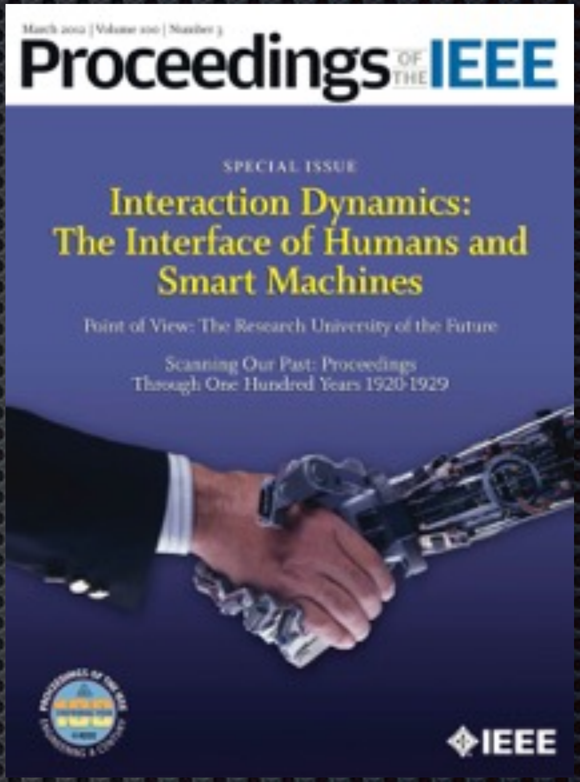
Topological Aspects of Optimal Information Acquisition

John Baillieul
Intelligent Mechatronics Lab
Boston University
johnb@bu.edu

10-th Summer School on Geometry,
Mechanics and Control June, 2016

Talk Texts:



“Decision Making for Rapid Information Acquisition in the Reconnaissance of Random Fields,” D. Baronov and J.B., *Proceedings of the IEEE*, March 2012.

“A Motion Description Language for Robotic Reconnaissance of Unknown Fields,'' D. Baronov and J.B., *European J. Control*, Sept.-Dec. 2011,Vol. 17:5-6, pp. 512-525. DOI:10:3166/EJC.17.512-525.

H. Edelsbrunner, D. Letscher and A. Zomorodian. “Topological persistence and simplification,” *Discrete Comput. Geom.* 28 (2002), 511-533.

H. Edelsbrunner and J. Harer, *Computational Topology. An Introduction,* Amer. Math. Soc., Providence, Rhode Island, 2010.

Talk Outline:

- Introduction - why did we begin to think about topology and information?
- The information theory of functions
- Motion primitives for robotic reconnaissance
- Reconnaissance as information acquisition
- The topology of unknown fields
- Data induced partitions and topology induced partitions
- Topology guided information acquisition

What is information?

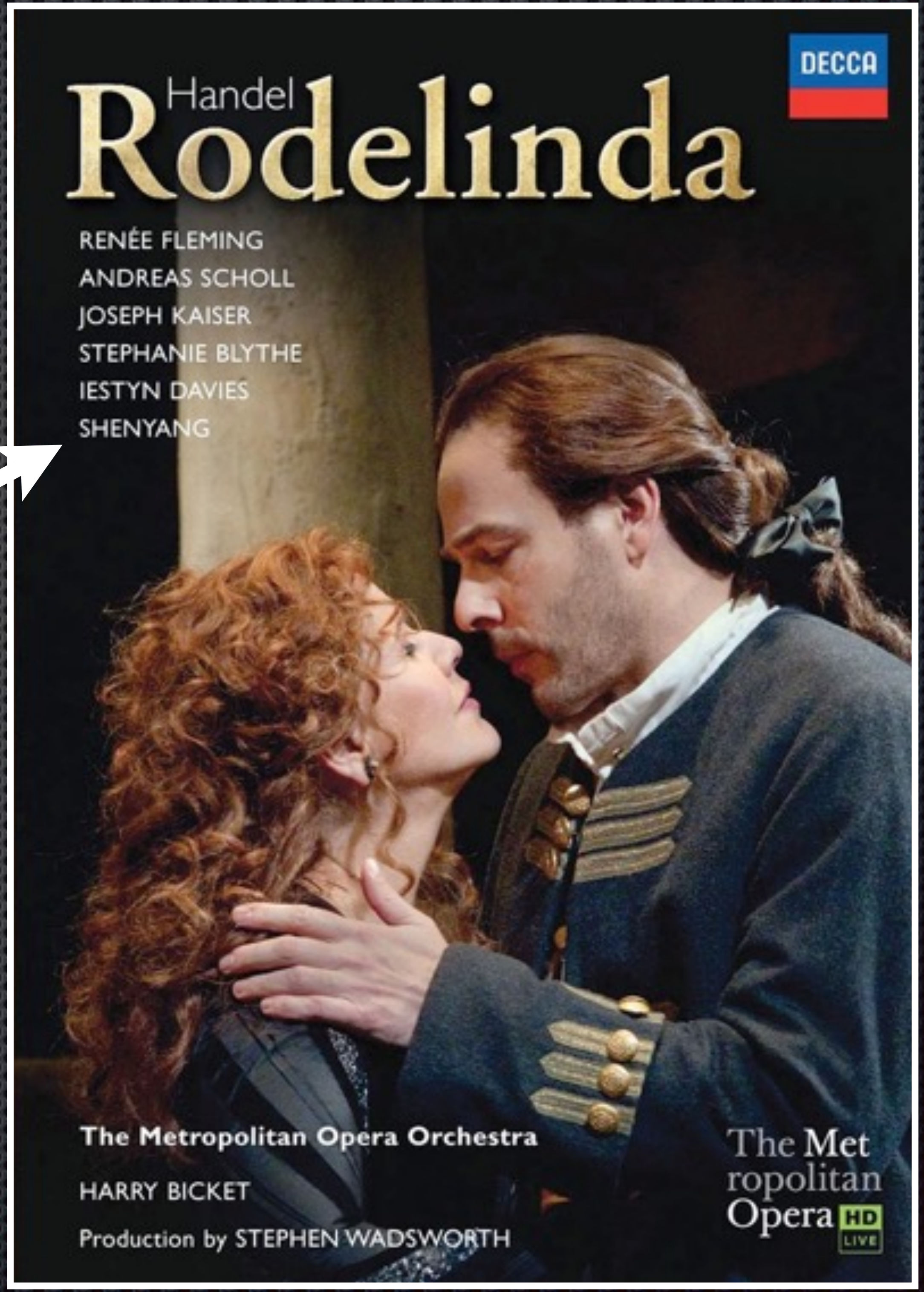
What is information?

The number of “yes-no” questions that must be answered to answer a given question.

What is information?

Typical (simple) question decomposition:

Where is Shenyang?

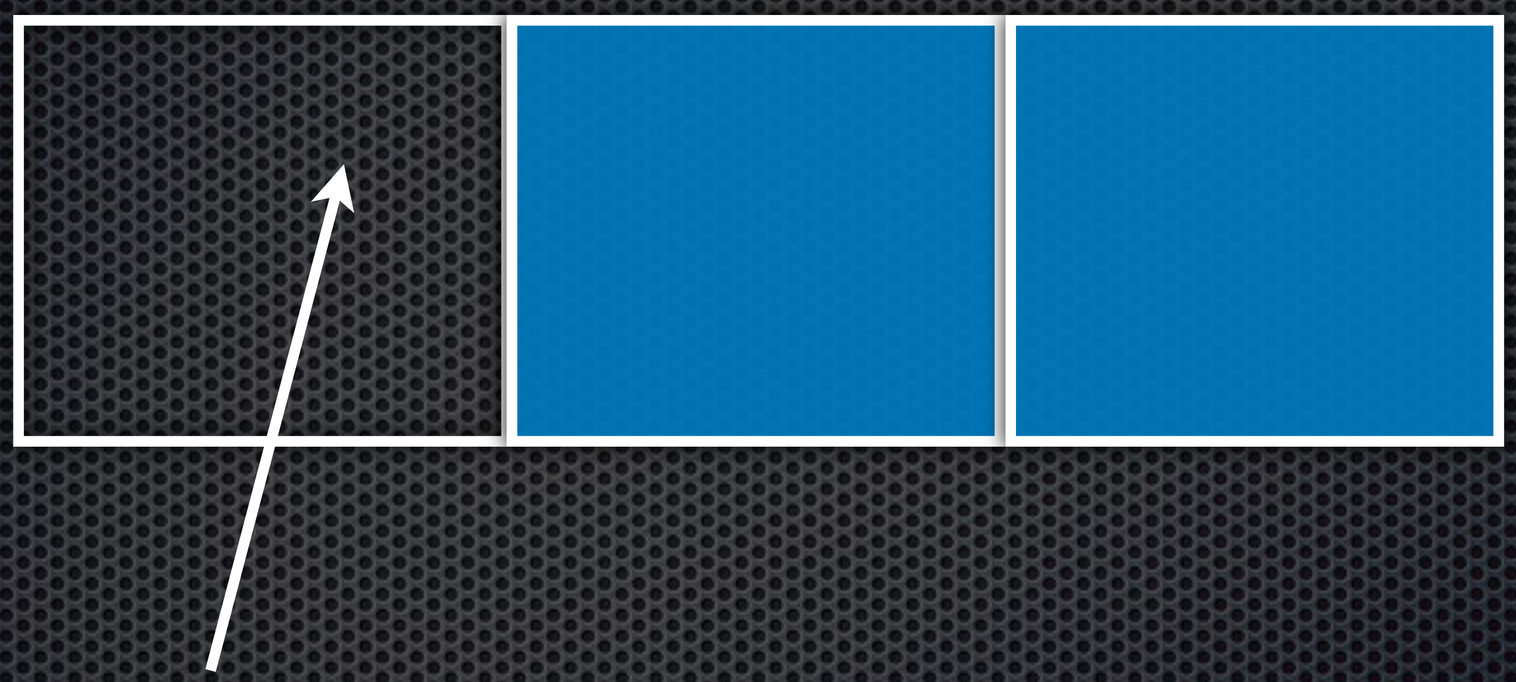


What is information?

Where is Shenyang?

What is information?

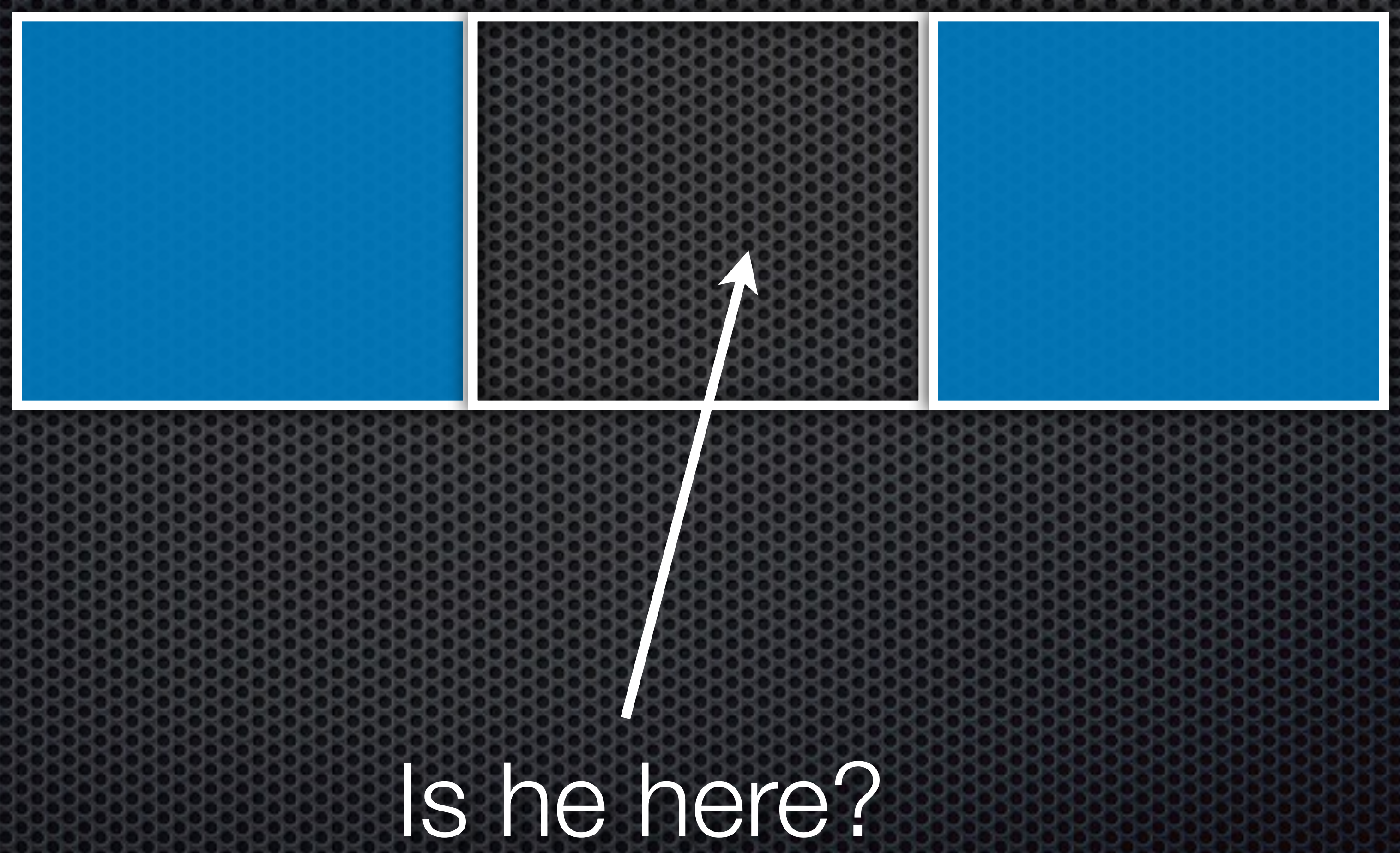
Where is Shenyang?



