $\delta_1{:=}\,0\,_{\frac{1}{13}}{\oplus}2$.

For $h{=}2$ the a posteriori distribution of p is $\delta_2{:=}\,0\,_{\frac{1}{7}}{\oplus}1$.

Thus C's knowledge of p, gained from H, can be summarised as $\Delta_0{:=}\,\delta_1\ {}_{.65}{\oplus}\ \delta_2$. We call Δ_0 a *hyperdistribution* because it is a distribution of distributions (on $\{0,1,2\}$). The *outer* distribution $({}_{0.65}{\oplus})$ of Δ_0 is *resolved*, known by C because he has already seen whether h is 1 or 2. The two "inner" distributions $\delta_{\{1,2\}}$ are *suspended* from C's point of view, quantifying his remaining uncertainty of p, that is δ_1 if he saw $h{=}1$ and δ_2 if he saw $h{=}2$. Whichever uncertainty that turns out to be, his best policy is to change his guess to the most likely value of p associated with it: thus (re-)choose $p{=}2$ for δ_1, and $p{=}1$ for δ_2. Doing so, he will at the end have guessed p correctly with probability $0.65{\times}\frac{12}{13}{+}0.35{\times}\frac{6}{7} = 0.9$, and he has used the policy "always change my guess."

The number 0.9 calculated just above is the "Bayes Vulnerability" of Δ_0, obtained by determining the *maximum a posteriori probability* MAP: given a known but unresolved probability over a hidden value, the best guess is the value that has the highest probability; then one calculates an overall probability of correct guessing by a weighted sum of the MAP's according to the probabilities assigned to them by the outer distribution in Δ_0.[3]

Let $X{=}\{0,1,2\}$ so that $\mathbb{D}X$ is the type of discrete distributions over X and $\mathbb{D}^2X$ is the type of hypers. In general the *Bayes Vulnerability* of a hyper $\Delta{:}\,\mathbb{D}^2X$ is $\sum_{\delta\in\lceil\Delta\rceil}\Delta(\delta){\times}{\sqcup}\delta$ where $\lceil\Delta\rceil$ is the support of Δ, those δ for which $\Delta(\delta){\neq}0$; and ${\sqcup}\delta$ is the largest probability assigned by distribution δ; it is one minus the Bayes Risk [30].

— C guesses $c{:=}\,1$ Here h is 0 or 2, and similar calculations give $\delta_3{:=}\,1_{\frac{1}{5}}{\oplus}2$ and $\delta_4{:=}\,0_{\frac{2}{5}}{\oplus}1$ and $\Delta_1{:=}\,\delta_3\ {}_{.75}{\oplus}\ \delta_4$, with Vulnerability $0.75{\times}\frac{4}{5} + 0.25{\times}\frac{3}{5} = 0.75$. But C's policy is now different: after $c{:=}\,1$ he changes only when $h{=}0$.

— C guesses $c{:=}\,2$ Here $\delta_5{:=}\,1_{\frac{1}{2}}{\oplus}2$ and $\delta_6{:=}\,0_{\frac{1}{4}}{\oplus}2$ and $\Delta_2{:=}\,\delta_5\ {}_{.6}{\oplus}\ \delta_6$ for $h{=}0,1$, with Vulnerability $0.6{\times}\frac{1}{2} + 0.4{\times}\frac{3}{4} = 0.6$ corresponding to a "don't change" policy.

A security analysis of this system might consider C to be the attacker, so that C's initial choice of c is demonic in the sense of trying to learn as much as possible about p, as in the three possibilities 0,1,2 we analysed separately above. Taken all together, this system is then the set $\mathbf{\Delta}{:=}\,\{\Delta_0,\Delta_1,\Delta_2\}$ of hypers, a "power-hyper" of type $\mathbb{PD}^2X$. It models the "internal," hidden probability in p with $\delta_{\{1..6\}}$ and "external," visible probability in h with $\Delta_{\{1,2,3\}}$ and finally the demonic nondeterminism in c by collecting the three Δ's into a single set. With a worst-case view, the Vulnerability of power-hyper $\mathbf{\Delta}$ is the maximum of its three constituent Vulnerabilities, since C's strategy will be to choose c in a way that makes

[3] We use Vulnerability here because it is a more intuitive term for the problem we describe; but we use its $(1{-})$-complement Bayes *Risk* in the sequel. Like other entropies, the Risk increases with increasing disorder.

his subsequent policy decision most likely to succeed: that is the 0.9 that reflects the overall attack strategy for C of "First choose $c{:=}0$, and then always change."

Observe that the demonic nondeterminism sets out an attacker's policy *options*, how he exploits knowledge of which distribution is suspended; but he cannot "look inside" the suspensions to see how they were resolved.

III. Notations and Conventions for the Kantorovich Monads

A. The Kantorovich monads $\mathbb{K}$ and $\underline{\mathbb{K}}$

Monads $\langle \mathbb{K}, \boldsymbol{\eta}, \boldsymbol{\mu}\rangle$ and $\langle \underline{\mathbb{K}}, \boldsymbol{\eta}, \boldsymbol{\mu}\rangle$ of measures and submeasures resp. are defined on the category KMet_1 of 1-bounded, compact metric spaces and non-expansive functions [33, there written $\mathcal{B}, \mathcal{B}'$]; they are extensions of the probability monad Π of Giry [8] on the category Mes of measurable spaces and measurable functions [34], [33].

For A in KMet_1 let α, α' be two submeasures over the Borel algebra induced by its metric. The *Kantorovich distance* between them is given by [4]

$$\boldsymbol{k}_A(\alpha,\alpha') \quad := \quad (\sqcup f{:}\, A \xrightarrow{1} [0,1] \bullet |\textstyle\int f \,\mathrm{d}\alpha - \int f \,\mathrm{d}\alpha'|) \quad (1)$$

where by $\sqcup$ we mean supremum, and by $\xrightarrow{1}$ we mean non-expansive functions. We usually omit the A-subscript, writing just $\boldsymbol{k}$.

Let B be a second metric space in KMet_1. The effect of $\underline{\mathbb{K}}$ (similarly $\mathbb{K}$) on a morphism $f{:}\, A \xrightarrow{1} B$ is then given by $\underline{\mathbb{K}}f(\alpha)(\mathbf{B}) := \alpha(f^{-1}(\mathbf{B}))$, for all α in $\underline{\mathbb{K}}A$ (sim. $\mathbb{K}A$) and measurable set $\mathbf{B}$ in the Borel algebra induced on B. Thus $\underline{\mathbb{K}}f(\alpha)$ is the push-forward measure of α via f [7].

For any $A{:}\,\mathsf{KMet}_1$, the unit $\boldsymbol{\eta}_A{:}\, A \xrightarrow{1} \underline{\mathbb{K}}A$ makes a *point* distribution, so that $\boldsymbol{\eta}_A(a)(\mathbf{A}) := 1$ if $a{\in}\mathbf{A}$ else 0, for each $\mathbf{A}$ in the induced Borel algebra of A.

The multiplication $\boldsymbol{\mu}_A{:}\, \underline{\mathbb{K}}^2 A \xrightarrow{1} \underline{\mathbb{K}}A$ takes a subprobability measure of subprobability measures $\boldsymbol{\beta}$, over the algebra of $B{=}\underline{\mathbb{K}}A$, and *averages* it together to form a single measure over A again: thus we have $\boldsymbol{\mu}_A(\boldsymbol{\beta})(\mathbf{A}) := \int \epsilon_{\mathbf{A}} \,\mathrm{d}\boldsymbol{\beta}$, where *evaluation* function $\epsilon_{\mathbf{A}}$ is defined $\epsilon_{\mathbf{A}}(\alpha){:=}\alpha(\mathbf{A})$.

B. Our use of Kantorovich structures for domains

Throughout we write X for a finite set of states; by $\mathbb{D}X$ we mean the (discrete) probability distributions on X.

As a special case for X we write $\mathbb{D}X$ equivalently for $\mathbb{K}(X, \mathsf{dis})$ where dis is the discrete metric; because X is finite $\delta{\in}\mathbb{D}X$ continues to mean that δ is a discrete distribution on X, and constructions like $\underline{\mathbb{K}}\mathbb{D}X$ are easier to read (than $\underline{\mathbb{K}}\mathbb{K}(X, \mathsf{dis})$ would be). The metric on $\mathbb{D}X$ is topologically equivalent to the Manhattan metric (Lem. 19).

By $\Delta{\in}\underline{\mathbb{K}}\mathbb{D}X$ we therefore mean that Δ is a submeasure on $\mathbb{D}X$ with respect to the Borel algebra of $\mathbb{K}(X, \mathsf{dis})$. [5]

Finally, we annotate structure-preserving functions: thus $(\xrightarrow{1})$ means non-expansive, and $(\xrightarrow{\mathsf{B}})$ means measurable.

[4] This is equivalent to van Breugel's presentation [33] (App. A).

[5] See Fig. 2 in App. D for an illustration of these sets.

IV. The Space of Hypermeasures and its Testing Order

A. The hypermeasure space $\underline{\mathbb{K}}\mathbb{D}X$

Our example of §II motivated a power-hyper structure, i.e. a power set $\mathbb{P}$ of distributions $\mathbb{D}$ of distributions $\mathbb{D}$, thus $\mathbb{P}\mathbb{D}^2X$ in short; but we deal with the $\mathbb{P}$ later, in §VII. Fix a finite state space X, and give it the discrete metric dis.

Instead of defining the hypers to be $\mathbb{D}^2X$ as in §II, we consider more generally the metric space of *hypermeasures* $\underline{\mathbb{K}}\mathbb{D}X$, recalling that $\underline{\mathbb{K}}$ includes $\mathbb{D}$ as a discrete/full special case. We continue to refer to these measures as hypers, i.e. submeasures over discrete distributions. If the measure $\Delta(\mathbb{D}X)$ assigned by hyper Δ to the whole space $\mathbb{D}X$ is less than 1, then we say that the deficit is Δ's probability of *diverging*. (Note that *elements* δ of $\mathbb{D}X$ are 1-summing.)

Now we begin to define an order between hypers based on information flow: given hypers $\Delta_{1,2}$ we would like to say that Δ_1 is refined by Δ_2 whenever Δ_2 "leaks no more information about X" (to an observer) than does Δ_1, where it is the distributions in $\mathbb{D}X$ that are the observables, similar to the belief states of *POMDP*'s. We will use a general notion of (conditional) entropy to formalise leakage.

B. Definition of disorder

There are a number of entropy-related testing definitions of security (see §IV-C immediately below) which we will generalise to "disorder" (employing a distinct word only to avoid overuse of "entropy").

Definition 1: Disorder test A *disorder test* on X is a non-negative function in $\mathbb{D}X \xrightarrow{\mathsf{D}} \mathbb{R}^{\geq}$, where $(\xrightarrow{\mathsf{D}})$ imposes the conditions concave, bounded, and continuous. (Recall that $\mathbb{D}X$ is a metric space, from §III-B.) The *disorder* of a hyper $\Delta{:}\,\underline{\mathbb{K}}\mathbb{D}X$ with respect a disorder test t is then $\int t \,\mathrm{d}\Delta$, that is the average of $t(\delta)$, weighted by Δ, over the distributions δ in $\mathbb{D}X$. □

Lemma 2: For any finite set of real-valued functions $F{\subseteq}X{\to}\mathbb{R}^{\geq}$ on X, the function $t{:}\,\mathbb{D}X{\to}\mathbb{R}^{\geq}$ defined as the infimum $t(\delta) = (\sqcap f{:}\, F \bullet \int f \,\mathrm{d}\delta)$ is a disorder test. (Recall that X is finite, so that $\int f \,\mathrm{d}\delta$ is in fact $\sum_{x:X} f(x){\times}\delta(x)$; the latter is more elementary, the former easier to read.) □

C. Relation of disorder to existing entropy tests

Since the 1980's, researchers such as Millen [23], Wittbold and Johnson [36], Gray [9] and others have described program security in terms of measures of uncertainty from Information Theory such as *Conditional Shannon Entropy* [29], *Guessing Entropy* [15] and *Bayes Risk* [30], [5], [2], [3], and each of these is a disorder test in the sense above. For Shannon Entropy use $\mathrm{H}(\delta){:=}-\sum_{x:\,X} \delta(x) \lg \delta(x)$; for Guessing Entropy use $\mathrm{W}(\delta){:=}\sum_{1\leq i\leq \#X} \sqcap^i\delta$ where $\sqcap^i\delta$ is the sum of the i smallest probabilities in δ; and for Bayes Risk use $\mathrm{B}(\delta){:=}1{-}\sqcup\delta$.

D. Definition of testing-based disorder refinement ($\sqsubseteq_{\mathcal{D}}$)

Definition 3: Disorder Refinement
For two hypers $\Delta_{\{1,2\}}$ we say that $\Delta_1 \sqsubseteq_{\mathcal{D}} \Delta_2$ just when $\int t \,\mathrm{d}\Delta_1 \leq \int t \,\mathrm{d}\Delta_2$ for all disorder tests t (Def. 1). □

Disorder Refinement is reflexive and transitive: both are consequences of reflexivity and transitivity of $(\leq)$ on the reals. Antisymmetry for $(\sqsubseteq_{\mathcal{D}})$ will follow from our main result in §VIII. For the moment we do not assume it.

V. The Space of Hypermeasures and its Structural Orders

The testing order of Def. 3 gives a negative witness, i.e. for the *failure* of refinement, a single test t for which $\int t \,\mathrm{d}\Delta_1 \not\leq \int t \,\mathrm{d}\Delta_2$; but to *establish* a refinement that way would require in principle quantifying over all tests. Here we present a structural order on $\underline{\mathbb{K}}\mathbb{D}X$, equivalent (we will later show) to Def. 3, and which provides a positive witness.

A. The divergence-refinement order $(\leq)$ over $\underline{\mathbb{K}}\mathbb{D}X$

The first structural order $\Delta_1 \leq \Delta_2$ holds just when diverging behaviour in Δ_1 has been replaced by "proper," i.e. arbitrary but non-diverging behaviour in Δ_2; it is based on a similar order in probabilistic/demonic semantics [12].

Definition 4: Divergence refinement
For hypers $\Delta_{\{1,2\}} \colon \underline{\mathbb{K}}\mathbb{D}X$ we say that $\Delta_1 \leq \Delta_2$ just when

$$\Delta_1(\mathbf{A}) \;\leq\; \Delta_2(\mathbf{A}) \qquad \text{for all measurable sets } \mathbf{A} \text{ in the Borel algebra of } \mathbb{D}X. \text{ (Recall §III-B.)}$$

□

B. The secure-refinement order $(\preceq)$ over $\underline{\mathbb{K}}\mathbb{D}X$

The second structural order $(\preceq)$ between hypers allows observer's ignorance, expressed as suspended inner distributions, to be increased by merging some of those inners; it is based on a similar order given earlier for the discrete, deterministic and non-diverging case [19], [21] that was synthesised from Bayes Risk via compositionality (see §X. Related work). We summarise its definition here, using an example for the discrete case $\mathbb{D}^2X$, and return to its measure-theoretic definition in the next section §VI below.