Is he here?

What is information?

Where is Shenyang?



What is information?

Where is Shenyang?

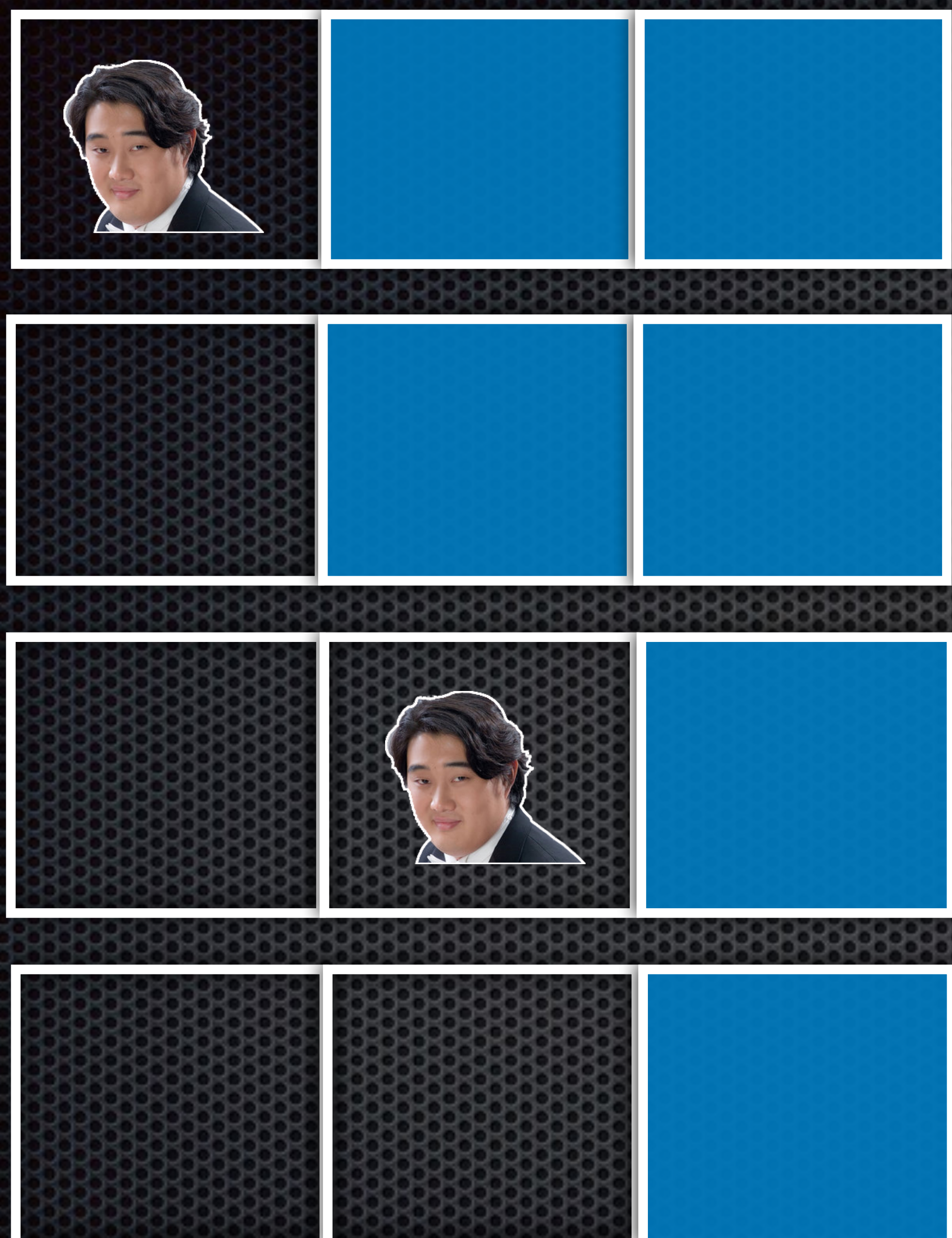
What is information? *Shannon's answer* (three properties)

Suppose an experiment can have any of *n* outcomes, and the probabilities of the outcomes are $p_1, p_2, \cdots, p_n$. If there is a measure of the amount of information contained in observing an outcome, say $H(p_1, p_2, \cdots, p_n)$, it is reasonable to assume that H satisfies the following.

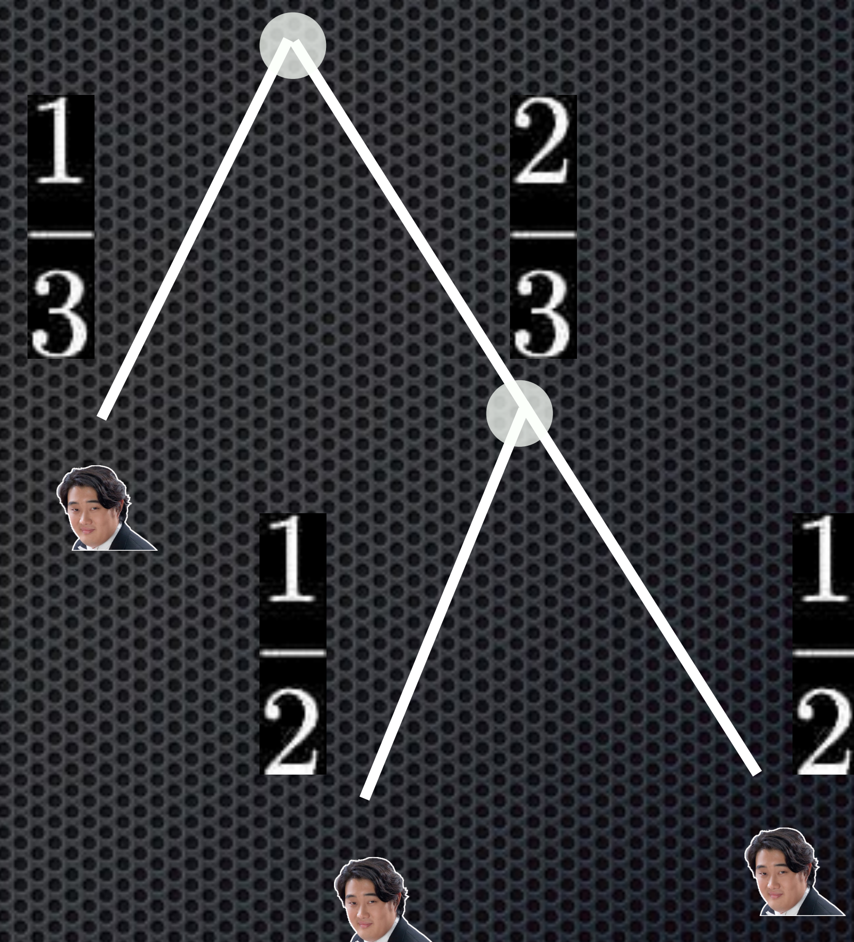
1. H should be continuous in the p_i.
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n. With equally likely events there is more choice, or uncertainty, when there are more possible events.
3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H.

$$\Rightarrow \quad H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

What is information?



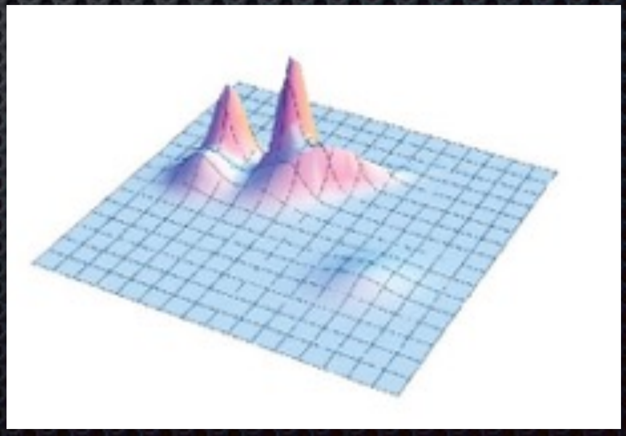
Shannon's third property:



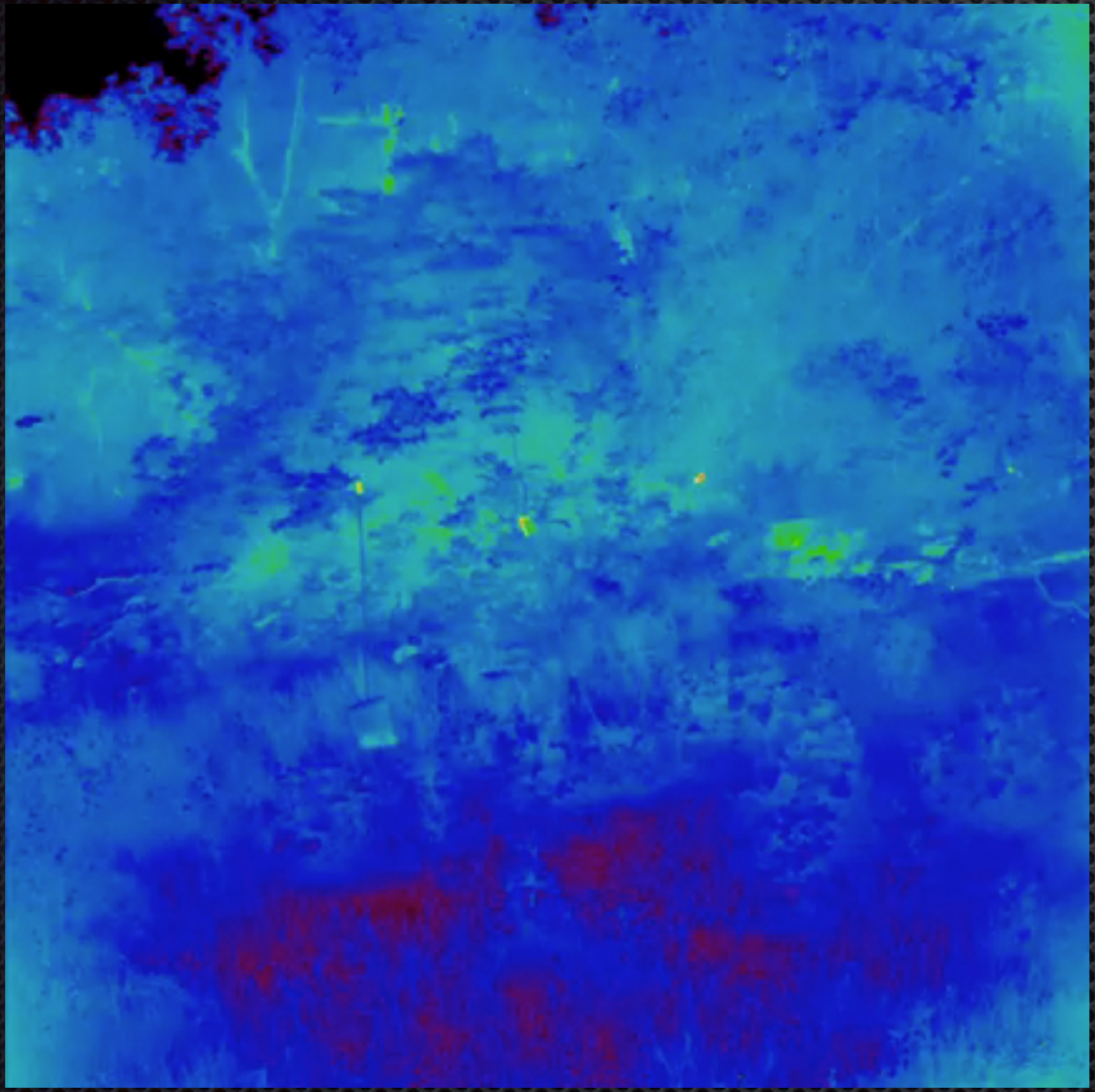
$$H(\frac{1}{3},\frac{2}{3})+\frac{2}{3}H(\frac{1}{2},\frac{1}{2})=\log_2 3$$

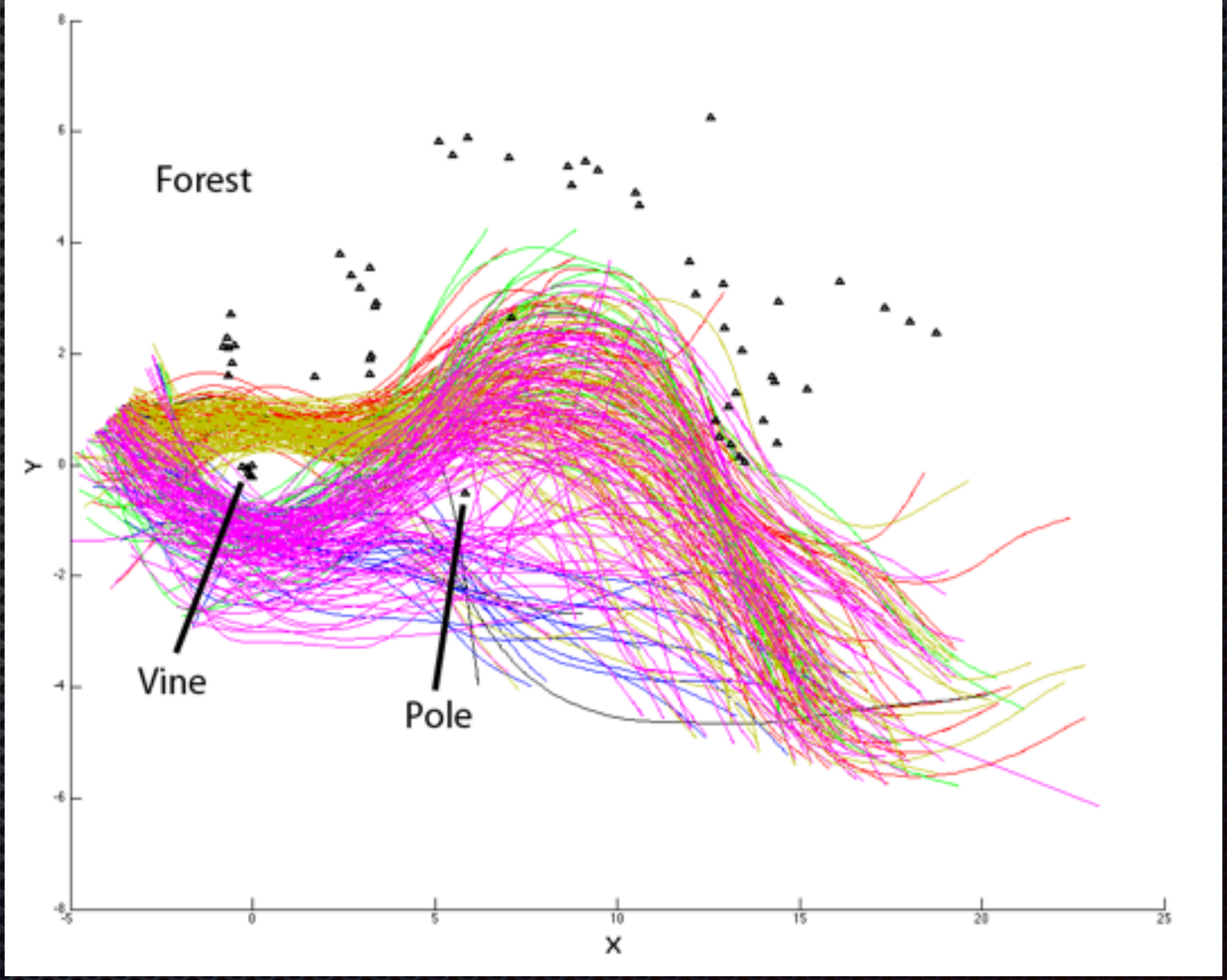
Motivation for connecting information theory and topology

- The desire to understand decision making in tasks involved in search, surveileince, reconnaissance.
- The desire to understand the cognitive psychology of perception in humans and animals.
- The desire to understand compression and sensor fusion for continuous data.
- The desire to autonomously recognize regions of interest in a data set:

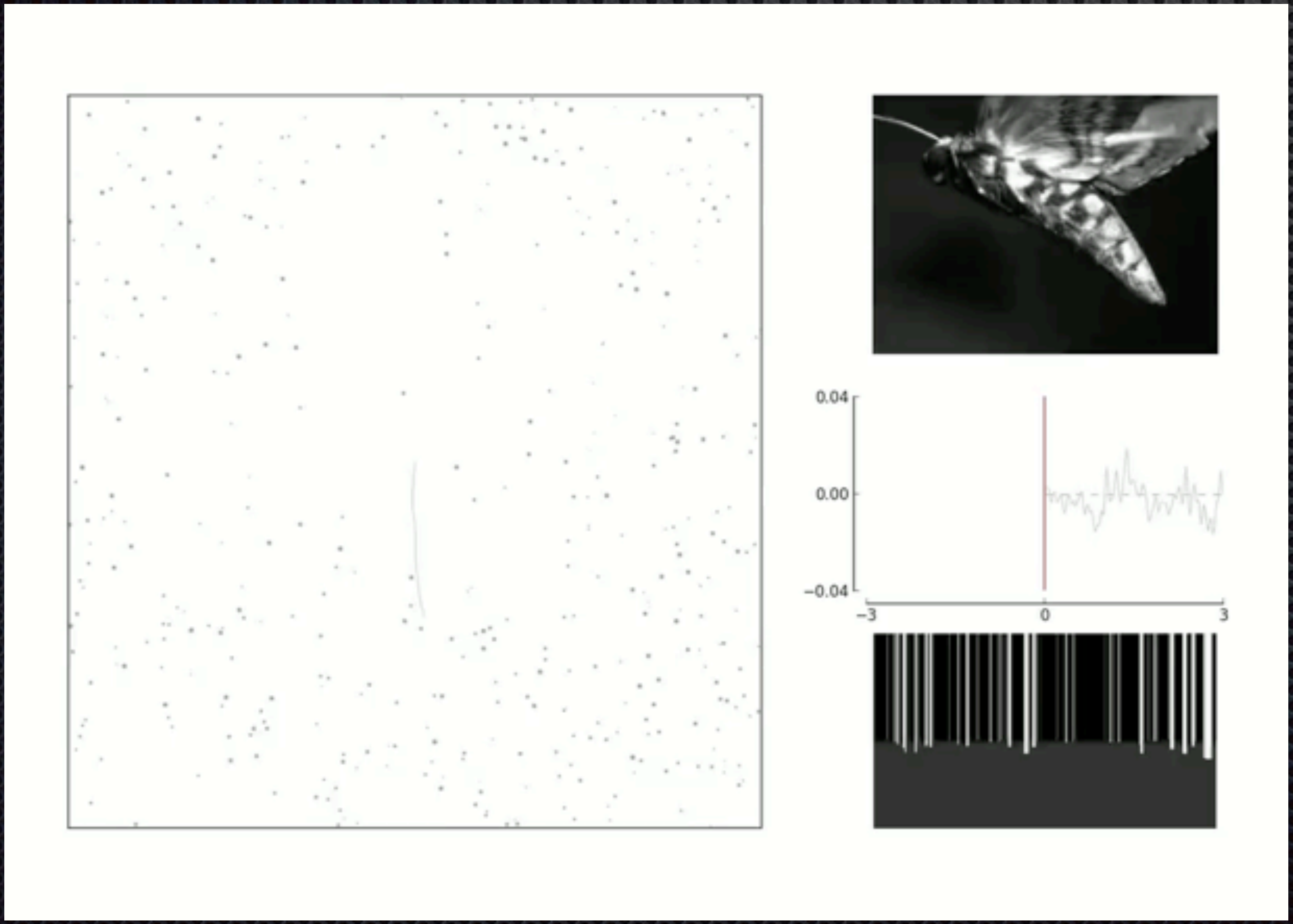


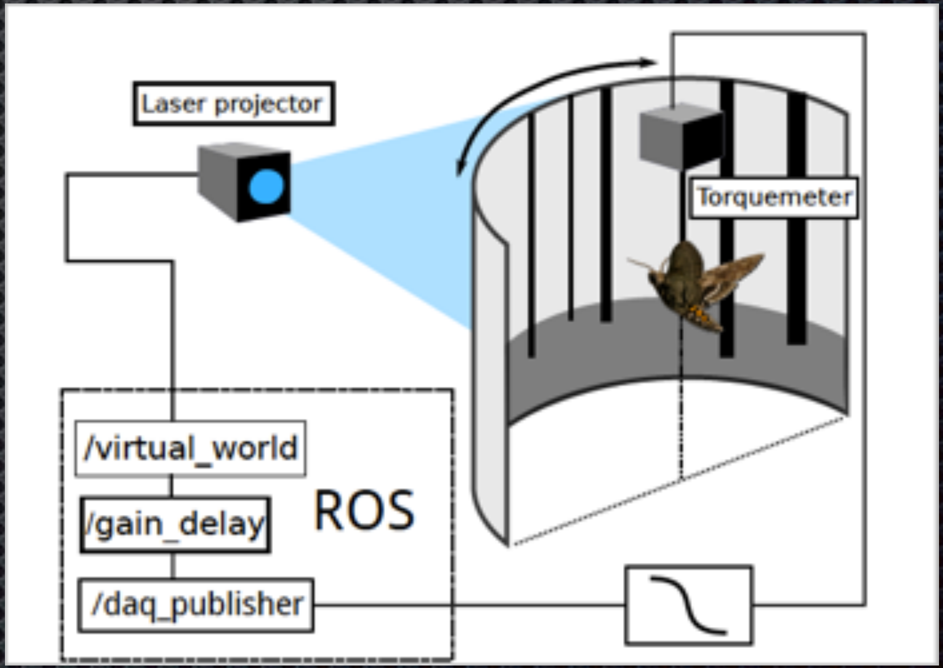
Animals are attracted to feature-rich backgrounds - the case of bats





Animals are attracted to feature-rich backgrounds - the case of Manduca sexta





Tom Daniels Lab, U. Washington

What is fundamental about human knowledge of space?

Spatial knowledge consist[s] of several quite different types of knowledge. Some is procedural, “how-to” knowledge about getting from one place to another. Some consist[s] of **topological connections** between places and travel paths. And some consist[s] of metrical layouts approximately analogous to the environment itself or to a printed map. But it is clear that accurate metrical layout descriptions come last, if at all, and depend on the earlier types of knowledge. Furthermore, spatial reasoning methods vary across individuals, with developmental stage, with experience in a particular environment, or simply with individual cognitive style.

--Benjamin Kuipers
An Intellectual History of the
Spatial Semantic Hierarchy

Functions and mappings as information channels

Suppose X is a random variable taking on values $X_1, \ldots, X_N$ with probabilities $p_1, \ldots p_N$.

The *entropy*

$$H(X) = -\sum_{k=1}^{N} p_k \log_2 p_k$$

measures the amount of information needed to for knowledge of X.

Functions and mappings as information channels

A function $f : X \to Y$ is a communication channel that provides information in the *range* Y about the structure of the *domain* X.

The function f is *informative* about X if the *mutual information*

$$I(X\,; f(X)) = H(f(X))$$

is large.