In §II the prize was initially hidden behind one of the three curtains p in $P{=}\{0,1,2\}$ with a known (a priori) distribution δ_P. Since we know what the initial distribution is, the hyper that describes the initial situation is $\{\!\{\delta_P\}\!\}$, which is how we write the point measure centred on that distribution δ_P.[6] That is, the initial distribution of p is known wprob. 1 to be δ_P. Name that hyper Δ_P.

Informally we regard Δ_P as at least as secure as any of $\Delta_{\{0,1,2\}}$ from §II since the latter are formed by the hosts's having leaked information from Δ_P via h, and indeed Δ_P can be constructed from any of the hypers $\Delta_{\{0,1,2\}}$ by weighted averaging of distributions within: we can take $\delta_{\{5,6\}}$ from Δ_2 and construct the weighted sum $0.6{\times}\delta_5 + 0.4{\times}\delta_6 = \delta_P$, since the probabilities assigned to 0,1,2 are $0.4{\times}\frac{1}{4}$, $0.6{\times}\frac{1}{2}$ and $0.6{\times}\frac{1}{2} + 0.4{\times}\frac{3}{4}$, that is the 0.1, 0.3 and 0.6 respectively of δ_P. Then if we *replace* the two inners $\delta_{\{5,6\}}$ in Δ_2 by δ_P, we get Δ_P as claimed.

Such merges can be carried out among many inners simultaneously (or not); and that is an informal description of secure refinement in the discrete case.

[6] Writing $\{\!\{\cdot\}\!\}$ is by analogy with singleton set $\{\cdot\}$.

C. The structural-refinement order $(\sqsubseteq_{\mathcal{S}})$ over $\underline{\mathbb{K}}\mathbb{D}X$

Structural refinement combines the two orders immediately above.

Definition 5: Structural refinement For hypermeasures $\Delta_{\{1,2\}} \colon \underline{\mathbb{K}}\mathbb{D}X$ say that $\Delta_1 \sqsubseteq_{\mathcal{S}} \Delta_2$ just when there is an intermediate hyper Δ with $\Delta_1 \leq \Delta$ and $\Delta \preceq \Delta_2$. □

Thus structural refinement composes the other two, allowing divergence to be replaced by proper behaviour and then allowing insecurity to be made more secure.

VI. Secure Refinement in Detail

A. Secure refinement for measures: formal definition

We saw in §V-B just above that secure refinement is a matter of merging together inner (discrete) distributions within a hyper. When that hyper is a proper measure, we use the following algebraic definition of $(\preceq)$, based on the Kantorovich construction from §III.

Definition 6: Secure refinement for measures Given hypers $\Delta_{\{1,2\}} \colon \underline{\mathbb{K}}\mathbb{D}X$, we say that $\Delta_1 \preceq \Delta_2$ just when there exists a so-called *super*-distribution $\mathbf{\Delta}$ in $\underline{\mathbb{K}}\underline{\mathbb{K}}\mathbb{D}X$, thus one level further up, that can be averaged and $\underline{\mathbb{K}}$-averaged onto Δ_1 and Δ_2 respectively.[7] That is, we need $\mathbf{\Delta}$ with

$$\Delta_1 = \boldsymbol{\mu}_{\mathbb{D}X}(\mathbf{\Delta}) \qquad \text{and} \qquad \underline{\mathbb{K}}\boldsymbol{\mu}_X(\mathbf{\Delta}) = \Delta_2 \;. \tag{2}$$

□

B. Secure refinement for measures: example

We argued informally in §V-B that $\Delta_2 \preceq \Delta_P$. Here we give a super that establishes that refinement via Def. 6: it is just the point-super $\mathbf{\Delta} := \{\!\{\Delta_2\}\!\} = \boldsymbol{\eta}_{\underline{\mathbb{K}}\mathbb{D}X}(\Delta_2)$ that assigns probability 1 to the hyper Δ_2. It has this particularly simple form here because we are merging *all* of the inners in Δ_2 together; and Def. 6 applies because $\boldsymbol{\mu}_{\mathbb{D}X} \circ \boldsymbol{\eta}_{\underline{\mathbb{K}}\mathbb{D}X}$ is the identity $1_{\underline{\mathbb{K}}\mathbb{D}X}$ (for the Δ_2 case in (2) left) and $\underline{\mathbb{K}}\boldsymbol{\mu}_X \circ \boldsymbol{\eta}_{\underline{\mathbb{K}}\mathbb{D}X} = \boldsymbol{\eta}_{\mathbb{D}X} \circ \boldsymbol{\mu}_X$ (for the Δ_P case (2) right).

C. Structural refinement for measures: properties

The space $(\underline{\mathbb{K}}\mathbb{D}X, \sqsubseteq_{\mathcal{S}})$, with its structural refinement combining divergence- and secure refinement, satisfies some basic properties we have proved elsewhere [21].

Theorem 7: Basic order properties For finite state-space X we have the following:

(i) As a relation $(\sqsubseteq_{\mathcal{S}})$ is reflexive, antisymmetric and transitive [21, App. B].
(ii) $\underline{\mathbb{K}}\mathbb{D}X$ is compact [33], [21, App. A].
(iii) $(\underline{\mathbb{K}}\mathbb{D}X, \sqsubseteq_{\mathcal{S}})$ is chain-complete [21, App. D].

□

[7] In a general monad $\mathbb{M}$ the transformation $\boldsymbol{\mu}_Y$ has type $\mathbb{M}^2Y \to \mathbb{M}Y$. Here we are effectively using $\underline{\mathbb{K}}^2(\underline{\mathbb{K}}X) \to \underline{\mathbb{K}}(\underline{\mathbb{K}}X)$ and $\underline{\mathbb{K}}^2X \to \underline{\mathbb{K}}X$, because both $\mathbb{K}$ and $\mathbb{D}$ are substructures of $\underline{\mathbb{K}}$.

D. Main result: disorder- versus structural refinement

Our main result, in §VIII, will be that our two refinement orders, the testing-based disorder refinement and the merging-based structural refinement, are identical over a powerdomain $\mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$ that we will construct over $\underline{\mathbb{K}}\mathbb{D}X$ to introduce demonic nondeterminism. It has as a corollary that the two refinement orders agree on $\underline{\mathbb{K}}\mathbb{D}X$ itself.

A component of the more general proof, for sets of hypers, is the following theorem for individual hypers:

Theorem 8: Structural refinement implies disorder refinement For any hypers $\Delta_{\{1,2\}}: \underline{\mathbb{K}}\mathbb{D}X$, if $\Delta_1 \sqsubseteq_{\mathcal{S}} \Delta_2$ (recall §V-C) then $\Delta_1 \sqsubseteq_{\mathcal{D}} \Delta_2$ (recall §IV-D).

Proof: By the transitivity of $(\leq)$ between real values and the definition of structural refinement (Def. 5) it is enough to show separately that for any (concave, bounded) $t: \mathbb{D}X \xrightarrow{\mathsf{D}} \mathbb{R}^{\geq}$ each of (i) $\Delta_1 \leq \Delta_2$ and (ii) $\Delta_1 \preceq \Delta_2$ implies that the inequality holds. The first claim (i) follows since t takes only non-negative values.

For (ii) we reason as below. In general by $\int_Y F\, \Phi(\mathrm{d}y)$ we mean the integral of the expression F, understood as a (measurable) function in y, over the measure Φ restricted to the (measurable) set Y. If Y is the whole of the sample space, we omit it; and if F is the expression $f(y)$ for some function f then we write $\int f \,\mathrm{d}\Phi$ [7].

We take any hypermeasures $\Delta_{\{1,2\}}$ in $\underline{\mathbb{K}}\mathbb{D}X$ such that $\Delta_1 \preceq \Delta_2$, and let $\blacktriangle$ be the super in $\underline{\mathbb{K}}\underline{\mathbb{K}}\mathbb{D}X$ supplied by Def. 6, i.e. with $\Delta_1 = \boldsymbol{\mu}_{\mathbb{D}X}(\blacktriangle)$ and $\underline{\mathbb{K}}\boldsymbol{\mu}_X(\blacktriangle) = \Delta_2$. Then

$$
\begin{array}{lll}
& \int t \,\mathrm{d}\Delta_1 & \\
= & \int t \,\mathrm{d}(\boldsymbol{\mu}_{\mathbb{D}X}(\blacktriangle)) & \text{“assumption } \Delta_1 = \boldsymbol{\mu}_{\mathbb{D}X}(\blacktriangle)\text{”} \\
= & \int \left(\int t \,\mathrm{d}\Delta\right) \blacktriangle(\mathrm{d}\Delta) & \text{“ [8, Thm.1(d)] ”} \\
\leq & \int t(\boldsymbol{\mu}_X(\Delta)) \,\blacktriangle(\mathrm{d}\Delta) & \text{“}t \text{ is concave and Jensen's inequality [22, Thm.2] (App. C)”} \\
= & \int (t \circ \boldsymbol{\mu}_X) \,\mathrm{d}\blacktriangle & \text{“rewrite”} \\
= & \int t \,\mathrm{d}(\underline{\mathbb{K}}\boldsymbol{\mu}_X(\blacktriangle)) & \text{“[8, Thm.1(a)]”} \\
= & \int t \,\mathrm{d}\Delta_2 \ , & \text{“assumption } \underline{\mathbb{K}}\boldsymbol{\mu}_X(\blacktriangle) = \Delta_2\text{”}
\end{array}
$$

as required. □

We now introduce the powerdomain.

VII. Powerhypers and their Refinement Orders

Our example of §II had nondeterminism, implicit in the contestant C's choice of strategy. We introduce that into our model by taking *sets* of hypers, rather than single hypers alone. Thus we introduce $\mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$, with the powerset constructor $\mathbb{P}$ expressing demonic choice, and we lift our two refinement orders $(\sqsubseteq_{\mathcal{S}})$ and $(\sqsubseteq_{\mathcal{D}})$ to the powersets. Our main result will be that the orders are identical when restricted to those sets satisfying certain *healthiness* properties defined below.

A. Disorder refinement for powerhypers

We extend the notion of disorder test from hypers to healthy sets of hypers: the *disorder* of a subset $Z \in \mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$, wrt. a disorder test t as given in §IV, is the infimum $(\sqcap)$ of the t-disorders of its constituent hypers.

$$Z \star t \quad := \quad (\sqcap \Delta{:}\, Z \bullet \textstyle\int t \,\mathrm{d}\Delta) \ . \tag{3}$$

Using that, we define a (pre-)order on subsets generalising the pre-order given at Def. 3.

Definition 9: Disorder refinement for subsets Subset $Z_2 \in \mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$ is “at least as disordered as” $Z_1 \in \mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$ when for any test t we have $Z_1 \star t \leq Z_2 \star t$. We write that $Z_1 \sqsubseteq_{\mathcal{D}} Z_2$, using the same symbol as for individual hypers. □

Def. 9 is the formalisation of what we mean by “quantitative security”: if $Z_1 \sqsubseteq_{\mathcal{D}} Z_2$ then Z_2 is “no less secure than” Z_1 under any measure of disorder as we have defined it. Our earlier work [19] shows that, in the context of programming language denotations, this more general approach based on arbitrary disorder tests is consistent with entropies like those mentioned in §IV-C.

As with its specialisation Def. 3 to individual hypers, we find that Def. 9 on sets of hypers is reflexive and transitive. We will see from our principal theorem in §VIII that it is antisymmetric as well, thus a partial order, when restricted to certain “healthy” sets as in the next section.

B. Healthiness conditions

Observe that the following two conditions, as well as closure (in $\underline{\mathbb{K}}\mathbb{D}X$), are invariant under (3), in the sense that none of convex-closing, $(\sqsubseteq_{\mathcal{D}})$-closing nor limit-point closing (wrt. the metric) a subset Z changes its disorder.

Definition 10: Convexity A set $Z: \mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$ of hypermeasures is *convex* when, for all $\Delta_{\{1,2\}}: Z$ and probabilities $p: [0,1]$, the weighted average $\Delta_1 \,{}_p\!\oplus\, \Delta_2$ is also in Z. □

Definition 11: $(\sqsubseteq_{\mathcal{D}})$-closed A set $Z: \mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$ of hypermeasures is $(\sqsubseteq_{\mathcal{D}})$-*closed* if $\Delta_1 \in Z$ and $\Delta_1 \sqsubseteq_{\mathcal{D}} \Delta_2$ implies $\Delta_2 \in Z$ also. □

Definition 12: Healthy sets of hyperdistributions A set of hypermeasures is *healthy* iff it is non-empty, convex, limit-point closed, hence compact by Thm. 7(ii), and $(\sqsubseteq_{\mathcal{D}})$-closed. □

We denote healthy sets of hypers on X by $\mathbb{H}X$, and call them *powerhypers*. And noninterference with information hiding, probability and external nondeterminism is then represented as single powerhyper.

As an illustration we again refer to Monty Hall. The basic strategies result in a set of three output behaviours $\{\Delta_0, \Delta_1, \Delta_2\}$. The smallest healthy set containing $\{\Delta_0, \Delta_1, \Delta_2\}$ consists of the set of all possible hypers made from probabilistic combinations of the basic strategies and secure refinements of them.

C. Structural refinement for powerhypers

We now turn to our structural order, extending it also from hypers to healthy sets of hypers. Because we have imposed $(\sqsubseteq_{\mathcal{D}})$-closure on the powerhypers $\mathbb{H}X$, and we have Thm. 8, we can see that healthy sets are $(\sqsubseteq_{\mathcal{S}})$-closed.

Definition 13: Structural refinement for powerhypers Powerhyper Z_2 is a *structural refinement* of powerhyper Z_1 just when $Z_1 \supseteq Z_2$. We write that $Z_1 \sqsubseteq_{\mathcal{S}} Z_2$, using the same symbol as for individual hypers. □

We can show immediately that structural refinement is a complete partial ordering over powerhypers.

Theorem 14: $(\mathbb{H}X, \sqsubseteq_{\mathcal{S}})$ *is a cpo*

Proof: That $(\sqsubseteq_{\mathcal{S}})$ defines a partial order over sets of hypers is because subset inclusion does. To prove that it is also complete we must show that the infinite intersection of any chain of powerhypers, i.e. healthy sets of hypers, is also healthy (Def. 12). Observe that the defining closures: convex, $(\sqsubseteq_{\mathcal{D}})$-closure and limit-point closure are preserved by arbitrary intersection. Non-emptiness follows from compactness of $\underline{\mathbb{K}}\mathbb{D}X$, Thm. 7(ii), and the "finite-intersection property" of chains. □

VIII. Structural Refinement and Disorder Refinement are Equivalent

In this section we sketch our principal result, that disorder-testing of powerhypers characterises structural refinement exactly: we show for any powerhypers Z_1 and Z_2 that $Z_1 \sqsubseteq_{\mathcal{S}} Z_2$ if and only $Z_1 \sqsubseteq_{\mathcal{D}} Z_2$. (The proof details are supported by material in the appendices.) We discuss the significance of this result in the conclusion.