Theorem. $$H(f(X)) \leq H(X).$$

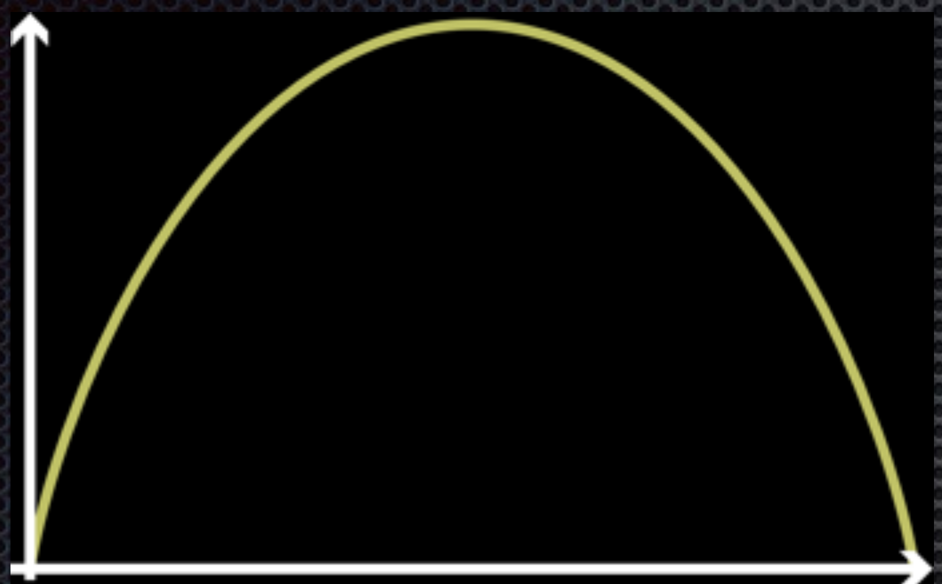
Functions and mappings as information channels

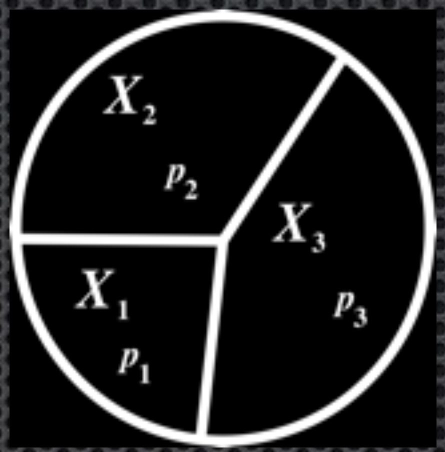
Theorem. The function that maximizes information preservation, $H(f(X))$, minimizes the conditional entropy $H(X \mid f(X))$.

Proof.

$$\begin{aligned} I(X, f(X)) &= H(X) - H(X|f(X)) \\ &= H(f(X)). \end{aligned}$$

Functions and mappings as information channels - diversity and noise





$$p_1 < p_2 < p_3$$

Among all functions $f : X \to \{0,1\}$

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

$$f(X_j) = \begin{cases} 0 & j = 1, 2 \\ 1 & j = 3 \end{cases}$$

maximizes $H(f(X))$.

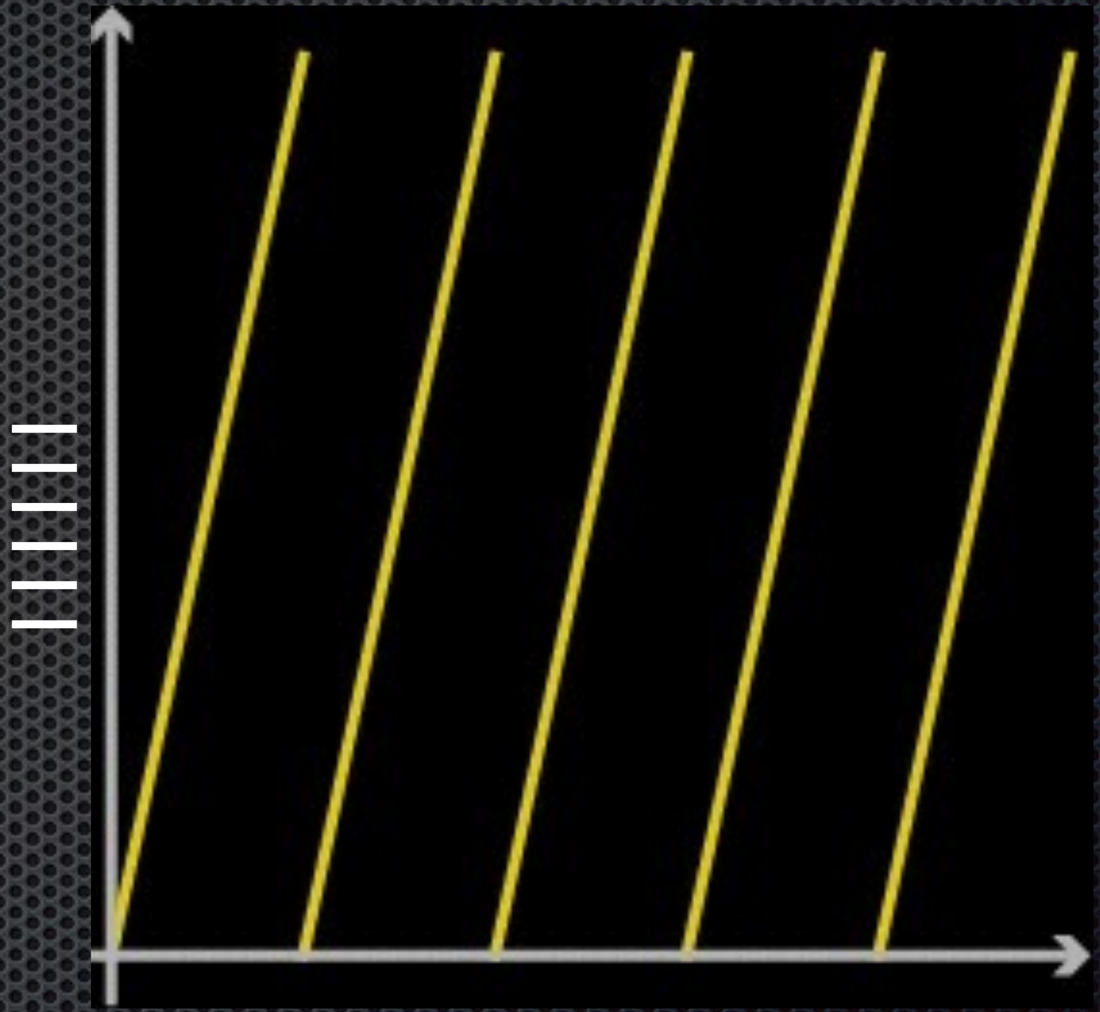
Functions and mappings as information channels - topological entropy

Consider the comb function:

$$f(x) = kx \pmod 1$$

In the equation $y = f(x)$

how much information do you have about *x* if you know *y*?



Consider a uniform partition into n subintervals of the range. The partition entropy is

$$-\sum_{k=1}^{n} \frac{\mu([y_{k-1}, y_k])}{n} \log_2 \left(\frac{\mu([y_{k-1}, y_k])}{n} \right) = \log_2 n$$

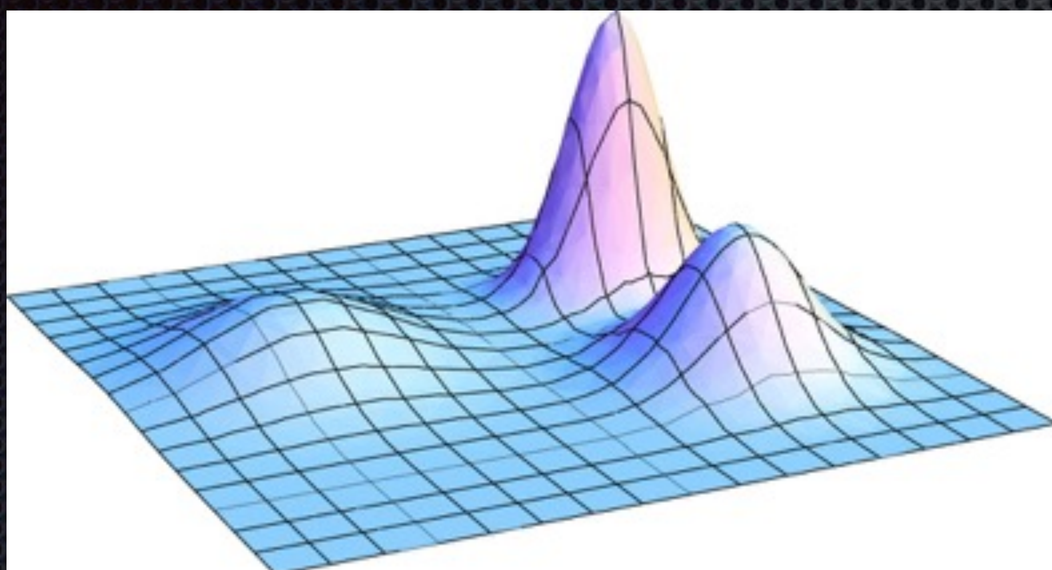
Mathematical Quantification of "Being Interesting"

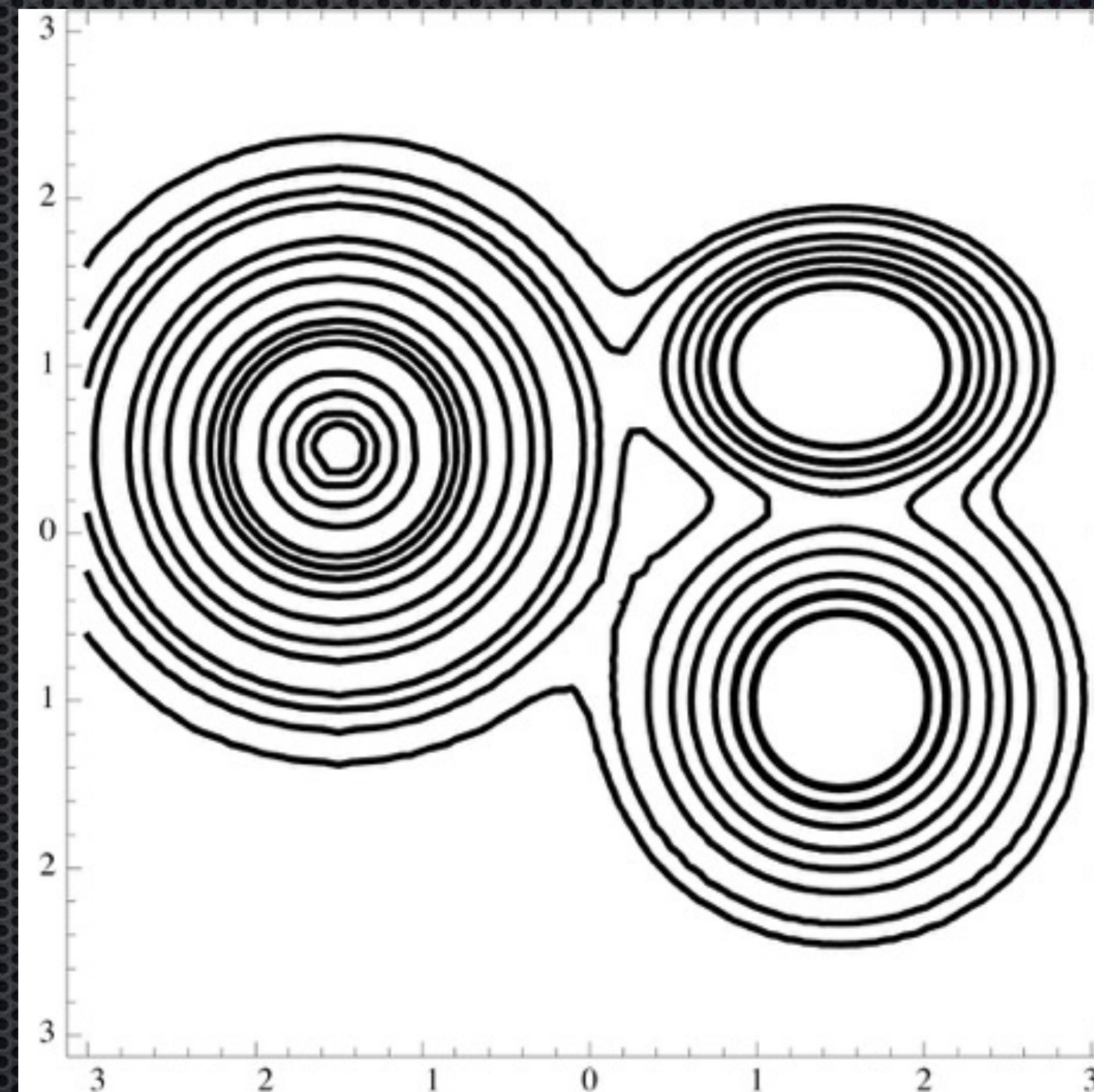
Things are interesting if they are *not* predictable.