A. Structural refinement implies disorder refinement

This is immediate, since an infimum over a subset cannot have decreased.

Corollary 15: Structural refinement is consistent with preserving disorder in the special cases of Shannon Entropy, Guessing Entropy and Bayes Risk.

Proof: Follows from the above and from §IV-C where all three entropies are given in terms of disorder tests. □

B. Disorder refinement implies structural refinement

We establish the contrapositive in the following theorem.

Theorem 16: Disorder refinement is complete If structural refinement fails, there is a single disorder test t demonstrating that. That is, for powerhypers $Z_{\{1,2\}} : \mathbb{H}X$ with $Z_1 \not\sqsubseteq_{\mathcal{S}} Z_2$ we have $Z_1 \star t > Z_2 \star t$ for some disorder test $t : \mathbb{D}X \xrightarrow{\mathsf{D}} \mathbb{R}^{\geq}$, whence $Z_1 \not\sqsubseteq_{\mathcal{D}} Z_2$.

Proof: The proof here follows the structure of an earlier, less general result, for single, full, discrete hypers $\Delta_{\{1,2\}}$ and secure refinement $(\preceq)$ between them [19]. The extra generality here is that we are concerned with *sets* of possibly *partial*, *non-discrete* hypers, and *structural* refinement $(\sqsubseteq_{\mathcal{S}})$. The earlier result was for discrete hypers $\Delta_{\{1,2\}}$ that $\int t \ \mathrm{d}\Delta_1 \leq \int t \ \mathrm{d}\Delta_2$ for all disorder tests t implied $\Delta_1 \preceq \Delta_2$.

That earlier proof assumed $\Delta_1 \not\preceq \Delta_2$, projected $\Delta_{\{1,2\}}$ into a finite-dimensional Euclidean space where the $(\preceq)$-closure of Δ_1 determined a closed, convex set not containing the projection of Δ_2, and finally induced a distinguishing test t from a hyperplane separating the projection of Δ_2 from that closure.

The result here is more complex, since in dealing with sets of hypers we need closure of those sets, hence a (Kantorovich) metric wrt. which closure can be defined, and then technical results concerning continuity of the projections into the finite-dimensional space. Furthermore, since we are now dealing with measures rather than discrete distributions, extra care is required when defining those continuous-to-discrete projections. Here we provide an overview of the full proof in the appendix.

We begin by assuming $Z_1 \not\sqsubseteq_{\mathcal{S}} Z_2$.

—Step 1 By Def. 13 we have $Z_1 \not\supseteq Z_2$, and so there is $\Delta_2 \in Z_2$ with $\Delta_2 \notin Z_1$. In fact *wlog* we can assume Δ_2 to have finite support, and so treat it as a discrete distribution. We show that as follows:

- Consider our Δ_2 with $Z_2 \ni \Delta_2 \notin Z_1$. It can be shown that for any $\varepsilon > 0$ there is a finitely supported hyper $\Delta_2 \sqsubseteq_{\mathcal{S}} \Delta_2^{\varepsilon}$ with $\boldsymbol{k}(\Delta_2, \Delta_2^{\varepsilon}) \leq \varepsilon$ (Lem. 21 in App. D). And $\Delta_2^{\varepsilon} \in Z_2$ from Thm. 8 and $(\sqsubseteq_{\mathcal{D}})$-closure, hence $(\sqsubseteq_{\mathcal{S}})$-closure of Z_2.
- If closed set Z_1 also contained all those Δ_2^{ε}'s, then it would contain their limit point Δ_2; but it does not. Thus take *wlog* some Δ_2^{ε} with $Z_2 \ni \Delta_2^{\varepsilon} \notin Z_1$ instead of Δ_2.

We fix the finitely supported Δ_2 from now on, and assume $\lceil \Delta_2 \rceil$ is non-empty; the empty case can be handled by similar but simpler techniques.

—Step 2 For that Δ_2, we explicitly index its support by defining non-empty finite set $I := \{1..\#\lceil \Delta_2 \rceil\}$ and an arbitrary injection $\hat{\imath} : \lceil \Delta_2 \rceil \to I$. Using $\hat{\imath}$, we project $\Delta_2 \in Z_2$ onto a point p_2^E within the set $\underline{\mathbb{D}}(X \times I)$ of discrete *sub*-distributions on $X \times I$, isomorphic to a right tetrahedron in the Euclidean unit cube E of $\#X \times \#I$-dimensions [18, Chap. 5]. (See Fig. 2 in App. D for an illustration.)

The $\hat{\imath}$-projection p_2^E of Δ_2 into E is $p_2^E := \mathsf{flatten}_{\hat{\imath}}(\Delta_2)$ where function $\mathsf{flatten}_{\hat{\imath}} : \underline{\mathbb{K}}\mathbb{D}X \to \underline{\mathbb{D}}(X \times I)$ is defined as

$$\mathsf{flatten}_{\hat{\imath}}(\Delta_2)(x,i) := \int_{\hat{\imath}^{-1}(i)} \delta(x) \ \Delta_2(\,\mathrm{d}\delta) \qquad \text{for } (x,i) : X \times I.$$

(See Fig. 3 in App. E for an equivalent and simpler definition of flatten in the discrete case.)

In the reverse direction we define function $\mathsf{lift} : \underline{\mathbb{D}}(X \times I) \to \underline{\mathbb{K}}\mathbb{D}X$ to take any δ' in Euclidean $E \approx \underline{\mathbb{D}}(X \times I)$ back to a (discrete) hyper in $\underline{\mathbb{K}}\mathbb{D}X$. First, write $\delta'_{\downarrow I} : \underline{\mathbb{D}}I$ for the marginal distribution of δ' onto I (i.e. $\underline{\mathbb{K}}\pi_2(\delta')$ for $\pi_2(x,i) := i$), and define $\delta'_I : I \to \mathbb{D}X$ so that $\delta'_I(i)$ is the conditioning of δ' wrt. i. Now define $\mathsf{lift}(\delta') := \underline{\mathbb{K}}\delta'_I(\delta'_{\downarrow I})$. It is then a routine calculation to show that we have $\mathsf{lift}(p_2^E) = \Delta_2$, since $\hat{\imath}$ is injective (Lem. 24 in App. E). We (usually) omit the $\hat{\imath}$ subscript of flatten. (See Fig. 4 in App. E for lift in the discrete case.)

—Step 3 We embed the whole of Z_1 also into E, as a non-empty, convex, closed and $(\leq)$-closed set Z_1^E not containing p_2^E. The embedding of Z_1 is via the inverse image through lift, that is $Z_1^E := \{\delta' : E \mid \mathsf{lift}(\delta') \in Z_1\}$. Since $p_2^E \in Z_1^E$ would imply $\Delta_2 = \mathsf{lift}(p_2^E) \in \mathsf{lift}(Z_1^E) \subseteq Z_1$, from our assumption we indeed have $p_2^E \notin Z_1^E$. Non-emptiness of Z_1^E can be shown from the non-emptiness and $(\sqsubseteq_{\mathcal{S}})$-closure of Z_1.

Closure of Z_1^E follows from closure of Z_1 and continuity of lift (Lem. 25 in App. F). Convexity and $\leq$-closure of Z_1^E follow from the fact that lift is "super-linear" (Lem. 26 in App. F) in the sense that it satisfies $c_1 \times \mathsf{lift}(\delta_1) + c_2 \times \mathsf{lift}(\delta_2) \ \sqsubseteq_{\mathcal{S}} \ \mathsf{lift}(c_1 \times \delta_1 + c_2 \times \delta_2)$ for all constants $c_{\{1,2\}} : \mathbb{R}^{\geq}$ with $c_1 + c_2 \leq 1$ and $\delta_{\{1,2\}} : \underline{\mathbb{D}}(X \times I)$, and Z_1 is convex and $(\sqsubseteq_{\mathcal{S}})$-closed.

—Step 4 The Separating Hyperplane Lemma [32] requires convexity and closure of a non-empty set to separate it from a point using an intervening plane. We have established those characteristics of Z_1^E, and so we are guaranteed the existence of at least one hyperplane in E with a normal H in $X\times I\to\mathbb{R}$ so that all of Z_1^E is strictly above the plane and p_2^E is strictly below the plane, i.e.

$$\int H\ \mathrm{d}p_2^E \quad < \quad (\sqcap p_1^E{:}\, Z_1^E \bullet \int H\ \mathrm{d}p_1^E)\ , \tag{4}$$

Recall that X, I are finite, so that the integrals are $\sum_{x,i} H(x,i)\times p_1^E(x,i)$ etc. Since Z_1^E is $(\leq)$-upclosed we have that at least one such H is non-negative (Lem. 27 in App. G). Fix that $H{:}\, X\times I\to\mathbb{R}^{\geq}$ as a separating hyperplane.

—Step 5 We now use H and Lem. 2 to construct a test t in $\mathbb{D}X\to^{\mathsf{D}}\mathbb{R}^{\geq}$ via the definition

$$t(\delta) \quad := \quad (\sqcap i{:}\, I \bullet \textstyle\int H(x,i)\,\delta(\mathrm{d}x))\ , \tag{5}$$

so inducing a disorder $\int t\ \mathrm{d}\Delta$ for any $\Delta{:}\,\underline{\mathbb{K}}\mathbb{D}X$.

This test t has two crucial properties derived from $\mathsf{flatten}$ and lift: the <u>first</u> one is that $\int t\ \mathrm{d}\Delta_2 \leq \int H\ \mathrm{d}(\mathsf{flatten}_{\hat{\imath}}(\Delta_2)) = \int H\ \mathrm{d}(p_2^E)$, since the *lhs* takes a minimal choice of indices while $\mathsf{flatten}_{\hat{\imath}}$ takes an arbitrary choice $\hat{\imath}$ (Lem. 28 in App. H). The <u>second</u>, more significantly, is that for any Δ in Z_1 we have $\int t\ \mathrm{d}\Delta = \int H\ \mathrm{d}(p_1^E)$ for some $p_1^E\in Z_1^E$ (Lem. 31 in App. H). With those we conclude our proof as follows.

—Step 6 Using test t constructed (5) from the normal H, and the two crucial properties of it, we calculate

$$\begin{array}{lll} & (\sqcap\Delta{:}\, Z_1 \bullet \int t\ \mathrm{d}\Delta) & \\ \geq & (\sqcap p_1^E{:}\, Z_1^E \bullet \int H\ \mathrm{d}(p_1^E)) & \text{“}\underline{\text{second}}\text{ property above”} \\ > & \int H\ \mathrm{d}(p_2^E) & \text{“strict-separation property of } H\text{”} \\ \geq & \int t\ \mathrm{d}\Delta_2 & \text{“}\underline{\text{first}}\text{ property above”} \\ \geq & (\sqcap\Delta{:}\, Z_2 \bullet \int t\ \mathrm{d}\Delta)\ . & \text{“}\Delta_2\in Z_2\text{”} \end{array}$$

Thus this t is the test we require. □

This gives us immediately the analogue of Thm. 14: not only is $(\sqsubseteq_{\mathcal{D}})$ antisymmetric, but $(\mathbb{H}X, \sqsubseteq_{\mathcal{D}})$ is a cpo as well.

IX. *POMDP*'s and an Example

A. Conventional POMDP's based on states

We now sketch how our definitions relate to standard *POMDP*'s [13], and how in particular they abstract from them. A *POMDP* over finite state space X comprises a finite set of labels $\mathcal{L}$ with for each label $l{:}\,\mathcal{L}$ a Markov transition matrix $T_l{:}\, X\to\mathbb{D}X$, and a finite observation-set Ω with map $\mathcal{O}{:}\,\mathcal{L}\times X\to\mathbb{D}\Omega$. From initial state $x{:}\, X$, a label $l{:}\,\mathcal{L}$ is selected and the final state x' is chosen according to distribution $T_l(x)$; the observable output ω is chosen according to distribution $\mathcal{O}(l,x')$.

We assume the observer knows the details of matrix T, map $\mathcal{O}$ and label l when he observes the output ω. From that ω, he computes a (conditional) probability over the final state x' summarising his “belief” of its value: that distribution in $\mathbb{D}X$ is called a belief state.

For a *particular* $l{:}\,\mathcal{L}$ and an initial state x the combined action above thus produces a joint distribution in $\mathbb{D}(X\times\Omega)$ defined by $D_{x,l}(x',\omega) := T_l(x)(x')\times\mathcal{O}(l,x')(\omega)$; thus varying $l{:}\,\mathcal{L}$ results in a *set* of (joint) distributions.

B. POMDP's lifted to belief states

Our alternative presentation suppresses the observations' *values* (typically written ω) while retaining their effect on information flow: this is the sense in which we abstract. Attention moves from states in X to *belief* states in $\mathbb{D}X$. We discuss the strategic reason for this in the final paragraph of this section, and further in the conclusion §XI.

Our *POMDP*-like structure is a function P of type $\mathcal{L}\times\mathbb{D}X\to\mathbb{K}\mathbb{D}X$, where from initial belief state δ_0 and label l, a hyper $\Delta := P(l,\delta_0)$ results. First we lift T_l and $D_{x,l}$ above to depend on input belief state δ_0: they become

$$\begin{array}{rcl} \widehat{T_l}(\delta_0)(x') & := & \sum_{x:\,X}\delta_0(x)\times T_l(x)(x') \\ \widehat{D_{\delta_0,l}}(x',\omega) & := & \sum_{x:\,X}\delta_0(x)\times T_l(x)(x')\times\mathcal{O}(l,x')(\omega)\ . \end{array}$$

Then we convert that joint distribution $\widehat{D_{\delta_0,l}}(x',\omega)$ on $X\times\Omega$ to a distribution of distributions, a hyper, on X alone.[8]

To eliminate the observations' values, for each value $\omega{:}\,\Omega$ define belief state $\delta_\omega{:}\,\mathbb{D}X$ as $\delta_\omega(x') := \widehat{D_{\delta_0,l}}(x',\omega)/\widehat{D_{\delta_0,l}}(X,\omega)$, i.e. the conditional probability of x' given that ω. Then form hyper Δ by taking those δ_ω's as its support and assigning probability $\Delta(\delta_\omega) := \widehat{D_{\delta_0,l}}(X,\omega)$ to each one.

C. Iterated POMDP's and nondeterminism

For multiple executions of P we must further lift the action from belief states to hypers, i.e. distributions on them. That is done in the usual Kleisli style using our monadic constructions (as in [21] but without nondeterminism).

But when we allow labels l to vary, we are adding a further level, one extra level of nondeterminism that correponds to the outer $\mathbb{P}$ in our powerhypers $\mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$. We define a choice function $f{:}\,\mathbb{D}X\to\mathcal{L}$ whose domain is $\mathbb{D}X$ rather than X because the choice of l, equivalently the resolution of the nondeterminism, has no access to the actual value of the state in X — the choice is based on the *belief*, as a distribution, of the state value; and that in turn is determined by the sequence of observations made up to this point. The abstraction is that we do not retain those sequences in our model: we keep only the belief states that result from them, assigning to each one the probability of the observation sequence that generated it. Further work is needed to use the constructions we have given here to define a monad on the full $\mathbb{P}\underline{\mathbb{K}}\mathbb{D}X$, thus removing the need to refer explicitly to l.

D. Strategic reasons for choosing hypers as a representation

By suppressing the observations, we have represented the output as a two-level hyper structure rather than what would be a conceptually simpler joint distribution. (The functions $\mathsf{flatten}$ and lift from §VIII-B, and illustrated in Figs. 3,4 of App. E, effectively translate between the two.) Intuitively, we do this

[8] In essence this converts a joint-distribution matrix to a distribution on its normalised columns (themselves distributions). The columns' names (the observations) are no longer needed.

because the observations are not themselves of interest if we are considering flow of information about X — only what we learn about X's belief state is important. Mathematically, retaining the observations would incur two penalties: the first is that with successive executions of the *POMDP* they would become *sequences* of observations; and the second is that comparison of information flow would then become a proper equivalence relation instead of an equality, complicating its use as the semantics of a programming language. (See §XI.)

E. An example

Let X be $\{a, b\}$ and write δ_p for the discrete distribution $a_p \oplus b$ in $\mathbb{D}X$. Let P_n be a *POMDP* with input X, observables T, F and labels $\{1..2^n{-}1\}$ with the property that executing P_n from belief state $\delta_{l/2^n}$ with label l produces observables F, T with equal probability: thus if F is observed, then the outgoing belief state is $\delta_{(2l-1)/2^{n+1}}$; and if T is observed, then it is $\delta_{(2l+1)/2^{n+1}}$. (See App. I.)

Thus in the notation of §IX-B we have $P_1(1, \delta_{1/2}) = \delta_{1/4\ \frac{1}{2}\oplus}\ \delta_{3/4}$; and $P_2(3, \delta_{3/4}) = \delta_{5/8\ \frac{1}{2}\oplus}\ \delta_{7/8}$. By repeated executions $P_1; P_2 \cdots$ with labels chosen as above we can produce a tree like pattern of belief states. (See (13) and Fig. 5 of App. I.) If we do it long enough, the coverage of the belief states' probability (the p in δ_p) can be made arbitrarily fine over $[0, 1]$.

Now suppose we are given an iterated system $P_1; \cdots; P_n; \cdots; P_N$ with the labels chosen as above but with the N unknown to us beforehand. The aggregate observations will be a sequences of values T, F, one emitted by each P_n, and we would like to bound from below the entropy remaining in X after all those observations. Without knowing N, and taking a conventional approach, we would: first decide which entropy we wanted (Shannon, Bayes-Risk, other); then calculate that entropy as a function of N, observing that it is decreasing (since each state releases more information); and finally take the limit of that entropy as $N{\to}\infty$. The result would give us a "least assured entropy" (of whichever type we chose) for any N.

Our alternative approach is to observe that the output hypers $\Delta_1, \Delta_2 \cdots$ say form a $(\sqsubseteq_{\mathcal{S}})$-decreasing sequence which, though with discrete elements in $\mathbb{D}^2X$, has $(\sqsubseteq_{\mathcal{S}})$-infimum $\Delta_{[0,1]}$, the *continuous* hyper in $\underline{\mathbb{K}}\mathbb{D}X$ where the p in δ_p is chosen uniformly from $[0, 1]$. From our principal result in §VIII we know that any disorder measure on Δ_N is bounded from below by that same disorder measure applied to $\Delta_{[0,1]}$. Thus in particular we have

- The Conditional Shannon Entropy of any Δ_N is at least the conditional Shannon Entropy of $\Delta_{[0,1]}$, and that is $\int_0^1 -x \log x - (1-x)\log(1-x)\ \mathrm{d}x = 1/2$.
- The Bayes Risk of any Δ_N is at least the Bayes Risk of $\Delta_{[0,1]}$, and that is $\int_0^1 x\ \mathbf{max}\ (1-x)\ \mathrm{d}x = 1/4$.

In a program-refinement setting, testing provides a very robust form of equivalence. The measurement of information leaked by successive Δ_n's is reduced to a proof of structural refinement between $\Delta_{[0,1]}$ and Δ_N, together with a *single* integration over $\Delta_{[0,1]}$ with respect to the relevant entropy. In this case, both verification tasks are routine (App. I).

X. Related Work

A number of notable investigations treat probability and nondeterminism from a domain-theoretic perspective, with the aim of obtaining equational theories which suitably generalise that of the set-theoretic power constructor. The probabilistic powerdomain constructor $\mathcal{V}$ [12] generates an order on distributions, which for us would be $\underline{\mathbb{K}}\mathbb{D}X$, based on the Scott-open sets induced by an underlying order on points, for us $\mathbb{D}X$: but there does not seem to be any order on $\mathbb{D}X$ that would generate our testing- or structural orders for information flow in that way. [9]

He, McIver et al. [11], [17] as well as Tix et al. [31] and Mislove [24] identify convex closure as an important healthiness condition on models justifying $Z\ {}_p\!\oplus\ Z = Z$, i.e. that the probabilistic choice between two identical (denotations) of programs is indistinguishable from the program itself. Varacca and Winskel [35] do not do this, instead introducing *indexed valuations* as a more general constructor that can be combined with the probabilistic power domain.

Our principal aim here is to ensure that the power domain constructions correspond to an observation space in which program refinement can be interpreted as preservation of information flow; however our convex-closed healthiness condition implies the usual probability law for external probability. The observational equivalence for *POMDP*'s has been investigated by Castro et al. [4], but from the perspective of the conditional probability distributions rather than via entropy measurement.

In the context of security semantics, similar domain-theoretic techniques have been used by Sabelfeld and Sands for example [28], although with a different underlying order from ours. Indeed there remains a range of information orders on which to base measurement of programs' information leakage, originally using Shannon Entropy [23], [36], [9], but more recently e.g. Bayes Risk [30], [5], [2], [3], and Marginal Guesswork [27], [14] have been considered. A quantitative information flow relation between programs has been defined by Yasuoka and Terauchi [37] but not yet investigated in full detail.

However a general "refinement" order characterising a single measure of information flow for defining program semantics remains problematic: if two programs are deemed equivalent with respect to Bayes Risk (say), they might still be inequivalent with respect to Shannon Entropy. Braun et al. [2] for example get around this lack of *compositionality* with respect to Bayes Risk by identifying a set of operators and contexts which can be used in combination to ensure that a required Bayes Risk is not violated by a program. Our own complementary approach is to construct the compositional closure of Bayes-Risk comparisons wrt. a fixed programming language: it effectively synthesises the secure-refinement order

[9] The various entropies in §IV-C do determine orders on $\mathbb{D}X$; but they are total orders and, moreover, they disagree with each other in general.

($\preceq$) of §V-B, which stronger relation we can then use as a compositional comparison for all operators. But our earlier work did not treat nondeterminism [19].

Finally, Pfitzmann, Waidner and Backes [26], [1] have combined the notion of simulation with concurrency, channels, probability and computational-based cryptography. An interesting topic of future research is to investigate how the refinement ordering in our domain-theoretic model relates to such simulation relations and, for example, whether a simulation relation can similarly be explained in terms of information flow expressed as a disorder test.

In our own earlier work, we gave explicit constructions for models of probabilistic and nondeterministic programs (but no hidden state) [11], [25], [18], [17], and with probability and hidden state (but no nondeterminism) [19], [21]. For both constructions we provided an equivalent testing-based characterisation. The present work is an extension of both.

XI. Conclusion

We have given an explicit power-domain construction for models of hidden state, including both probability and nondeterminism: inspired by *POMDP*'s, it is based on a novel security refinement order [19], [21], which however did not include nondeterminism. Our principal technical contribution here is to add nondeterminism to that earlier order, and then to extend to measures the equivalence between structural- and testing refinement we earlier proved for the discrete case [19]. The testing refinement is general enough to include well-known entropies for information flow, such as Shannon Entropy, Bayes Risk and Guessing Entropy.

Our strategic goal for the research is to give a *quantitative denotational semantics for a programming language* with noninterference-style information hiding, probability and nondeterminism. The semantics of [19] accepted a straight-line language (i.e. without loops) with probabilistic but not demonic choice, and gave a discrete semantics; in [21] we extended that to loops, including their possible divergence, and gave a hyperdistribution-based semantics but still without demonic choice. Here we add demonic choice, preparing a semantic domain of *sets* of measures that will be suitable for a programming language with demonic choice as well. Our observations are determined by *visibility* declarations on the program variables as well as the structure of the program [19, Sec. 8] [21, Sec. 8.5].

A crucial feature of our planned work, though not stressed here, is compositionality of refinement. As we reviewed in [19], the popular measures of Shannon Entropy etc. are not compositional in even the simple framework we chose, and not in others either [2]. Our structural order was synthesised as the weakest strengthening of the Bayes-Risk order that *was* compositional, the so-called *compositional closure* of Bayes Risk [19]. More generally our focus on compositionality can be seen as a preliminary for program algebra where equality of algebraic terms means they have exactly the same information flow properties as measured by disorder tests. That remains a topic for future work.

Although the compositionally closed refinement was sound for Shannon-Entropy- and Guessing-Entropy comparisons as well, we left as an open question whether it was complete for those (as it is for Bayes Risk). With nondeterminism (for which the $\mathbb{PKD}X$ domain here has been constructed), we hope the resulting more-discriminating contexts will make the structural order complete for any disorder test, thus settling that open question.

Acknowledgements

We thank James Worrell, David Fremlin and Yuxin Deng for helpful correspondence during this work, and Frits Vaandrager who hosted our six-month stay at Radboud University in Nijmegen. The research was supported by the Australian ARC (Grant DP1092464) and the Dutch NWO (Grant 040.11.303).

References

[1] M. Backes and B. Pfitzmann, "A general composition theorem for secure reactive systems," in *Theory of Cryptography*, ser. LNCS, M. Naor, Ed., vol. 2951. Springer, 2004, pp. 336–354.

[2] C. Braun, K. Chatzikokolakis, and C. Palamidessi, "Compositional methods for information-hiding," in *Proc. FOSSACS'08*, ser. LNCS, vol. 4962. Springer, 2008, pp. 443–57.

[3] ——, "Quantitative notions of leakage for one-try attacks," in *Proc. MFPS*, ser. ENTCS, vol. 249. Elsevier, 2009, pp. 75–91.

[4] P. S. Castro, P. Panangaden, and D. Precup, "Equivalence relations in fully and partially observable Markov Decision Processes," in *IJCAI'09 Proceedings of the 21st international jont conference on Artifical intelligence*, 2009, pp. 1653–1658.

[5] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden, "Probability of error in information-hiding protocols," in *Proc. CSF*. IEEE Computer Society, 2007, pp. 341–354.

[6] Y. Deng and W. Du, "The Kantorovich Metric in computer science: A brief survey," *Electron. Notes Theor. Comput. Sci.*, vol. 253, no. 3, pp. 73–82, Nov. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.entcs.2009.10.006

[7] D. Fremlin, *Measure Theory*. Torres Fremlin, 2000.

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

[9] J. Gray, "Toward a mathematical foundation for information flow security," in *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 1991, pp. 21–35.

[10] Probabilistic Systems Group, "Collected publications," http://www.cse.unsw.edu.au/∼carrollm/probs.

[11] J. He, K. Seidel, and A. McIver, "Probabilistic models for the guarded command language," *Science of Computer Programming*, vol. 28, pp. 171–92, 1997, first presented at FMTA '95, Warsaw.

[12] C. Jones and G. Plotkin, "A probabilistic powerdomain of evaluations," in *Proceedings of the IEEE 4th Annual Symposium on Logic in Computer Science*. Los Alamitos, Calif.: Computer Society Press, 1989, pp. 186–95.

[13] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, pp. 99–134, 1998.

[14] B. Köpf and D. Basin, "An information-theoretic model for adaptive side-channel attacks," in *Proceedings of the 14th ACM conference on Computer and communications security*, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 286–296. [Online]. Available: http://doi.acm.org/10.1145/1315245.1315282

[15] J. Massey, "Guessing and entropy," in *Proc. IEEE International Symposium on Information Theory*, 1994, p. 204.

[16] A. McIver, L. Meinicke, and C. Morgan, "Full version of this paper with appendices,," April 2012, available at [10].

[17] A. McIver and C. Morgan, "Partial correctness for probabilistic demonic programs," Oxford University, PRG Technical Report PRG-TR-35-97, 1997, available at [10].

[18] ——, *Abstraction, Refinement and Proof for Probabilistic Systems*, ser. Tech Mono Comp Sci. New York: Springer, 2005. [Online]. Available: http://www.cse.unsw.edu.au/ carrollm/arp/

[19] A. McIver, L. Meinicke, and C. Morgan, "Compositional closure for Bayes Risk in probabilistic noninterference," in *Proceedings of the 37th international colloquium conference on Automata, languages and programming: Part II*, ser. ICALP'10, vol. 6199. Berlin, Heidelberg: Springer, 2010, pp. 223–235, full version available at [20].

[20] ——, "Compositional closure for Bayes Risk in probabilistic noninterference," 2011, extended version of [19]. [Online]. Available: http://arxiv.org/pdf/1007.1054v1.pdf

[21] ——, "Hidden-Markov program algebra with iteration," 2011, at arXiv:1102.0333v1; to appear in Mathematical Structures in Computer Science in 2012.

[22] E. McShane, "Jensen's inequality," *Bull. Amer. Math. Soc.*, vol. 43, 521-527., no. 8, pp. 521–7, 1937.

[23] J. Millen, "Covert channel capacity," in *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 1987, pp. 60–66.

[24] M. Mislove, "Nondeterminism and probabilistic choice: Obeying the laws," in *In Proc. 11th CONCUR, volume 1877 of LNCS*. Springer, 2000, pp. 350–364.

[25] C. Morgan, A. McIver, and K. Seidel, "Probabilistic predicate transformers," *ACM Trans Prog Lang Sys*, vol. 18, no. 3, pp. 325–53, May 1996.

[26] B. Pfitzmann and M. Waidner, "A model for asynchronous reactive systems and its application to secure message transmission," in *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 2001, pp. 184–200.

[27] J. Pliam, "On the incomparability of entropy and marginal guesswork in brute-force attacks," in *Progress in Cryptology (INDOCRYPT 2000)*, ser. LNCS, vol. 1977. Springer, 2000, pp. 67–79.

[28] A. Sabelfeld and D. Sands, "A PER model of secure information flow in sequential programs," *Higher-Order and Symbolic Computation*, vol. 14, no. 1, pp. 59–91, 2001.