A function $f : X \to Y$ maximally preserves information if its image reflects the diversity of X.

$$\mathcal{C}(f, \mathcal{V}) = -\sum_{j=1}^{m} \frac{\mu(f^{-1}(V_j))}{\mu(X)} \log_2 \frac{\mu(f^{-1}(V_j))}{\mu(X)}$$

The information contained in a random field





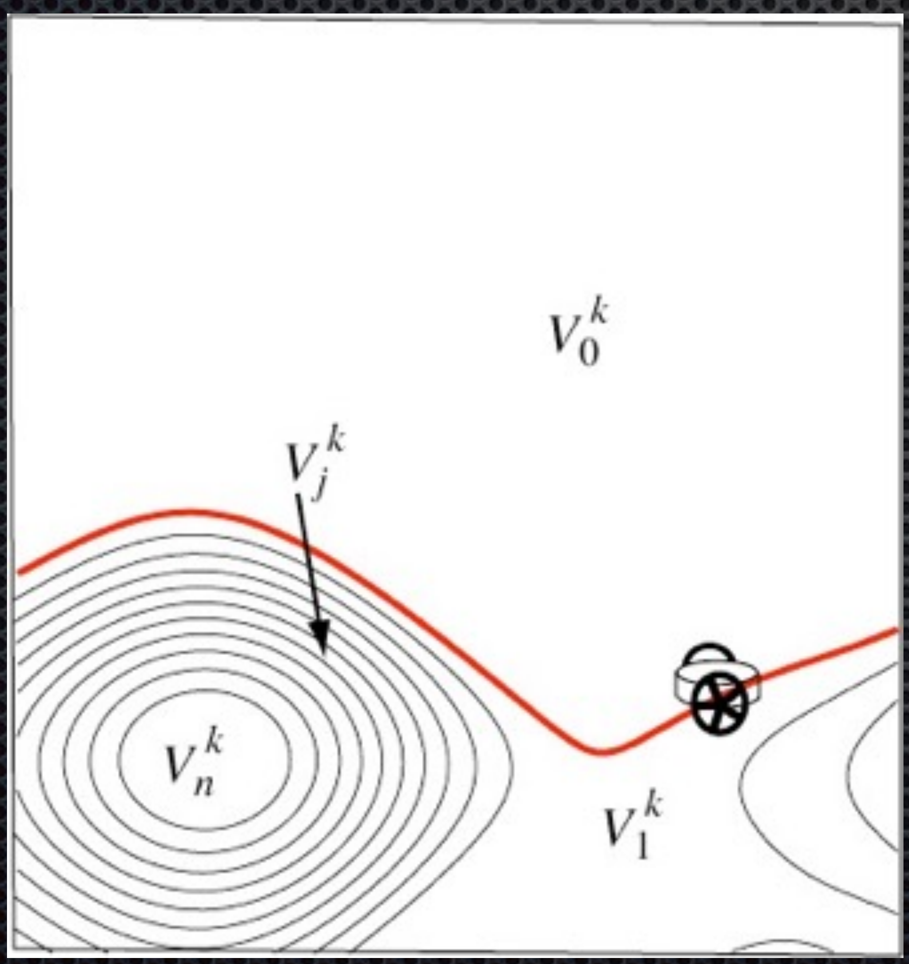
Thermal field, CO_2 concentration, chemical plume, magnetic field, . . .

Information Acquisition as a Search Metric

Let $X \subset \mathbb{R}^m$ be a compact, connected, simply connected domain, and let $[a, b]$ be the image of X under $f : \mathbb{R}^m \to \mathbb{R}$. Consider the partition

$$a = x_0 < x_1 < \cdots < x_m = b.$$

The corresponding partition $\mathcal{V}_n = \bigcup_{j=1}^{m} \left\{ \mathrm{cc}\left(f^{-1}\left([x_{j-1}, x_j]\right)\right)\right\}$ of X is called the *data-induced partition*..

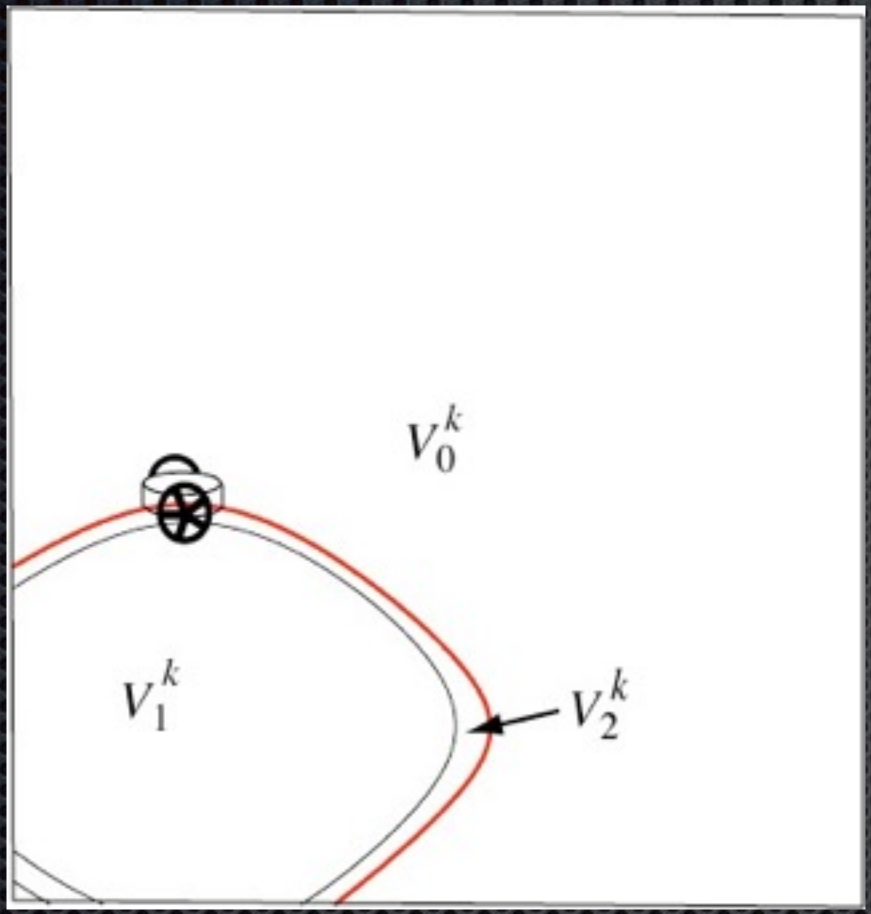


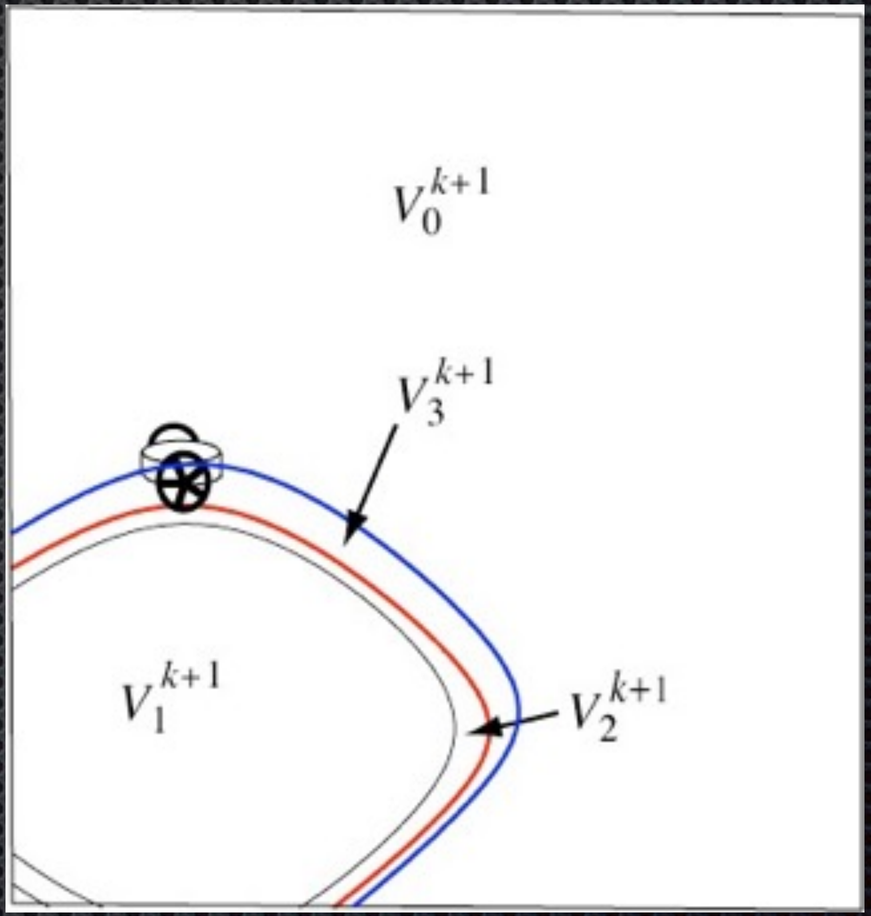
The *data-induced partition* will be deemed to be interesting or not according to the metric

$$H_k = -\sum_{j=0}^{n} \mathcal{A}(V_j^k) \log_2 \mathcal{A}(V_j^k).$$

Information Acquisition as a Search Metric

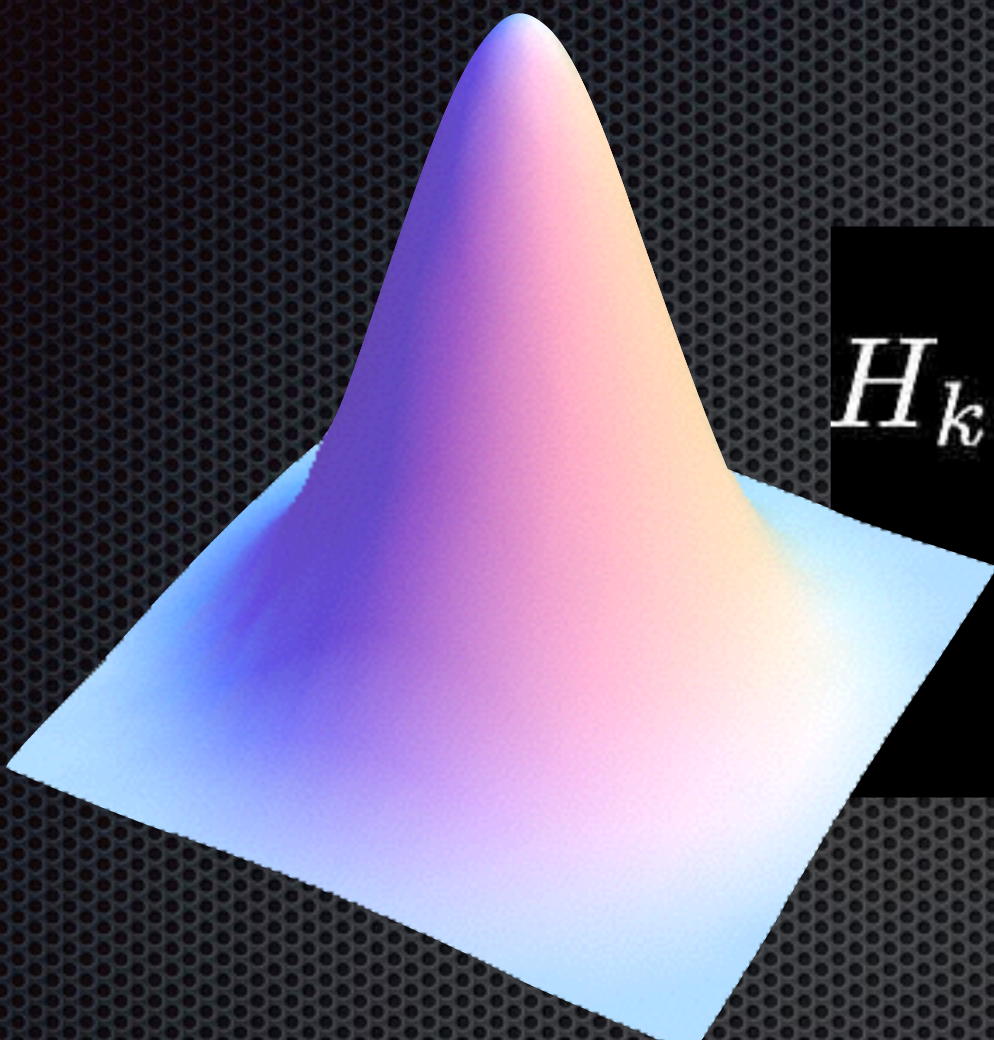
$$H_k = -\sum_{j=0}^{n} \mathcal{A}(V_j^k) \log_2 \mathcal{A}(V_j^k)$$





If $H_{k+1} - H_k > \eta$, continue *exploiting*.