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

[30] G. Smith, "Adversaries and information leaks (Tutorial)," in *Proc. 3rd Symp. Trustworthy Global Computing*, ser. LNCS, G. Barthe and C. Fournet, Eds., vol. 4912. Springer, 2007, pp. 383–400.

[31] R. Tix, K. Keimel, and G. Plotkin, "Semantic domains for combining probability and non-determinism," *Electron. Notes Theor. Comput. Sci.*, vol. 222, pp. 3–99, Feb. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.entcs.2009.01.002

[32] K. Trustrum, *Linear Programming*, ser. Library of Mathematics. London: Routledge and Kegan Paul, 1971.

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

[34] F. van Breugel, S. Shalit, and J. Worrell, "Testing labelled Markov Processes," in *ICALP*, ser. LNCS, P. Widmayer, F. Ruiz, R. Bueno, M. Hennessy, S. Eidenbenz, and R. Conejo, Eds., vol. 2380. Springer, 2002, pp. 537–548.

[35] D. Varacca and G. Winskel, "Distributing probability over non-determinism," *Math Struct Comp Sci*, vol. 16, no. 1, pp. 87–113, 2006. [Online]. Available: http://www.pps.jussieu.fr/ varacca/papers/distrib.pdf

[36] J. Wittbold and D. Johnson, "Information flow in nondeterministic systems," in *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 1990, pp. 144–161.

[37] H. Yasuoka and T. Terauchi, "Quantitative information flow — verification hardness and possibilities," in *Proc. 23rd IEEE CSF Symp.*, 2010, pp. 15–27.

APPENDIX

A. Subprobability measures, monads and metrics

This section supports Footnote 4 in §III-A, referring to the work of van Breugel [33].

In his definition of the Kantorovich-metric monad $\mathbb{K}$, van Breugel represents subprobability measures on a given metric space $A{\in}\mathsf{KMet}_1$ by (full) probability measures on $\mathbf{1}{+}A$, which is A with an extra element $\perp$ at a distance 1 from all others. A subprobability measure $\alpha{:}\,\underline{\mathbb{K}}A$ is then represented by the (full) probability measure $\alpha_\perp{:}\,\mathbb{K}(\mathbf{1}{+}A)$ that allocates the "missing" $1{-}\alpha(A)$ to the extra point $\perp$, and he defines the distance between $\alpha, \alpha'{:}\,\underline{\mathbb{K}}A$ to be $\boldsymbol{k}(\alpha_\perp, \alpha'_\perp)$.

Here we verify that our direct definition at (1) of §III-A is equivalent, in fact the same as the definition over (full) probability measures.

Lemma 17: Direct definition of Kantorovich metric for submeasures For any metric space $A{:}\,\mathsf{KMet}_1$, and submeasures α, α' in $\underline{\mathbb{K}}A$, we have that $\boldsymbol{k}(\alpha_\perp, \alpha'_\perp)$, that is

$$(\sqcup f{:}\,A \xrightarrow{1} [0,1] \bullet |\textstyle\int f \ \mathrm{d}(\alpha_\perp) - \int f \ \mathrm{d}(\alpha'_\perp)|) \ ,$$

is equal to $(\sqcup f{:}\,A \xrightarrow{1} [0,1] \bullet |\int f \,\mathrm{d}\alpha - \int f \,\mathrm{d}\alpha'|)$, which is our definition of $\boldsymbol{k}_A(\alpha, \alpha')$ at §III-A(1).

Proof: We appeal to (1) and argue that in $\boldsymbol{k}(\alpha_\perp, \alpha'_\perp)$ we can assume *wlog* that the functions f used in its definition (1) satisfy $f(\perp){=}0$. First note that because of $\perp$'s isolation we can always set $f(\perp)$ either to 0 or 1 without decreasing $|\int f \ \mathrm{d}(\alpha_\perp) - \int f \ \mathrm{d}(\alpha'_\perp)|$; whether we choose 0 or 1 depends on the sign of $\int f \ \mathrm{d}(\alpha_\perp) - \int f \ \mathrm{d}(\alpha'_\perp)$. Then we observe that we can replace f by $1{-}f$ without affecting $|\int f \ \mathrm{d}(\alpha_\perp) - \int f \ \mathrm{d}(\alpha'_\perp)|$, so converting the "set $f(\perp)$ to 1" cases into "set $f(\perp)$ to 0" cases.

Finally, if $f(\perp){=}0$ we have $\int f \,\mathrm{d}\alpha_\perp = \int f \,\mathrm{d}\alpha$. □

B. Equivalent metrics for discrete metric spaces

Here we support our remarks in §III-B about the Kantorovich metric on a discrete space being effectively the Manhattan metric. We first revisit its duality with "earth-moving."

For any full probability measures α, α' on a separable metric space A, the Kantorovich-Rubinstein duality [6] establishes the equality of the Kantorovich metric (1) and the so-called "earth-mover's" metric, defined by

$$\boldsymbol{e}(\alpha, \alpha') \quad := \quad (\sqcap \mu{:}\,\alpha \lll \alpha' \bullet \textstyle\int \boldsymbol{a}(a, a') \ \mu(\,\mathrm{d}(a, a'))) \ ,$$

where $\boldsymbol{a}$ is the underlying metric on A and $\alpha \lll \alpha'$ is the set of all joint measures on the product space whose marginals are the given α and α'.

For subprobability measures there is also an earth-mover's duality:

Lemma 18: Kantorovich-Rubenstein duality for submeasures Given a separable metric space $A{=}(A, \boldsymbol{a})$, and any two subprobability measures α, α' in $\underline{\mathbb{K}}A$,

$$\text{If } \alpha(A){=}\alpha'(A) \text{ then } \boldsymbol{k}(\alpha, \alpha') = \boldsymbol{e}(\alpha, \alpha') \ . \tag{6}$$

More generally, assume *wlog* that $\alpha(A) \leq \alpha'(A)$. Then we have that $\boldsymbol{k}(\alpha, \alpha')$ is

$$\begin{array}{ll} & \alpha'(A) - \alpha(A) \\ + & (\sqcap \alpha'' \mid \alpha(A){=}\alpha''(A) \wedge \alpha''{\leq}\alpha' \bullet \boldsymbol{e}(\alpha, \alpha'')) \ , \end{array} \tag{7}$$

the second summand being the infimum over all α'' with $\alpha(A){=}\alpha''(A)$ and $\alpha''{\leq}\alpha'$ of the distance $\boldsymbol{e}(\alpha, \alpha'')$. That is, to use the earth-mover's definition on submeasures of differing weight, consider "intermediate" submeasures α'' of the same weight as α and everywhere no more than α'. Use earth-moving via (6) between α and α'', and then add the difference in weight between the original submeasures.

Proof: The proof of (6) is straightforward: simply use the original duality over $\mathbf{1}{+}A$, noting that the minimum earth-mover's value is obtained by leaving the earth on $\perp$ alone: its cost is therefore zero.

For (7) a similar argument applies: the minimising strategy over $\mathbf{1}{+}A$ will involve transfer of weight $\alpha'(A){-}\alpha(A)$ from $\alpha_\perp$'s point $\perp$ to proper elements of a in $\alpha'_\perp$, and not in the other direction (i.e. no transfers *to* $\perp$), since transfers both to- and from $\perp$ could not achieve minimality. A minimal transfer strategy of that kind is minimal over A as well (since minimisation distributes through addition); and the extra transfer from $\perp$ is accounted for by the addition of the term $\alpha'(A){-}\alpha(A)$. □

Lemma 19: Equivalent metrics in $\mathbb{K}$ for discrete metric spaces Any finite metric space Y equipped with the discrete metric is in KMet_1, and the Kantorovich distance induced on $\mathbb{K}Y$ is equal to half the Manhattan metric.

Proof: Take two (discrete) distributions $\delta_{\{1,2\}}$ in $\mathbb{K}Y$. From Lem. 18 their Kantorovich distance is $(\sum y{:}\,Y \mid \delta_1(y){>}\delta_2(y) \bullet \delta_1(y){-}\delta_2(y))$ since to reach δ_2 from δ_1 it's sufficient (and clearly optimal) to take all the δ_1 surfiets and move them to the δ_2 surfiets, and each movement is a distance 1 because the underlying metric of Y itself is discrete. And the total weight that moves cannot exceed the weight of δ_1 (or δ_2), also 1. Thus the distance is 1-bounded.

The Manhattan metric $(\sum y{:}\,Y \bullet |\delta_1(y){-}\delta_2(y)|)$ is trivially twice the above, since it sums both surfiets and deficits and they must be equal because $\sum \delta_1{=}\sum \delta_2{=}1$. (Although the Manhattan metric within a unit cube is bounded by its number of dimensions, here an upper bound of 2 applies because the distributions are 1-summing.) □

Note that the Kantorovich-, Manhattan- and Euclidean metrics all generate the same topology, and hence the same Borel algebra, and hence the same measurable space on $\mathbb{K}Y$ when Y is finite.

The same remarks apply for the submeasure spaces. For an example, suppose we're operating in a state-space $X{=}\{x, y\}$, so that the subdistributions in $\underline{\mathbb{D}}$ are the (non-negative) points satisfying $x{+}y \leq 1$. We're interested in the Kantorovich distance on $\underline{\mathbb{D}}$, where X has the discrete metric.

Following Breugel [33], we add an extra point z that soaks up the deficit in an (x, y), converting it to $(x, y, 1{-}x{-}y)$. Geometrically the subdistributions (x, y) are in a right triangle A in the xy plane, and their corresponding totalled

representative is the point (x, y, z) lying directly above, in the $x{+}y{+}z = 1$ plane which is the base B of the right-tetrahedron in 3-space. The (two-dimensional) Kantorovich dustance between (x, y) and (x', y') is then by definition the (three-dimensional) distance between (x, y, z) and (x', y', z') where z, z' have been generated as above.

Because there's a Kantorovich-continuous isomorphism between A and B, we know the Kantorovich-topology in A is the projection of the Kantorovich-topology in B. But we already know that the topology in B is (equivalently) Euclidean. Hence the topology in A is as well.

Corollary 20: Equivalent metrics in $\mathbb{K}$ for discrete metric spaces For finite discrete metric space Y and $\delta_{\{1,2\}}{:}\,\underline{\mathbb{K}}Y$, the Kantorovich metric $\boldsymbol{k}_{\underline{\mathbb{K}}Y}(\delta_1, \delta_2)$ is half of $(\sum y{:}\,Y \bullet |\delta_1(y){-}\delta_2(y)|) + |\sum\delta_1{-}\sum\delta_2|$.

Proof: Immediate from Lem. 19 and Lem. 17. □

C. Supporting material for §VI-D — Thm. 8

In the proof of Thm. 8 an appeal is made to Jensen's inequality at [22, Thm.2]. We use that reference to show $\int t\ \mathrm{d}\Delta \leq t(\boldsymbol{\mu}_X(\Delta))$ via these correspondences:

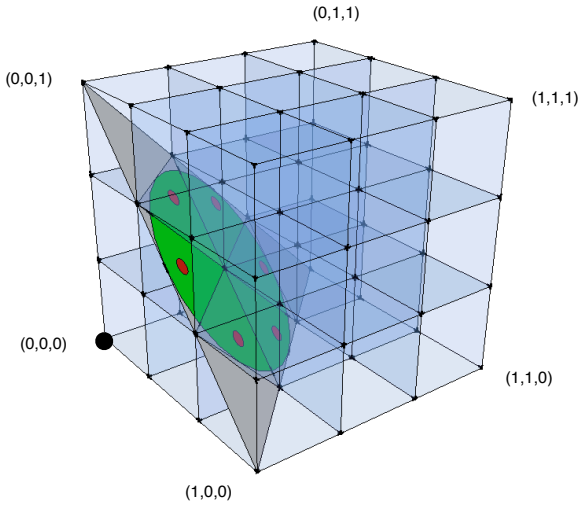
- Recall that our state-space X is finite; let N be its size. Then K from [22, Thm.2] is the closed, convex subset of $[0,1]^N$ lying on the one-summing plane: it is in fact $\mathbb{D}X$.
- Let $\phi{:}\,K{\to}\mathbb{R}$ be $-t$, continuous because t is and convex because t is concave.
- Let the class L be the measurable functions in $K{\to}\mathbb{R}$, trivially satisfying (1a) and (1b).
- Let $E{=}\mathbb{D}X$ (in fact the same as K) and define the functions $f_n{:}\,E{\to}\mathbb{R}$ to be the projection functions onto the n^{th} coordinate; note that they are in L. That means that $\mathbf{f}$ is (effectively) the identity function in $\mathbb{D}X{\to}[0,1]^N$.
- Fixing our hyper Δ, define the linear mean M so that in general $M(f) = \int f\,\mathrm{d}\Delta$. This satisfies (2b) and (2c) trivially, and (2a) because Δ is total, as in Thm. 8 it comes from the support of a super $\mathbf{\Delta}$. Because $\mathbf{f}$ is the identity, the condition that $x \mapsto \phi(\mathbf{f}(x))$ is in L amounts to $-t$ being measurable, which it is because it is continuous.
- The conclusion (J) of the theorem is then

$$-t\Big(\int \mathbf{f}\,\mathrm{d}\Delta\Big) \quad \leq \quad \int -t(\mathbf{f}(x))\,\mathrm{d}\Delta(x)\ ,$$

that is (since $\mathbf{f}$ is the identity) that $t(\boldsymbol{\mu}_X\Delta) \geq \int t\,\mathrm{d}\Delta$ as required.

D. Supporting material for §VIII-B — Thm. 16 Step 1

Here we elaborate the construction that in Step 1 of Thm. 16 allowed us to assume that Δ_2 had finite support. We relied on two things: that the set Z_1 from which we were hoping to separate Δ_2 was closed, and that Δ_2 itself was the Cauchy limit in the Kantorovich metric of a sequence of hypers Δ_2^ε each of which was within distance ε of Δ_2 and also satisfied $\Delta_2 \sqsubseteq_{\mathcal{S}} \Delta_2^\varepsilon$. Here we prove the existence of that limiting sequence; the construction is illustrated in Fig. 2.



The figure assumes a state space $X{=}\{x, y, z\}$ of size three, whose (full) distributions $\delta{\in}\mathbb{D}X$ are represented as points in the Euclidean unit cube $[0,1]^3$: they lie on the plane $x{+}y{+}z = 1$, here coloured grey, so that the x-coordinate of a distribution δ is the probability $\delta(x)$ it assigns to x, etc.

A (sub-)hyper in $\underline{\mathbb{K}}\mathbb{D}X{=}\underline{\mathbb{K}}\mathbb{K}(X, \mathsf{dis})$ on this space assigns a probability to each measurable subset of the cube, i.e. to sets of δ's, assigning no more than one to the whole cube. We fix a particular subhyper Δ, of total probability $W{\leq}1$ say, whose support is confined to the green circle lying in that grey plane. Deficit $1{-}W$ is the probability of Δ's diverging. For this example we assume that Δ is *uniform* on its support; note it is *not* discrete.