Single peaks have limited entropy



$$H_k = -\sum_{j=1}^{n} \mathcal{A}(V_j^K) \log_2 \mathcal{A}(V_j^K)$$

$$\leq \log_2(n)$$

The Complexity of Monotone Structures

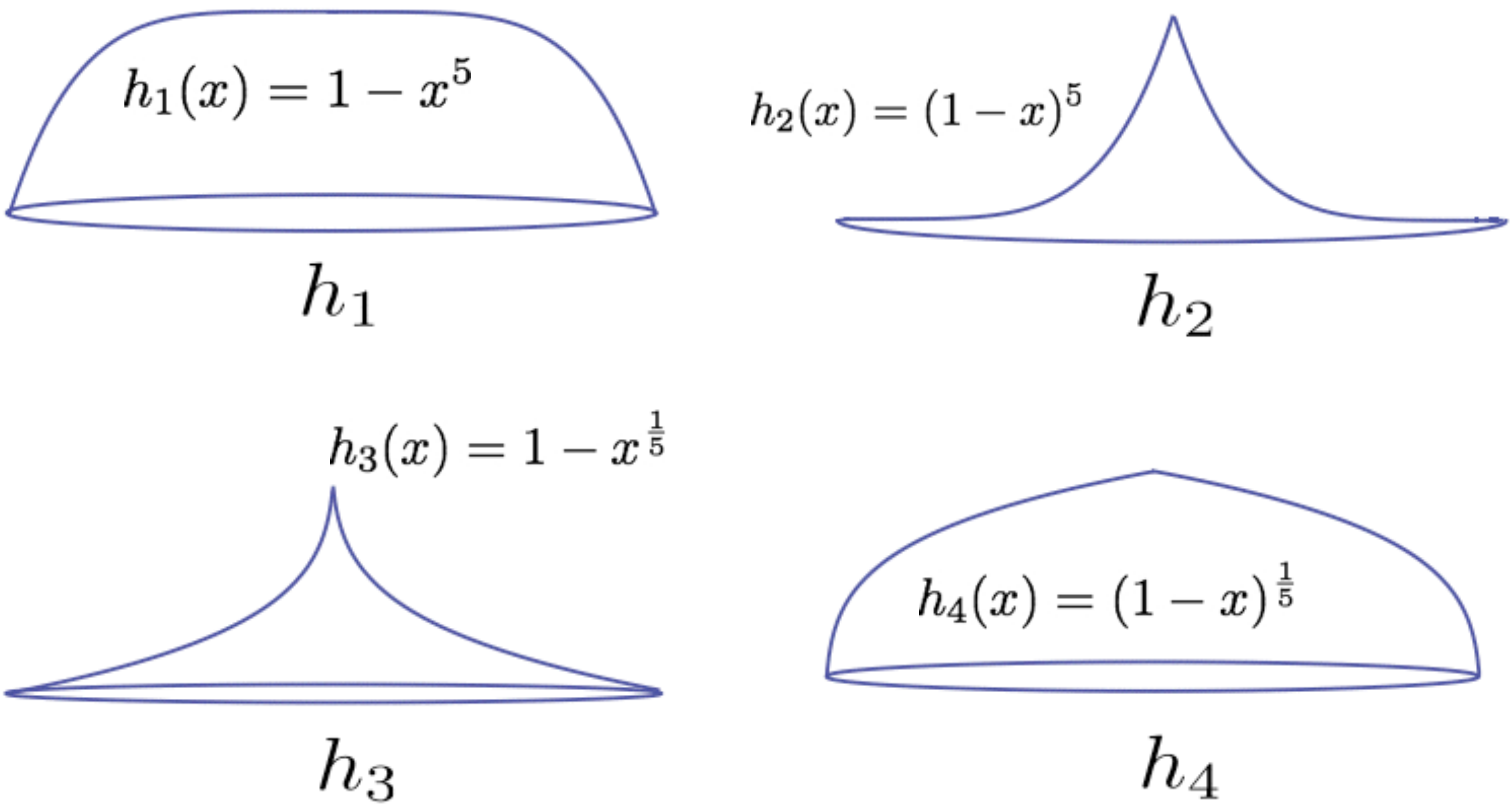


Fig. 6 The functions $h_k(\cdot)$ listed in Table 1 define radially symmetric functions on the unit circle in the way described in Section 4.1. The figures are the silhouettes of the surfaces defined by these functions $f_k(x,y) = h_k(\sqrt{x^2+y^2})$ for each function appearing in the table

The Complexity of Monotone Structures

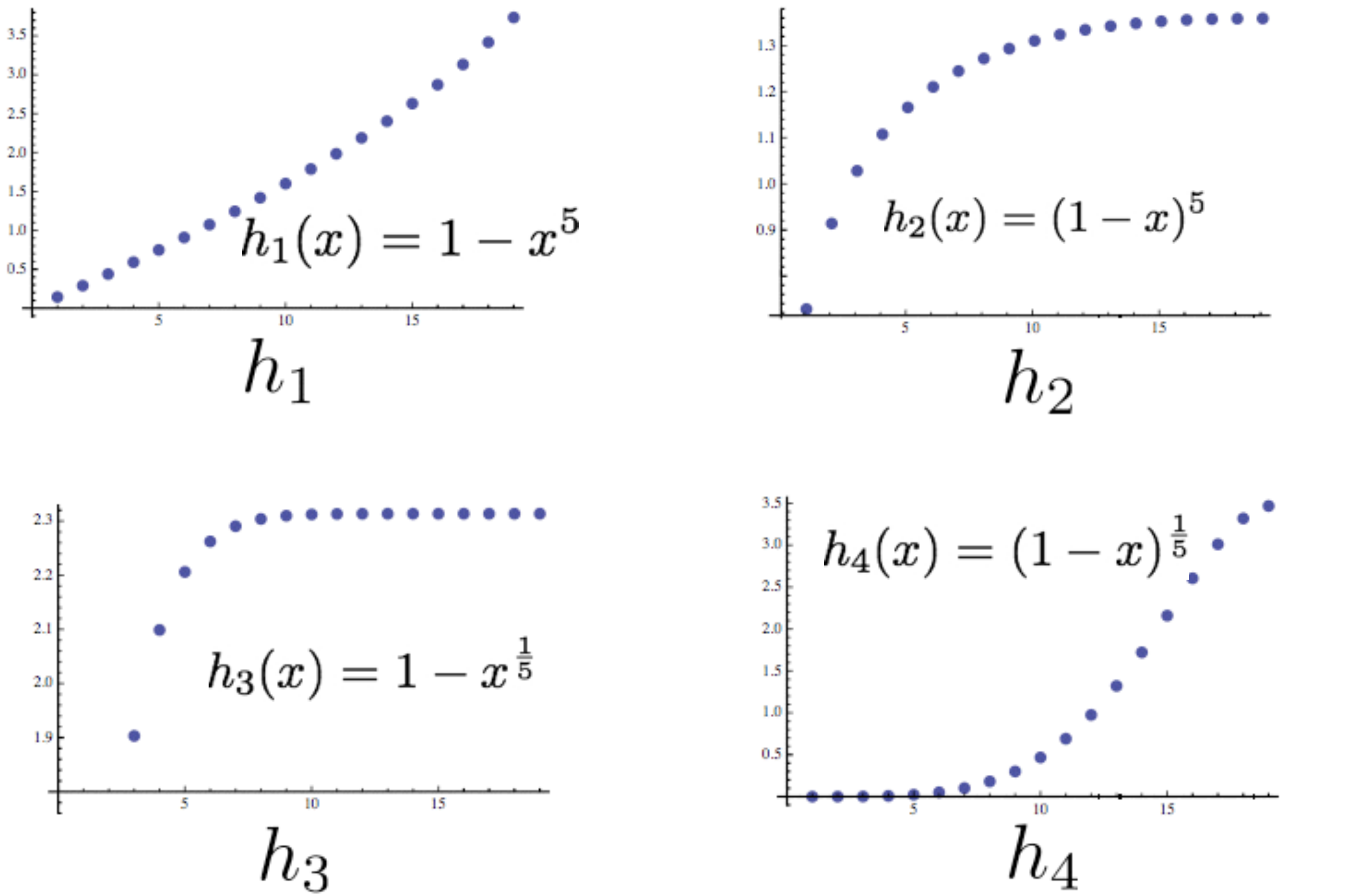
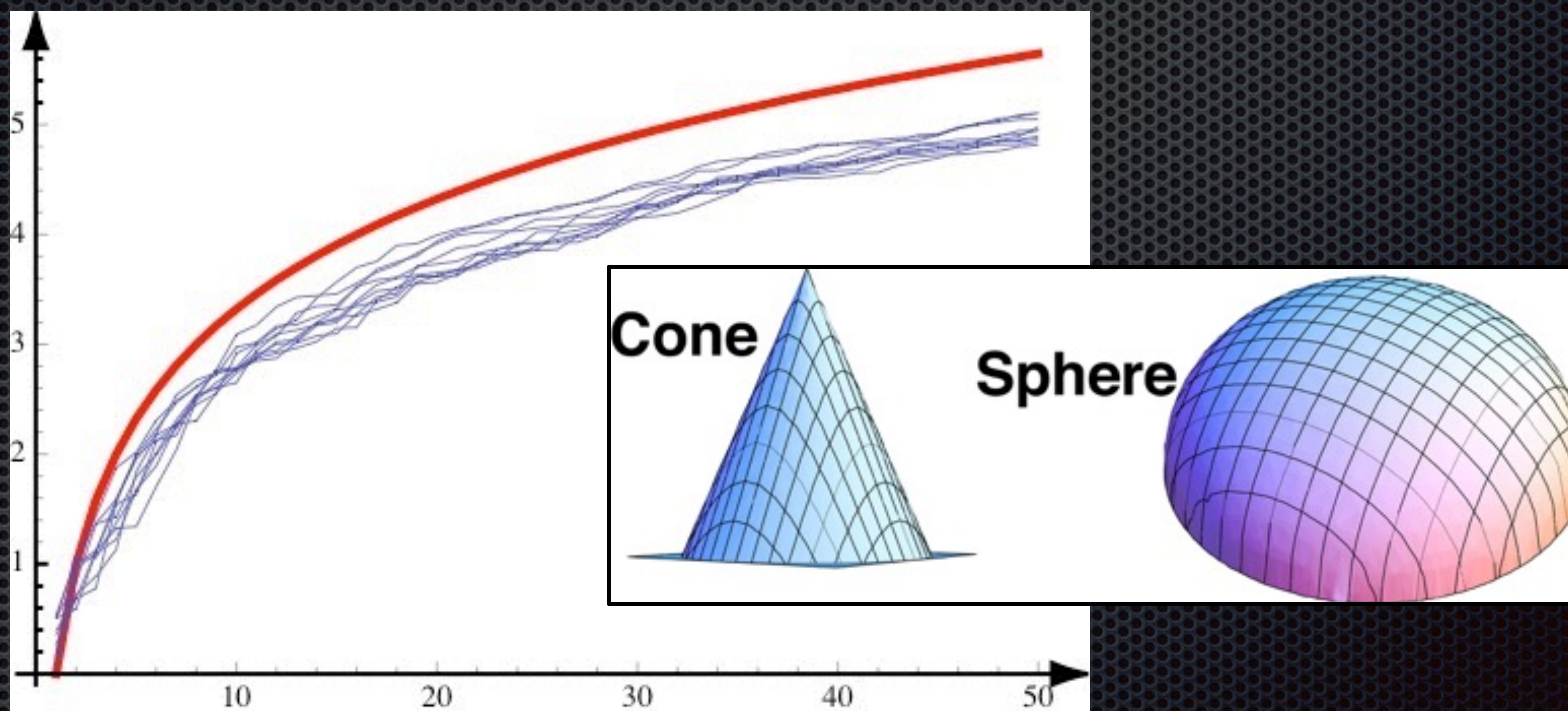


Fig. 7 The figures display the monotonic increase in partition entropy and partitions go through n successive refinements corresponding to the simple search chain and uniform twenty interval partition of the range of the monotone fields associated with the functions in the table and depicted in Fig. 6.