We divide the unit cube into cubies of side $1/3$, so that there are 27 of them; most do not intersect $\lceil\Delta\rceil$. Each of the six that does determines a "pie-wedge," and those six subhypers sum to the original subhyper Δ. Within each of those wedges we average the portion of Δ lying there to give the smaller red spot (drawn as a small circle so that it can be seen, but actually a single point).

Each red spot is a morsel, a point-subhyper having total probability $W/6$ (by symmetry) concentrated on a single δ. And each morsel is a $(\preceq)$-refinement, thus a $(\sqsubseteq_{\mathcal{S}})$-refinement of its wedge, the Kantorovich distance from that wedge being no more than the diameter of a cubie times the total probability $W/6$ of the wedge.

The sum of the six morsels is a single discrete approximant Δ^ε that $(\sqsubseteq_{\mathcal{S}})$-refines the sum of the six wedges, i.e. all of Δ, and whose Kantorovich distance from Δ is, by the earth-moving duality [6], no more than a cubie's diameter times the total $6{\times}W/6{=}W$ of the wedges' probabilities. Since the Manhattan diameter of a cubie is the dimension (3) times the side of the cubie $(1/3)$, that is 1, the Kantorovich diameter is half that (Lem. 19), thus $1/2$.

Hence Δ^ε is within Kantorovich distance $1/2$ of Δ. That distance can be decreased arbitrarily by taking smaller cubies, at a cost of more –but still finitely many– morsels.

Fig. 2. Illustration of discrete approximation, supporting the proof of Thm. 16 Step 1 in §VIII-B

Lemma 21: Approximating hypermeasures from above

For any finite set X and hypermeasure $\Delta{:}\,\underline{\mathbb{K}}\mathbb{D}X$ there is a structural refinement of it that is arbitrarily close to it and has finite support: for any $\varepsilon{>}0$ there is a hypermeasure Δ^ε with finite support such that $\Delta \sqsubseteq_{\mathcal{S}} \Delta^\varepsilon$ and $\boldsymbol{k}(\Delta, \Delta^\varepsilon) \leq \varepsilon$.

Proof: Let N be the size of X. Choose a constant K large enough that $\frac{N}{2K}{\leq}\varepsilon$, and partition the Euclidean cube $[0,1]^N$

into disjoint *cubies* each of side $1/K$ so that $\mathbf{C}_{k_1,\dots,k_N}$ is the cubie whose least vertex is the point $(k_1/K,\cdots,k_N/K)$; make them disjoint by assigning each shared vertex or edge to just one of its adjacent cubies. Thus the cubies will not be closed, but that does not matter: they need only be measurable. There are $M{:=}K^N$ cubies in all, and they cover the set of (full) distributions in $\mathbb{D}X$ thought of as points in $[0,1]^N$. The Kantorovich distance between any two points in the same cubie is no more than $N{\cdot}\frac{1}{K}/2{\leq}\varepsilon$ (Lem. 19).

The required secure refinement Δ^ε of Δ is then constructed by taking the cubies one-by-one and, for each, relativising Δ to that cubie so that $\Delta_{k_1,\dots,k_N}$ is the submeasure of Δ lying within $\mathbf{C}_{k_1,\dots,k_N}$ and having some weight $W_{k_1,\dots,k_N}$ over total distributions in $\mathbb{D}X$. Note that the sum of all those submeasures' weights equals the weight W of Δ itself. We then replace each submeasure $\Delta_{k_1,\dots,k_N}$ by a point submeasure, that is a submeasure whose support is a single distribution in $\mathbb{D}X$. It is concentrated on a single full discrete distribution $\delta_{k_1,\dots,k_N}$ that is the average of $\Delta_{k_1,\dots,k_N}$, but normalised. The weight assigned to the point is $W_{k_1,\dots,k_N}$, the same weight as the portion of $\Delta_{k_1,\dots,k_N}$ it came from. We call that point submeasure a *morsel*, and it is a secure refinement of $\Delta_{k_1,\dots,k_N}$.

The Δ^ε we seek is then the (finite, discrete) hyper obtained by adding up all the morsels, and its earth-moving distance from the original Δ cannot be more than ε times the sum of all morsels' weights, that is $\varepsilon W{\leq}\varepsilon$ since $W{\leq}1$.

The details are as follows. For each cubie $\mathbf{C}_{k_1,\dots,k_N}$ define the submeasure

$$\Delta_{k_1,\dots,k_N}(\mathbf{A}) \quad := \quad \Delta(\mathbf{A}\cap\mathbf{C}_{k_1,\dots,k_N})$$

and then construct that submeasure's average, a discrete subdistribution on X, as $\delta'_{k_1,\dots,k_N}{:=}\boldsymbol{\mu}\Delta_{k_1,\dots,k_N}$, where $\boldsymbol{\mu}$ is the $\mathbb{K}$-multiplier of the Kantorovich monad (§III). Determine the weight $W_{k_1,\dots,k_N}{=}\sum\delta'_{k_1,\dots,k_N}{=}\sum\Delta_{k_1,\dots,k_N}$, and use it to normalise to $\delta_{k_1,\dots,k_N}{:=}\delta'_{k_1,\dots,k_N}/W_{k_1,\dots,k_N}$, which is a point in $\mathbf{C}_{k_1,\dots,k_N}$.

The scaled-down point hyper $\{\!|\delta_{k_1,\dots,k_N}|\!\}{\times}W_{k_1,\dots,k_N}$, a *morsel*, is trivially a secure refinement of $\Delta_{k_1,\dots,k_N}$ because *all* of $\Delta_{k_1,\dots,k_N}$ has been merged; formally this is shown by taking $\mathbf{\Delta}_{k_1,\dots,k_N}{:=}\{\!|\Delta_{k_1,\dots,k_N}/W_{k_1,\dots,k_N}|\!\}{\times}W_{k_1,\dots,k_N}$ in Def. 6.[10] Since secure refinement is linear in its use of $\mathbf{\Delta}$, the single super that shows $\Delta{\sqsubseteq_{\mathcal{S}}}\Delta^\varepsilon$ is simply the sum of all the (finitely many) separate $\mathbf{\Delta}_{k_1,\dots,k_N}$'s.

It remains to bound the distance between Δ and Δ^ε. We have

$$\begin{array}{lll}
& \boldsymbol{k}(\Delta,\Delta_\varepsilon) & \\
= & (\sqcup f{:}\,\mathbb{D}X{\xrightarrow{1}}[0,1]\bullet|\int f\,\mathrm{d}\Delta-\int f\,\mathrm{d}\Delta^\varepsilon|) & \text{“definition }\boldsymbol{k}\text{”}\\
= & & \text{“sum over all cubies }\mathbf{C}\text{, covering }[0,1]^N\text{”}\\
& (\sqcup f{:}\,\mathbb{D}X{\xrightarrow{1}}[0,1]\bullet|(\textstyle\sum\mathbf{C}\bullet\int_{\mathbf{C}}f\,\mathrm{d}\Delta-\int_{\mathbf{C}}f\,\mathrm{d}\Delta^\varepsilon)|) & \\
\leq & (\textstyle\sum\mathbf{C}\bullet(\sqcup f{:}\,\mathbb{D}X{\xrightarrow{1}}[0,1]\bullet|\int_{\mathbf{C}}f\,\mathrm{d}\Delta-\int_{\mathbf{C}}f\,\mathrm{d}\Delta^\varepsilon|)) & \\
\leq & (\textstyle\sum\mathbf{C}\bullet\varepsilon\,W_{\mathbf{C}}) & \text{“}\boldsymbol{k}(\delta,\delta_{\mathbf{C}})\leq\varepsilon\text{ for all }\delta,\delta_{\mathbf{C}}{:}\,\mathbf{C}\text{”}\\
= & \varepsilon\,W & \\
\leq & \varepsilon\ . & \text{“Total weight }W\text{ of }\Delta\text{ is no more than 1”}
\end{array}$$

□

E. Supporting material for §VIII-B — Thm. 16 Step 2

Here we provide supporting material for the functions flatten and lift that take us back and forth between hypers and Euclidean space.

Any hyper $\Delta{:}\,\underline{\mathbb{K}}\mathbb{D}X$ (possibly a sub-measure) can be projected onto a *sub*-distribution in $\underline{\mathbb{D}}(X{\times}I)$ via an indexing map from $\mathbb{D}X$ to a finite set I. Distributions in $\underline{\mathbb{D}}(X{\times}I)$ can themselves be represented in the unit cube $[0,1]^{NM}$, where N is the size of X and M is the size of I.

If the support of Δ is finite, then I can be chosen large enough to make the indexing map an injection on that support; and, in that case, the original Δ can be recovered from its flatten'ing via lift (Def. 23 below).

For example, write $\{\!|x_1^{@p_1},\cdots,x_n^{@p_n}|\!\}$ for the finite (sub-)distribution assigning probability p_1 to x_1 etc. and take $x_{\{1,2\}}$ from a state-space X say of size 2. We could then represent the (sub-)hyper

$$\Delta \quad := \quad \{\!|\{\!|x_1^{@\frac12},x_2^{@\frac12}|\!\}^{@\frac14},\ \{\!|x_1|\!\}^{@\frac12}|\!\}$$

using a mapping

$$\{(\{\!|x_1^{@\frac12},x_2^{@\frac12}|\!\}\mapsto i_1),\ (\{\!|x_1|\!\}\mapsto i_2)\}$$

of distributions to indices $\{i_1,i_2\}$: it becomes the distribution

$$\{\!|(x_1,i_1)^{@\frac18},\ (x_2,i_1)^{@\frac18},\ (x_1,i_2)^{@\frac12}|\!\}\ ,$$

a point in $2N$ –that is 4-dimensional– Euclidean space: its coordinates are $(\frac18,\frac12,\frac18,0)$, the last one corresponding to x_2,i_2. Because the mapping is injective in this case, we can recover the original Δ from that point $(\frac18,\frac12,\frac18,0)$.

Definition 22: indexed flattening Write $(\xrightarrow{\mathsf{B}})$ for the Borel-measurable functions. Given a hyper $\Delta{:}\,\underline{\mathbb{K}}\mathbb{D}X$ and a measurable mapping $\iota{:}\,(\mathbb{D}X\xrightarrow{\mathsf{B}}I)$ of distributions to a finite set of indices in $I{:=}(I,\mathsf{dis})$ we define

$$\begin{array}{l}
\mathsf{flatten}{:}\ (\mathbb{D}X\xrightarrow{\mathsf{B}}I)\to\underline{\mathbb{K}}\mathbb{D}X\to\underline{\mathbb{D}}(X{\times}I)\\
\mathsf{flatten}_\iota(\Delta)(x,i)\ :=\ \int_{\iota^{-1}(i)}\delta(x)\ \Delta(\mathrm{d}\delta)\ .
\end{array}\tag{8}$$

□

Next we define a function lift that restores the hyper distribution structure from its encoding in terms of indices.

Definition 23: indexed lifting Given a measure $\delta{:}\,\underline{\mathbb{D}}(X{\times}I)$, we define the lifting of δ back to a hypermeasure in $\underline{\mathbb{K}}\mathbb{D}X$ by

$$\begin{array}{l}
\mathsf{lift}{:}\ \underline{\mathbb{K}}(X{\times}I)\to\underline{\mathbb{K}}\mathbb{D}X\\
\mathsf{lift}(\delta)\ :=\ \underline{\mathbb{K}}\delta_I(\delta_{\downarrow I})\ ,
\end{array}\tag{9}$$

[10] In general we have for arbitrary hyper $\Delta'{=}W{\times}\Delta$ with normalisation Δ and weight W that $\Delta'{\sqsubseteq_{\mathcal{S}}}W{\times}\{\!|\boldsymbol{\mu}\Delta|\!\}$ via $\mathbf{\Delta}{:=}W{\times}\{\!|\Delta|\!\}{=}W{\times}\boldsymbol{\eta}\Delta$

because $\boldsymbol{\mu}\mathbf{\Delta}{=}\boldsymbol{\mu}(W{\times}\boldsymbol{\eta}\Delta){=}W{\times}\boldsymbol{\mu}(\boldsymbol{\eta}\Delta){=}W{\times}\Delta{=}\Delta'$
and $\underline{\mathbb{K}}\boldsymbol{\mu}\mathbf{\Delta}{=}\underline{\mathbb{K}}\boldsymbol{\mu}(W{\times}\boldsymbol{\eta}\Delta){=}W{\times}\underline{\mathbb{K}}\boldsymbol{\mu}(\boldsymbol{\eta}\Delta){=}W{\times}\boldsymbol{\eta}(\boldsymbol{\mu}\Delta){=}W{\times}\{\!|\boldsymbol{\mu}\Delta|\!\}$.

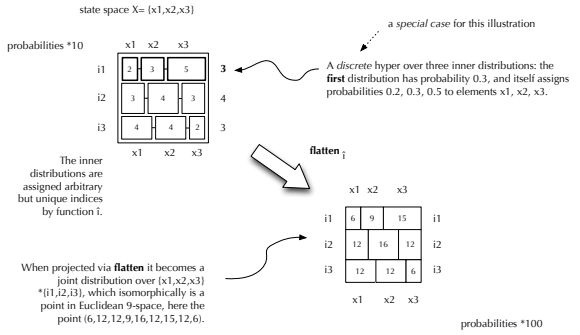


Fig. 3. Illustration of flatten

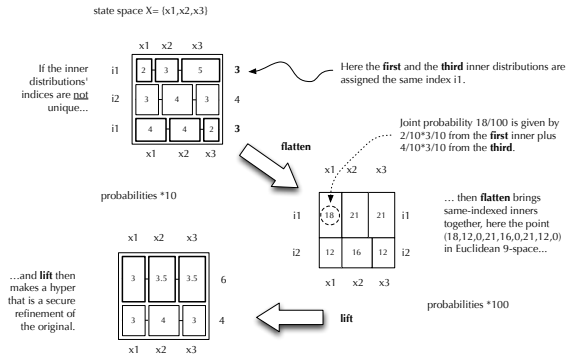


Fig. 4. Illustration of lift

The flatten function converts a measure over distributions on X into a joint sub-distribution on $X\times I$ for a finite indexing set I created specifically to give "names" to the interior distributions. Since both X and I are finite, the joint distribution can be considered to be a single point in finite-dimensional Euclidean space.

If flatten's argument is discrete and $\hat{\iota}$ is injective, e.g. as in §VIII-B(Step 2), then its action can be defined more directly as

$$\mathsf{flatten}_{\hat{\iota}}(\Delta)(x,i) \quad := \quad \Delta(\delta)\times\delta(x) \quad \text{for the } \delta \text{ st. } \hat{\iota}(\delta)=i\ .$$

In the proof of Lem. 31 (used in §VIII-B(Step 5)) however $\mathsf{flatten}_{\iota}$ is applied to measures, and with non-injective ι.

where $\delta_{\downarrow I}$ is the *right marginal* of δ and δ_I a the *right conditional* of δ, unique on the support of $\delta_{\downarrow I}$. □

Fig. 3 illustrates a discrete hyper (upper left), and its projection into finite-dimensional Euclidean space.