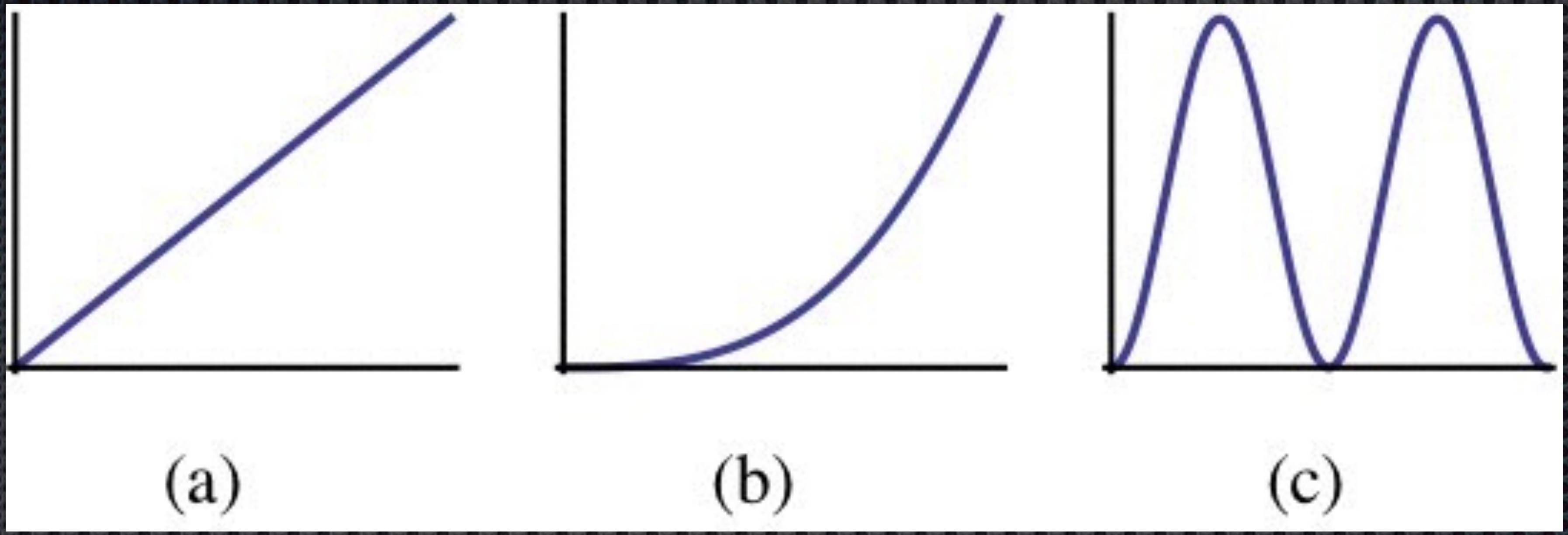
Optimal Random Reconnaissance

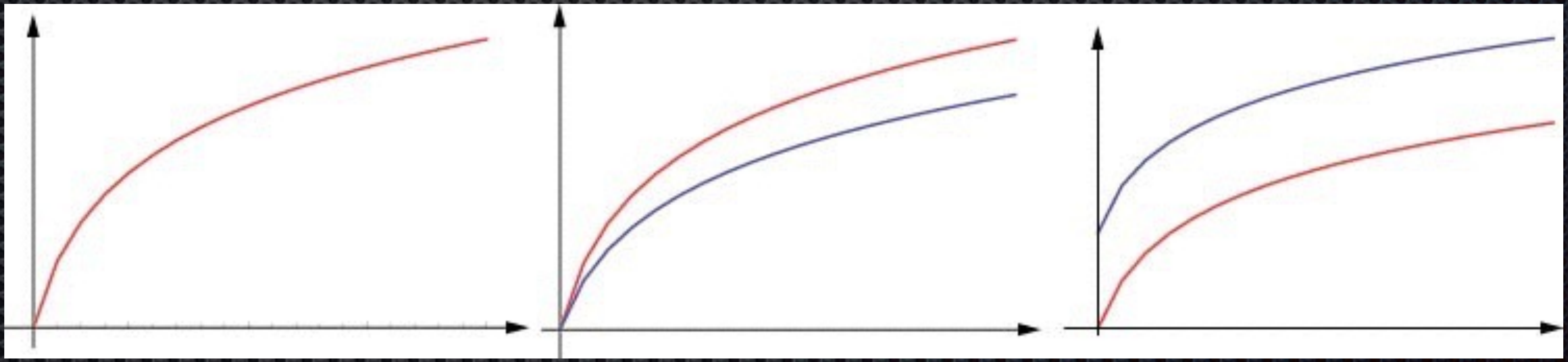
For a monotone structure, the maximum information in a data-induced partition containing n cells is $\log_2 n$.



Suitably randomized search strategies are nearly optimal in terms of the information metric.

Non-monotone functions have more entropy





Partition Entropy

To each partition, we associate a measure

$$H(\alpha) = -\sum_{A_i \in \alpha} \mu(A_i) \log_2 \mu(A_i)$$

called the *partition entropy.*

The *conditional entropy* of α conditioned on β is

$$\begin{aligned}
H(\alpha|\beta) &= \sum_{B_j \in \beta} \mu(B_j) H(\alpha|B_j) \\
&= -\sum_{B_j \in \beta} \mu(B_j) \sum_{A_i \in \alpha} \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \log_2 \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \\
&= -\sum_{B_j \in \beta} \sum_{A_i \in \alpha} \mu(A_i \cap B_j) \log_2 \frac{\mu(A_i \cap B_j)}{\mu(B_j)}.
\end{aligned}$$

Partition Entropy

Let α, β, γ be partitions of a domain X.

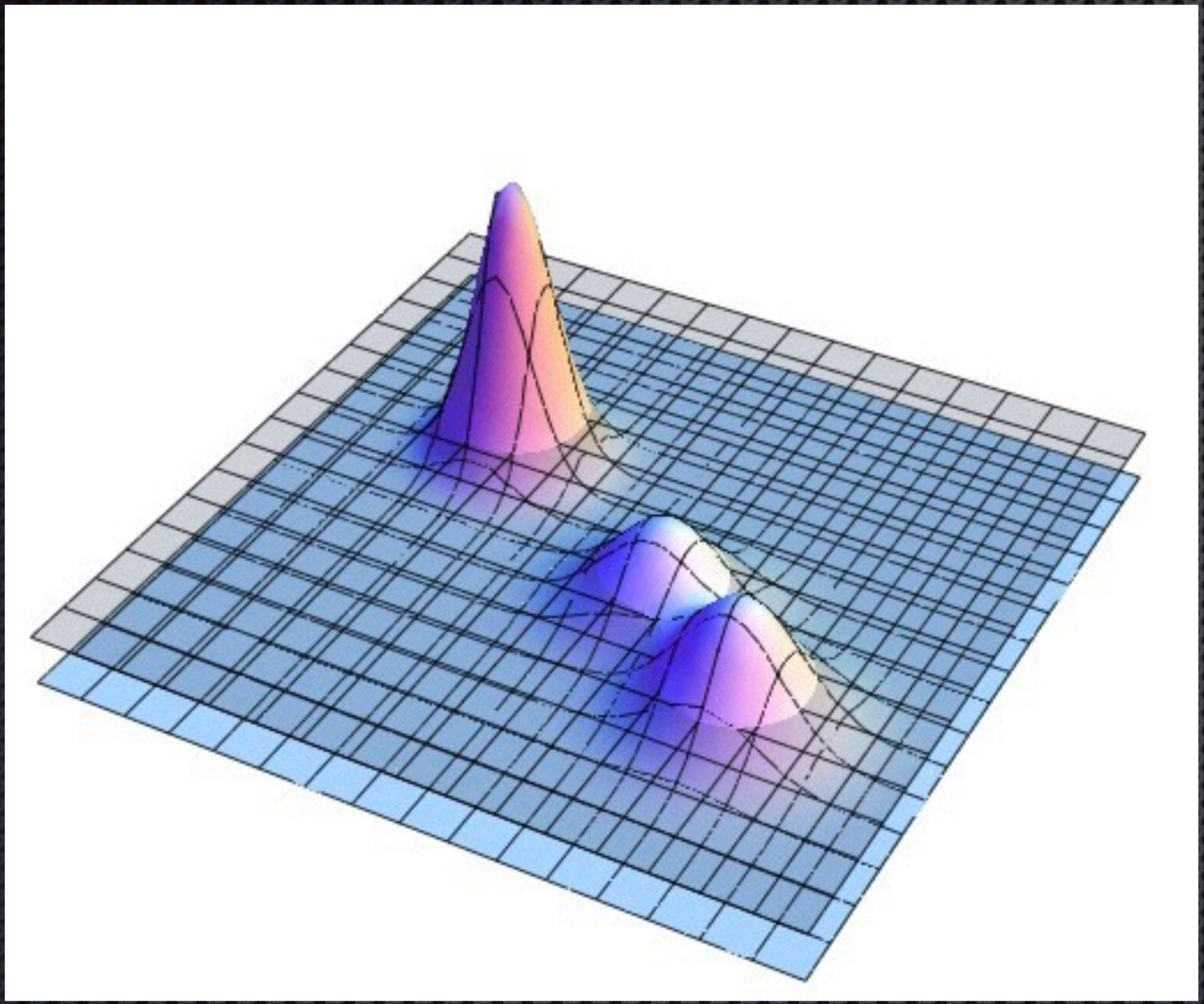
(i) $0 \leq H(\alpha|\beta) \leq H(\alpha)$ with $H(\alpha|\beta) = 0 \Leftrightarrow \beta$ is a refinement of α.

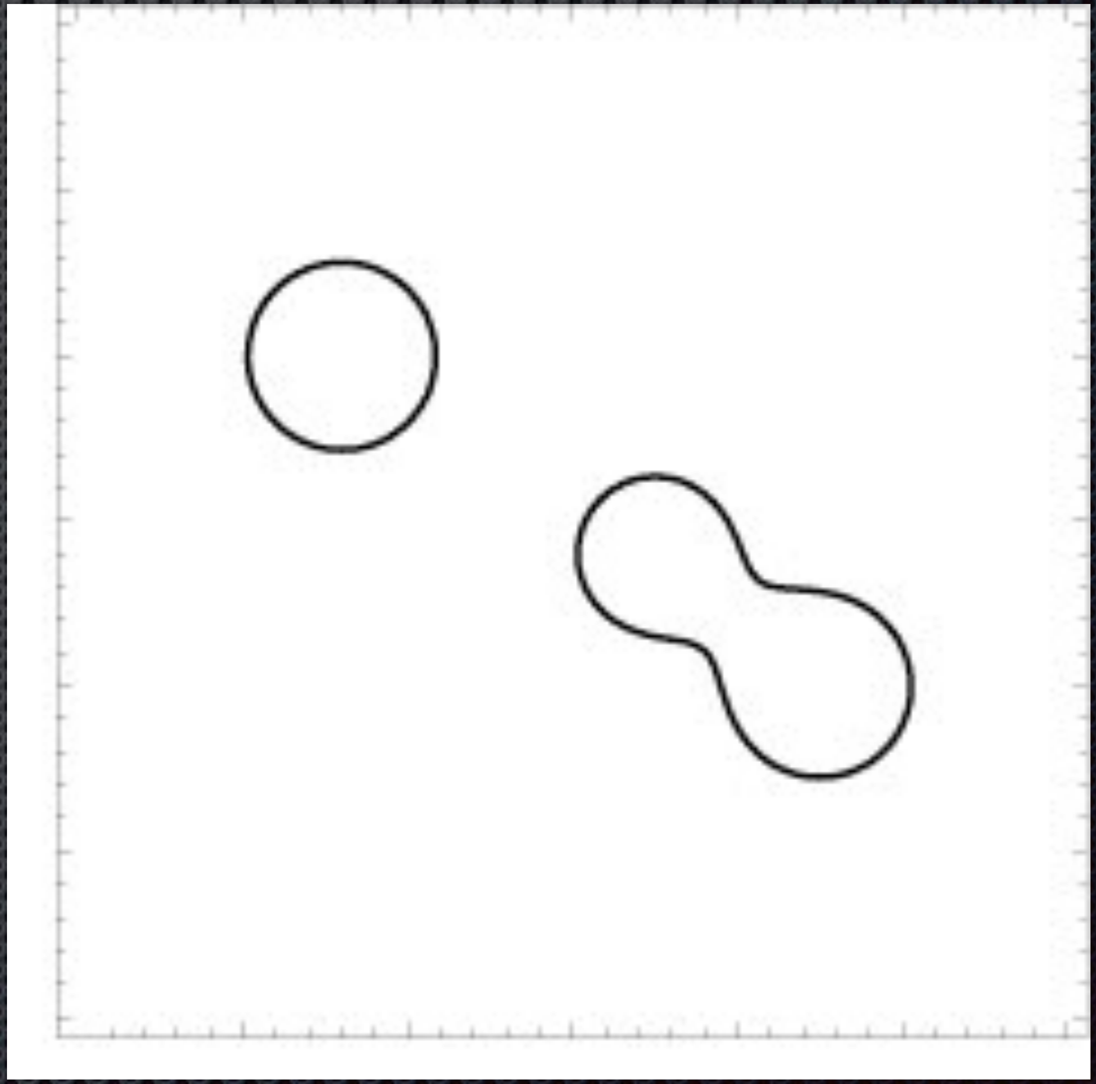
(ii) β a refinement $\alpha \Rightarrow H(\alpha|\gamma) \leq H(\beta|\gamma)$ " $<$ "
if β a proper refinement provided γ not a refinement β.