We now turn to function lift that takes us from Euclidean space back in to the hypers $\underline{\mathbb{K}}\mathbb{D}X$. It is illustrated for this discrete case in Fig. 4.

The following lemma gives us that any finitely-supported hyper $\Delta\colon\underline{\mathbb{K}}\mathbb{D}X$ can be recovered from its indexed flatten'ing $\delta\colon\underline{\mathbb{D}}(X\times I)$ providing the indexing is injective.

Lemma 24: For any finitely supported $\Delta\colon\underline{\mathbb{K}}\mathbb{D}X$ and mapping $\hat{\iota}\colon(\mathbb{D}X \overset{\mathsf{B}}{\to} I)$ of measures to indices that is injective over the support of Δ, we have that

$$\mathsf{lift}(\mathsf{flatten}_{\hat{\iota}}(\Delta)) \quad = \quad \Delta\ .$$

Proof: Fix hyper Δ and injective mapping $\hat{\iota}$, and for each $i\colon I$ set δ_i to be the unique distribution in the support of Δ such that $\hat{\iota}(\delta_i)=i$ so that

$$\Delta \quad = \quad (\sum i\colon I \bullet \Delta(\delta_i)\times\boldsymbol{\eta}_{\mathbb{D}X}(\delta_i))\ . \tag{10}$$

For $\delta:=\mathsf{flatten}_{\hat{\iota}}(\Delta)$ and $i\colon I$ we have that

$$\begin{array}{lll} & \delta_{\downarrow I}(i) & \\ = & \delta(X\times\{i\}) & \text{“marginal distribution definition”}\\ = & \Delta(\hat{\iota}^{-1}(i)) & \text{“}\mathsf{flatten}_{\hat{\iota}}\text{ (Def. 22)”}\\ = & \Delta(\delta_i) & \text{“}\hat{\iota}^{-1}(i)\cap\lceil\Delta\rceil=\{\delta_i\}\text{”}\end{array}$$

and

$$\begin{array}{lll} & \delta_I(i)(x) & \\ = & \frac{\delta(x,i)}{\delta_{\downarrow I}(i)} & \text{“discrete conditional distribution”}\\ = & \frac{\delta_i(x)\times\Delta(\delta_i)}{\delta_{\downarrow I}(i)} & \text{“}\mathsf{flatten}_{\hat{\iota}}\text{ for }\hat{\iota}^{-1}(i)\cap\lceil\Delta\rceil=\{\delta_i\}\text{”}\\ = & \delta_i(x) & \text{“}\delta_{\downarrow I}(i)=\Delta(\delta_i)\text{ (above)”}\end{array}$$

so that

$$\begin{array}{lll} & \mathsf{lift}(\mathsf{flatten}_{\hat{\iota}}(\Delta)) & \\ = & \underline{\mathbb{K}}\delta_I(\delta_{\downarrow I}) & \text{“lift (Def. 23)”}\\ = & (\sum i\colon I \bullet \delta_{\downarrow I}(i)\times\boldsymbol{\eta}_{\mathbb{D}X}(\delta_I(i))) & \text{“}\delta_{\downarrow I}\text{ is discrete”}\\ = & (\sum i\colon I \bullet \Delta(\delta_i)\times\boldsymbol{\eta}_{\mathbb{D}X}(\delta_i))\ . & \text{“(above)”}\\ = & \Delta\ . & \text{“(10)”}\end{array}$$

□

F. Supporting material for §VIII-B — Thm. 16 Step 3

This section supports the use of the flatten and lift functions in §VIII-B by detailing some of their further properties, including the continuity of lift (necessary to get a closed projection Z_1^E) and the $(\sqsubseteq_{\mathcal{S}})$-refining of a flatten followed by a lift that is important for showing that the separating hyperplane in Euclidean space gives a test in hyper space that also separates.

Lemma 25: lift *is continuous* Suppose we have a Cauchy sequence $\{\delta_n\}_n$ in the Euclidean space E that is isomorphic to $\underline{\mathbb{D}}(X\times I)$, and that it converges to some δ in the Euclidean metric there. We must show that $\{\mathsf{lift}(\delta_n)\}_n$ converges to $\mathsf{lift}(\delta)$ in $\underline{\mathbb{K}}\mathbb{D}X$ wrt. the Kantorovich metric on the latter. It may be calculated that $\boldsymbol{k}(\mathsf{lift}(\delta),\mathsf{lift}(\delta_n))$ is no more than

$$(\sum(x,i)\colon X\times I \bullet |\delta_n(x,i)-\delta(x,i)|)+|\textstyle\sum\delta-\sum\delta'|\ . \tag{11}$$

Since $X\times I$ is finite, the convergence of δ_n to δ implies uniform convergence in n of the individual probabilities $\delta_n(x,i)$ to $\delta(x,i)$ in $\mathbb{R}$, and so (11) itself tends to zero trivially, establishing the convergence in $\underline{\mathbb{K}}\mathbb{D}X$. □

The super-linearity of lift is used to establish that Z_1^E is convex, necessary for the Separating-Hyperplane Lemma.

Lemma 26: lift *is super-linear* For all constants $c_{\{1,2\}}{:}\,\mathbb{R}^{\geq}$ with $c_1{+}c_2{\leq}1$ and $\delta_{\{1,2\}}{:}\,\underline{\mathbb{D}}(X{\times}I)$,

$$c_1{\times}\mathsf{lift}(\delta^1){+}c_2{\times}\mathsf{lift}(\delta^2) \sqsubseteq_{\mathcal{S}} \mathsf{lift}(c_1{\times}\delta^1{+}c_2{\times}\delta^2)\ , \qquad (12)$$

holds.

Proof: Let $\delta{:=}\,c_1{\times}\delta^1{+}c_2{\times}\delta^2$. To verify the structural refinement (Def. 5) in (12) we construct super

$$\begin{array}{rcl} \blacktriangle & := & \underline{\mathbb{K}}(\boldsymbol{\lambda} i{:}\,I \bullet {\delta^1}_I(i)\ {}_{\frac{c_1\times\delta^1{}_{\downarrow I}(i)}{\delta_{\downarrow I}(i)}}\!\oplus \delta^2{}_I(i))(\delta_{\downarrow I}) \\ & = & (\sum i{:}\,I \bullet \delta_{\downarrow I}(i) \times \boldsymbol{\eta}_{\underline{\mathbb{K}}\mathbb{D}X}({\delta^1}_I(i)\ {}_{\frac{c_1\times\delta^1{}_{\downarrow I}(i)}{\delta_{\downarrow I}(i)}}\!\oplus \delta^2{}_I(i))) \end{array}$$

and show that

$$\begin{array}{rcll} & \boldsymbol{\mu}_{\mathbb{D}X}(\blacktriangle) & & \\ = & & \text{“}\boldsymbol{\mu}_{\mathbb{D}X}\text{ applied to discrete }\blacktriangle\text{”} & \\ & (\sum i{:}\,I \bullet \delta_{\downarrow I}(i) \times ({\delta^1}_I(i)\ {}_{\frac{c_1\times\delta^1{}_{\downarrow I}(i)}{\delta_{\downarrow I}(i)}}\!\oplus \delta^2{}_I(i))) & & \\ = & c_1 \times (\sum i{:}\,I \bullet {\delta^1}_{\downarrow I}(i) \times \boldsymbol{\eta}_{\mathbb{D}X}({\delta^1}_I(i))) + & & \text{“simplify”} \\ & c_2 \times (\sum i{:}\,I \bullet {\delta^2}_{\downarrow I}(i) \times \boldsymbol{\eta}_{\mathbb{D}X}({\delta^2}_I(i))) & & \\ = & c_1 \times \underline{\mathbb{K}}({\delta^1}_I)({\delta^1}_{\downarrow I}) + c_2 \times \underline{\mathbb{K}}({\delta^2}_I)({\delta^2}_{\downarrow I}) & & \text{“}\underline{\mathbb{K}}\text{”} \\ = & c_1 \times \mathsf{lift}(\delta^1) + c_2 \times \mathsf{lift}(\delta^2) & & \text{“lift (Def. 23)”} \end{array}$$

and

$$\begin{array}{rcll} & \underline{\mathbb{K}}\boldsymbol{\mu}_X(\blacktriangle) & & \\ = & & \text{“}\blacktriangle\text{ and functor composition”} & \\ & \underline{\mathbb{K}}(\boldsymbol{\mu}_X \circ (\boldsymbol{\lambda} i{:}\,I \bullet {\delta^1}_I(i)\ {}_{\frac{c_1\times\delta^1{}_{\downarrow I}(i)}{\delta_{\downarrow I}(i)}}\!\oplus \delta^2{}_I(i)))(\delta_{\downarrow I}) & & \\ = & & \text{“}\boldsymbol{\mu}_X\text{ applied to discrete distribution”} & \\ & \underline{\mathbb{K}}(\boldsymbol{\lambda} i{:}\,I \bullet \tfrac{c_1\times\delta^1{}_{\downarrow I}(i)}{\delta_{\downarrow I}(i)} \times {\delta^1}_I(i) + \tfrac{c_2\times\delta^2{}_{\downarrow I}(i)}{\delta_{\downarrow I}(i)} \times {\delta^2}_I(i))(\delta_{\downarrow I}) & & \\ = & \underline{\mathbb{K}}(\delta_I)(\delta_{\downarrow I}) & & \text{“simplify using }\delta = c_1{\times}\delta^1{+}c_2{\times}\delta^2\text{”} \\ = & \mathsf{lift}(\delta)\ . & & \text{“lift (Def. 23)”} \end{array}$$

□

G. Supporting material for §VIII-B — Thm. 16 Step 4

Here we justify the claim in that a separating hyperplane can be chosen with only non-negative coefficients.

Lemma 27: Non-negative hyperplane

In §VIII-B(Step 4) we noted that the separating hyperplane can always be chosen with non-negative coefficients. Informally that is because we are separating a point p_2^E from a set Z_1^E that is *up-closed* in Euclidean space — i.e. it is the up-closure of Z_1^E that introduces the asymmetry guaranteeing non-negative coefficients for H.

Proof: Assume that there exists a hyperplane with normal $H{:}\,X{\times}I \to \mathbb{R}$ such that $\int H\ \mathrm{d}p_2^E < (\sqcap p_1^E{:}\,Z_1^E \bullet \int H\ \mathrm{d}p_1^E)$, and assume that H is somewhere negative. Define $H'{:=}\,H{+}c$ where c is the smallest value making H' everywhere non-negative. We argue that

$$\begin{array}{lrcl} & (\sqcap p_1^E{:}\,Z_1^E \bullet \int H'\ \mathrm{d}p_1^E) & = & (\sqcap p_1^E{:}\,Z_1^E \bullet \int H\ \mathrm{d}p_1^E) + c\ , \\ \text{but} & \int H'\ \mathrm{d}p_2^E & \leq & \int H\ \mathrm{d}p_2^E + c\ , \end{array}$$

so establishing that H' will suffice.

The inequality is obvious because p_2^E sums to no more than one. The equality holds because both minima are attained on the one-summing subset of Z_1^E, due to its $\leq$-closure: any minimising point can be made one-summing by increasing its coordinate where H, H' are zero or negative, and that cannot increase the integral. □

H. Supporting material for §VIII-B — Thm. 16 Step 5

The following lemmas support the crucial properties of lift and flatten together, used in Step 5 of §VIII-B.

Lemma 28: First property from §VIII-B(Step 5)

Proof: If the definition of flatten from Step 2 is specialised to discrete distributions (as Δ_2 indeed is), it becomes

$$\mathsf{flatten}_i(\Delta_2)(x,i) \quad := \quad (\sum \delta{:}\,\lceil\Delta_2\rceil \mid \hat{\imath}(\delta){=}i \bullet \Delta_2(\delta){\times}\delta(x))$$

Then we calculate

$$\begin{array}{rcl} & \int H\ \mathrm{d}(\mathsf{flatten}_i(\Delta_2)) & \\ = & (\sum x, i \bullet H(x,i) \times \mathsf{flatten}_i(\Delta_2)(x,i)) & \\ = & (\sum x, i, \delta \mid \hat{\imath}(\delta) = i \bullet H(x,i) \times \Delta_2(\delta) \times \delta(x)) & \\ = & (\sum \delta \bullet \Delta_2(\delta) \times (\sum x \bullet H(x,i) \times \delta(x))) & \\ \geq & (\sum \delta \bullet \Delta_2(\delta) \times (\sqcap i \bullet (\sum x \bullet H(x,i) \times \delta(x)))) & \\ = & (\sum \delta \bullet \Delta_2(\delta) \times (\sqcap i \bullet \int H(x,i)\,\delta(\,\mathrm{d}x))) & \\ = & (\sum \delta \bullet \Delta_2(\delta) \times t(\delta)) & \\ = & \int t\ \mathrm{d}\Delta_2\ . & \end{array}$$

□

The second property is given as a sequence of three lemmas.

We begin with an important connection between flatten and lift and $(\sqsubseteq_{\mathcal{S}})$-refinement. In the discrete case it is illustrated Fig. 4.

Lemma 29: flatten*'ing then* lift*'ing induces a structural refinement*

For arbitrary hypermeasure Δ, finite metric space $I{:=}\,(I, \mathsf{dis})$, and Borel-measurable mapping $\iota{:}\,\mathbb{D}X{\to}I$, we have that $\Delta \sqsubseteq_{\mathcal{S}} \mathsf{lift}(\mathsf{flatten}_\iota(\Delta))$.

Proof: Fix Δ and ι and let $\delta{:=}\,\mathsf{flatten}_\iota(\Delta)$. For each $i{:}\,I$ such that $\delta_{\downarrow I}(i) \neq 0$ we define hyper $\Delta_i \in \underline{\mathbb{K}}\mathbb{D}X$ by

$$\Delta_i(\mathbf{A}) \quad := \quad \tfrac{\Delta(\iota^{-1}(i)\cap\mathbf{A})}{\delta_{\downarrow I}(i)} \qquad \begin{array}{l}\text{for each } \mathbf{A} \text{ in the} \\ \text{Borel-algebra of } \mathbb{D}X\end{array}$$

so that $\boldsymbol{\mu}_X(\Delta_i) = \delta_I(i)$. For i such that $\delta_{\downarrow I}(i) = 0$ we set Δ_i arbtitrarily. Using these I-indexed hypers we construct super

$$\blacktriangle \quad := \quad \underline{\mathbb{K}}(\boldsymbol{\lambda} i \bullet \Delta_i)\delta_{\downarrow I} \quad = \quad (\sum i{:}\,I \bullet \delta_{\downarrow I}(i) \times \boldsymbol{\eta}_{\mathbb{K}^2X}\Delta_i)\ ,$$

and show that

$$\begin{array}{rcll} & (\boldsymbol{\mu}_{\mathbb{D}X}(\blacktriangle))(\mathbf{A}) & & \\ = & & \text{“apply }\boldsymbol{\mu}_{\mathbb{D}X}\text{ to discrete }\blacktriangle\text{”} & \\ & (\sum i{:}\,I \bullet \delta_{\downarrow I}(i) \times \boldsymbol{\mu}_{\mathbb{D}X}(\boldsymbol{\eta}_{\mathbb{K}^2X}\Delta_i))(\mathbf{A}) & & \end{array}$$

$= \quad (\sum i{:}\, I \bullet \delta_{\downarrow I}(i) \times \Delta_i(\mathbf{A}))$ "$\boldsymbol{\mu}_{\mathbb{D}X} \circ \boldsymbol{\eta}_{\mathbb{K}^2 X} = 1$"
$= \quad (\sum i{:}\, I \bullet \Delta(\iota^{-1}(i) \cap \mathbf{A}))$ "simplify using Δ_i"
$= \quad \Delta(\mathbf{A})$ "countable additivity"

for all $\mathbf{A}$ in the Borel-algebra of $\mathbb{D}X$, and

$\underline{\mathbb{K}}\boldsymbol{\mu}_X(\boldsymbol{\Delta})$
$= \quad (\underline{\mathbb{K}}\boldsymbol{\mu}_X \circ \underline{\mathbb{K}}(\boldsymbol{\lambda} i \bullet \Delta_i))(\delta_{\downarrow I})$ "$\boldsymbol{\Delta}$"
$= \quad \underline{\mathbb{K}}(\boldsymbol{\lambda} i \bullet \boldsymbol{\mu}_X \Delta_i)(\delta_{\downarrow I})$ "functor composition"
$= \quad \underline{\mathbb{K}}\delta_I(\delta_{\downarrow I})$ "$\boldsymbol{\mu}_X \Delta_i = \delta_I(i)$ for $\delta_{\downarrow I}(i) \neq 0$"
$= \quad \mathsf{lift}(\delta)$. "lift (Def. 23))"

This demonstrates that Δ is securely refined (Def. 6), and hence structurally refined (Def. 5), by $\mathsf{lift}(\delta)$. □

Next we introduce a technical lemma relating the test t in the hyper space and the hyperplane H in Euclidean space.

Lemma 30: Given a hyper $\Delta{:}\,\underline{\mathbb{K}}\mathbb{D}X$ and a mapping $\iota{:}\,(\mathbb{D}X \overset{\mathsf{B}}{\to} I)$ such that $\lceil\Delta\rceil \subseteq \iota^{-1}(I)$, we have that for any hyperplane normal $H{:}\,X{\times}I \to \mathbb{R}$ we can relate its effect in the Euclidean space to its effect in the hyper space via

$$\int H\,\mathrm{d}(\mathsf{flatten}_\iota(\Delta)) \quad = \quad \int\!\!\int H(x, \iota(\delta))\ \delta(\mathrm{d}x)\ \Delta(\mathrm{d}\delta)\ .$$

Proof:

$\int H\,\mathrm{d}(\mathsf{flatten}_\iota(\Delta))$
$= \quad (\sum (x,i) \bullet H(x,i) \times \mathsf{flatten}_\iota(\Delta)(x,i))$ "discrete integration"
$= \quad (\sum (x,i) \bullet H(x,i) \times (\int_{\iota^{-1}(i)} \delta(x)\ \Delta(\mathrm{d}\delta)))$ "flatten (Def. 22)"
$= \quad (\sum i \bullet (\int_{\iota^{-1}(i)} (\sum x \bullet H(x, \iota(\delta)) \times \delta(x))\ \Delta(\mathrm{d}\delta)))$ "linearity of integration"
$= \quad (\int (\sum x \bullet H(x, \iota(\delta)) \times \delta(x))\ \Delta(\mathrm{d}\delta))$ "countable additivity"
$= \quad \int\!\!\int H(x, \iota(\delta))\ \delta(\mathrm{d}x)\ \Delta(\mathrm{d}\delta)\ .$ "discrete integration"

□

Finally we give the second crucial property of the flatten-then-lift combination.

Lemma 31: Second property from §VIII-B(Step 5)

In the terminology of Step 5 and using the t, H defined there, for any Δ in Z_1 we have $\int t\ \mathrm{d}\Delta = \int H\ \ \mathrm{d}(p_1^E)$ for some $p_1^E \in Z_1^E$.

Proof: Suppose we have our index set I and a hyperplane normal H; note that H is a function of pairs (x,i) with x coming from the space X and i an index from I. For notational convenience define now $H_i(x) := H(x,i)$ so that our test $t = (\sqcap i{:}\, I \bullet \int H(x,i)\ \delta(\mathrm{d}x))$ can equivalently be written $(\sqcap i{:}\, I \bullet \int H_i\ \mathrm{d}\delta)$. Seen this way H is an I-indexed collection of individual H_i's and to apply the induced test t to some particular δ in $\lceil\Delta\rceil$ you choose the H_i that gives the least outcome for that δ, and t applied to the whole of Δ is the integral over Δ of that procedure.

Define measurable function $\iota{:}\,\mathbb{D}X \to I$ by choosing for each $\delta{:}\,\mathbb{D}X$ an $\int H_i\,\mathrm{d}\delta$ -minimising i.[11]

Unlike the arbitrary 1-1 map $\hat{\iota}$ used earlier for Δ_2, this ι is not injective in general. Using $\mathsf{flatten}$ and the H-minimising ι we define $p_1^E := \mathsf{flatten}_\iota(\Delta)$, which satisfies $\int t\ \mathrm{d}\Delta = \int H\,\mathrm{d}p_1^E$ (from Lem. 30 and the fact that we chose ι to minimise).

Now define $\Delta_1 := \mathsf{lift}(p_1^E)$, and observe that it is a $(\sqsubseteq_{\mathcal{S}})$-refinement of Δ by Lem. 29: effectively any non-injectiveness of ι has merged inners in $\lceil\Delta\rceil$ just as $(\sqsubseteq_{\mathcal{S}})$-refinement does, as illustrated in Fig. 4 for the discrete case. By $(\sqsubseteq_{\mathcal{D}})$-closure of Z_1, hence $\sqsubseteq_{\mathcal{S}}$-closure (Thm. 8), we then have $\Delta_1 \in Z_1$, and thus $p_1^E \in Z_1^E$ is the point we seek. □

I. Anti-refinements of POMDP's example: a discrete chain with a continuous limit

In this section we give some background for the example in §IX-E. Write $\{\!\{x_1, x_2, \cdots, x_k\}\!\}$ when $k{>}0$ for the uniform distribution over the finitely many distinct values $x_1, x_2, \cdots, x_k$, of which the point distribution $\{\!\{x\}\!\}$ is a special case.

In §IX-E we referred to a *POMDP* P_n over state $X{=}\{a,b\}$ with labels in $\{1 \ldots 2^n{-}1\}$, and introduced the abbreviation $\delta_p{=}a_p{\oplus}b$. The *POMDP* was to have the property that with label l it takes an incoming belief state $\delta_{l/2^n}$ to outgoing hyper $\{\!\{\delta_{(2l-1)/2^{n+1}}, \delta_{(2l+1)/2^{n+1}}\}\!\}$, a fair choice between two outgoing belief states each a small perturbation of the incoming belief state. Existence of such a P_n is shown further below. Here we use the P_n's to demonstrate how being able to take limits of *POMDP*-chains themselves (i.e. in effect as a limit *POMDP*) is useful for bounding the security properties of elements in the chain.

We start with belief state $\delta_{\frac{1}{2}}$, and apply P_1 with label $l{=}1$. With probability $1/2$ we observe F, say, and move to belief state $\delta_{\frac{1}{4}}$; with probability $1/2$ we observe T and move to $\delta_{\frac{3}{4}}$.

Suppose we observed T on that first step, leading to $\delta_{\frac{3}{4}}$. In our second step we apply P_2 with label $l{=}3$ and the two equally likely observations (again T, F) are summarised as the hyper $\{\!\{\delta_{\frac{5}{8}}, \delta_{\frac{7}{8}}\}\!\}$.

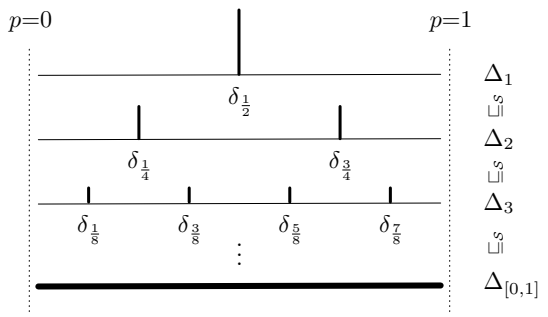
Suppose we observed F on that second step. Our belief state would now be $\delta_{\frac{5}{8}}$, and we would assign it the probability of the two observations in succession: the sequence $\langle \mathsf{T}, \mathsf{F} \rangle$ that occurs with probability $1/4$. If we abstract from the sequence(s) but retain the effects, we record the result as the hyper $\{\!\{\delta_{\frac{1}{8}}, \delta_{\frac{3}{8}}, \delta_{\frac{5}{8}}, \delta_{\frac{7}{8}}\}\!\}$.

A system programmed to implement that l-strategy, i.e. choosing l at each stage on the basis of all observations so far, would operate from hyper-to-hyper and would thus generate successively

$$\begin{array}{lcl} \Delta_1 & := & \{\!\{\delta_{1/2}\}\!\} \\ \Delta_2 & := & \{\!\{\delta_{1/4}, \delta_{3/4}\}\!\} \\ \Delta_3 & := & \{\!\{\delta_{1/8}, \delta_{3/8}, \delta_{5/8}, \delta_{7/8}\}\!\} \quad \cdots \end{array} \tag{13}$$

in which, as we saw, in Δ_3 the essential information is given without having to retain the observations themselves: belief

[11] Measurability of ι follows from continuity in δ of $\int H_i\,\mathrm{d}\delta$ and its preservation under finite minima.



Distribution δ_p is $a_p \oplus b$ over state space $X=\{a,b\}$.

Consider factories, of varying quality, using a physical process to manufacture oscillators of fixed but not wholly predictable ratios between the two states a, b. Factory Δ_1 is the highest quality, guaranteed (with probability 1) to produce oscillators that alternate perfectly fairly between a and b; Factory Δ_2, of slightly lower quality, produces with equal probability either oscillators dwelling three times as much at a as b or the reverse, three times as much at b as at a. Factory $\Delta_{[0,1]}$ of lowest quality produces oscillators whose whose fixed ratio p is distributed uniformly over the entire interval $[0,1]$.

All factories test their oscillators before delivery and label them with their fixed ratio p, whatever it turns out to be; thus p is known (afterwards) but not predictable (beforehand). An attacker defeats a system containing such an oscillator if at a given moment he can guess whether it's at a or b. For this he looks at the label p to decide what the best guess should be.

For Factory Δ_1's oscillators, a single guess of the attacker succeeds with probability $1/2$. For Factory Δ_2, the attacker reads the label and then guesses a if the label says $3/4$ and b if it says $1/4$; he succeeds with probability $3/4$. For Factory $\Delta_{[0,1]}$ he succeeds with probability $2\int_0^{\frac{1}{2}}(1-p)\,\mathrm{d}p = 3/4$.

Calculation shows that whereas the Bayes Risk stabilises at n=2, the Shannon Entropy continues strictly to decrease. The former tells a Bayes-Risk attacker that there is no point in proceeding beyond two iterations; the latter tells a Shannon-Entropy *defender* that no matter how many iterations are taken, his vulnerability is bounded (and gives that bound).

Fig. 5. A discrete-hyper anti-chain with continuous infimum

state $\delta_{5/8}$ is one of four elements of a uniform distribution, and thus occurs with probability $1/4$ as noted above.

Because each application of $P_{(\bullet)}$ releases (more) information, the sequence (13) is a $(\sqsubseteq_{\mathcal{S}})$-anti-chain $\Delta_1 \sqsupseteq_{\mathcal{S}} \Delta_2 \sqsupseteq_{\mathcal{S}} \cdots$, and the finite support is unbounded in size, doubling at each step; its infimum is the continuous hyper $\Delta_{[0,1]}$ in $\mathbb{K}\mathbb{D}X$ where we write $\Delta_{[q,r]}$ for the uniform hyper whose support is the set of belief states δ_p with $q \leq p \leq r$. This is illustrated in Fig. 5. [12]

For example (refer Def. 6) the refinement $\Delta_3 \sqsubseteq_{\mathcal{S}} \Delta_2$ is shown by the super $\{\!\{\{\!\{\delta_{1/8}, \delta_{3/8}\}\!\}, \{\!\{\delta_{5/8}, \delta_{7/8}\}\!\}\}\!\}$, and the refinement $\Delta_{[0,1]} \sqsubseteq_{\mathcal{S}} \Delta_3$ is shown by the super $\{\!\{\Delta_{[0,\frac{1}{4}]}, \Delta_{[\frac{1}{4},\frac{1}{2}]}, \Delta_{[\frac{1}{2},\frac{3}{4}]}, \Delta_{[\frac{3}{4},1]}\}\!\}$.

Finally, for completeness, we establish the existence of such P_n's by exhibiting the channel (stochastic) matrix for a particular P_n and label l: it is

$$\begin{array}{c|cc|c} & \mathsf{F} & \mathsf{T} & \\ \hline a & \frac{2l-1}{4l} & \frac{2l+1}{4l} & \sum\!=\!1 \\ b & \frac{(2^{n-1}-\frac{2l-1}{4})}{(2^n-l)} & \frac{(2^{n-1}-\frac{2l+1}{4})}{(2^n-l)} & \sum\!=\!1 \end{array} .$$

Multiplied through by incoming distribution (belief state) $\delta_{l/2^n}$ gives this joint distribution on $X \times \{\mathsf{F},\mathsf{T}\}$:

$$\delta_{l/2^n} \left\{ \begin{array}{c|cc} & \mathsf{F} & \mathsf{T} \\ \hline a & (2l-1)/2^{n+2} & (2l+1)/2^{n+2} \\ b & \frac{1}{2}-(2l-1)/2^{n+2} & \frac{1}{2}-(2l+1)/2^{n+2} \end{array} \right. .$$

Since each column sums to $1/2$, the probability of observing F,T is $1/2$ for each, and normalising those columns gives the resulting conditional belief states, that is

$$\begin{array}{c|c\|c} & \mathsf{F} & \mathsf{T} \\ \hline a & (2l-1)/2^{n+1} & (2l+1)/2^{n+1} \\ b & 1-(2l-1)/2^{n+1} & 1-(2l+1)/2^{n+1} \end{array} ,$$

where the first,second columns are then $\delta_{(2l-1)/2^{n+1}}$, $\delta_{(2l+1)/2^{n+1}}$ resp. as claimed.

[12] An example of a $(\sqsubseteq_{\mathcal{S}})-$*ascending* discrete chain with a continuous limit is given in [21, Sec. 6.1].