(iii) γ a refinement of $\beta \Rightarrow H(\alpha|\beta) \geq H(\alpha|\gamma)$.

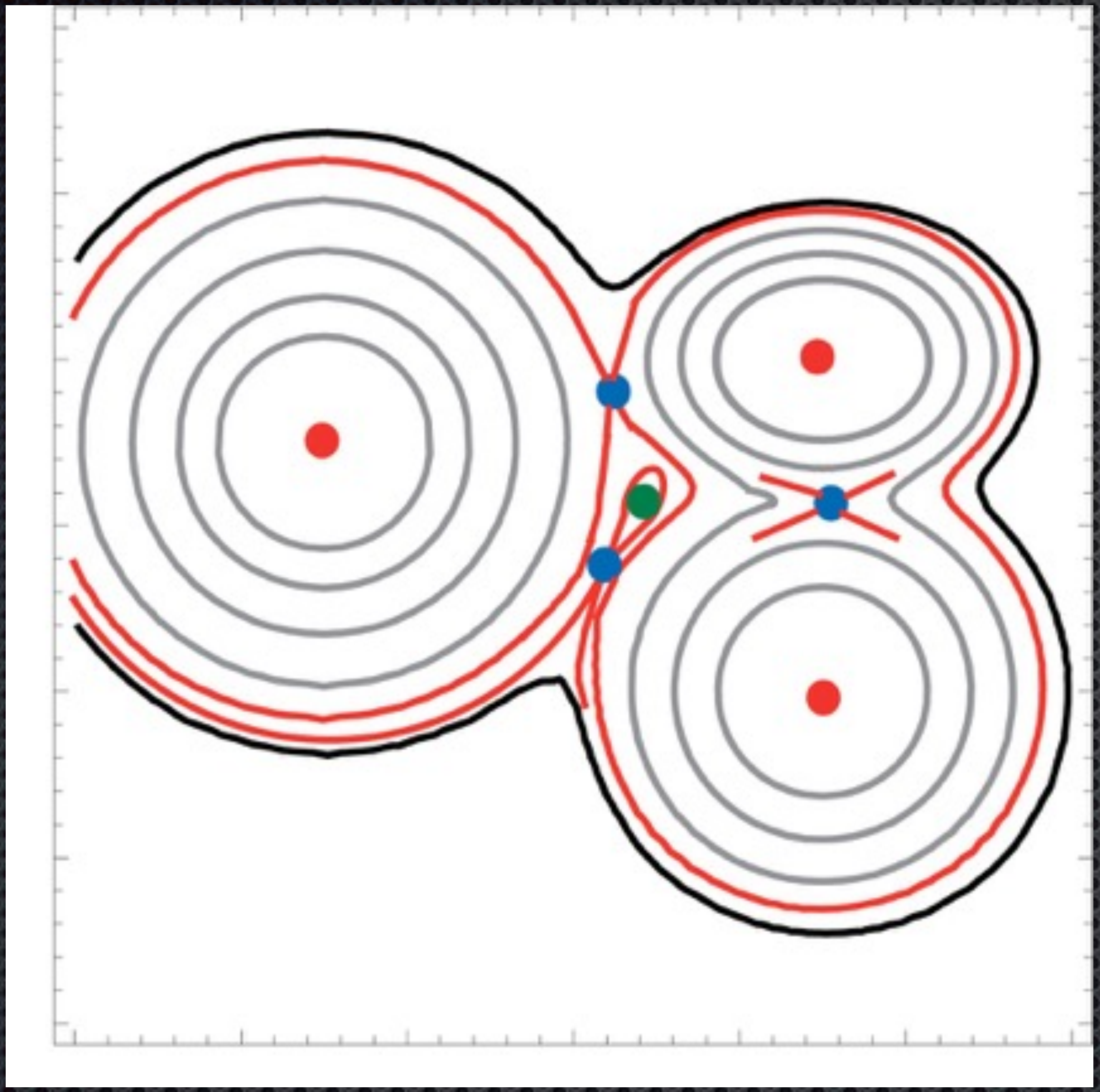
(iv) $\beta \equiv X \Rightarrow H(\alpha|\beta) = H(\alpha)$.

Level crossings, excursion sets, and the height map of random fields





Critical Level Sets Are Essential Objects in Potential Field Reconnaissance



- Index zero critical point
- Index one critical point
- Index two critical point

The topology-induced partition

Definition: A *critical level set* of f on a compact domain is a connected component of values

$$\xi(c^*) = cc(\{\mathbf{r} \in X | f(\mathbf{r}) = c^*\}),$$

$f(\mathbf{r}^*) = c^*$, with $\mathbf{r}^*$ being a critical point.

Notation: The set of all *critical level sets* is denoted

$$\mathrm{Cr}(f, X)$$

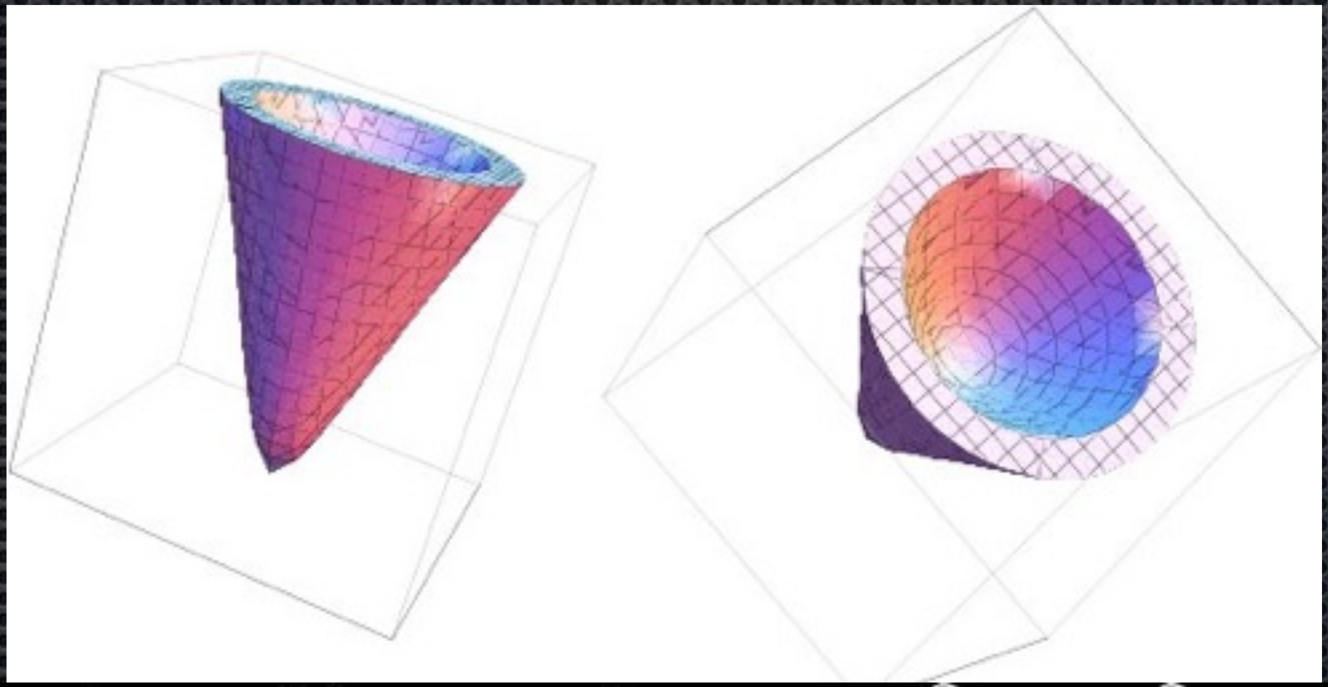
Definition: The *topology induced partition* is the domain partition

$$\mathcal{M}(f, X) = cc\left(X \setminus \mathrm{Cr}\left(f, X\right)\right)$$

The topology-induced partition

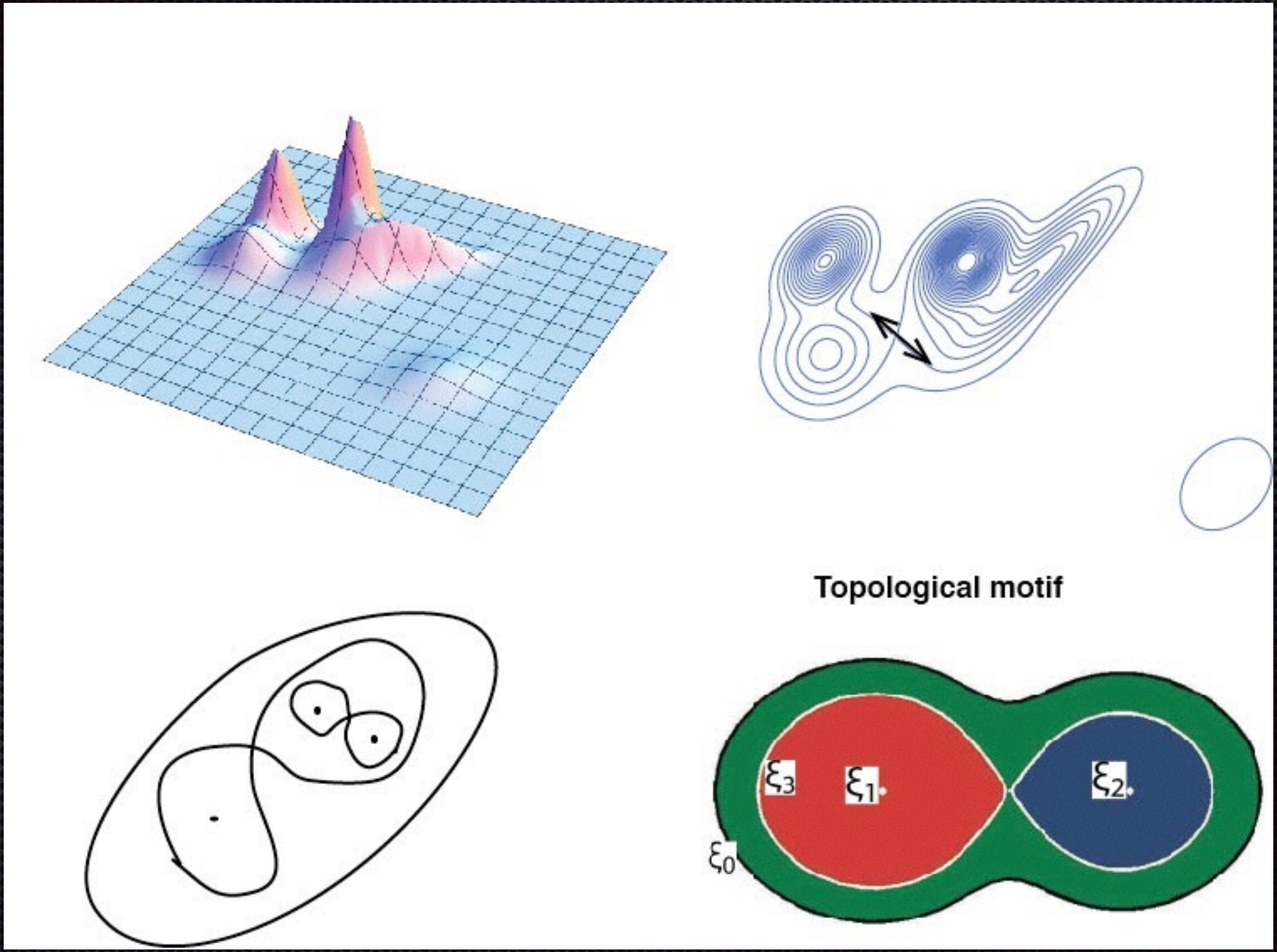
Definition: The *topology induced partition* is the domain partition

$$\mathcal{M}(f, X) = cc\left(X \setminus \mathrm{Cr}\,(f, X)\right)$$



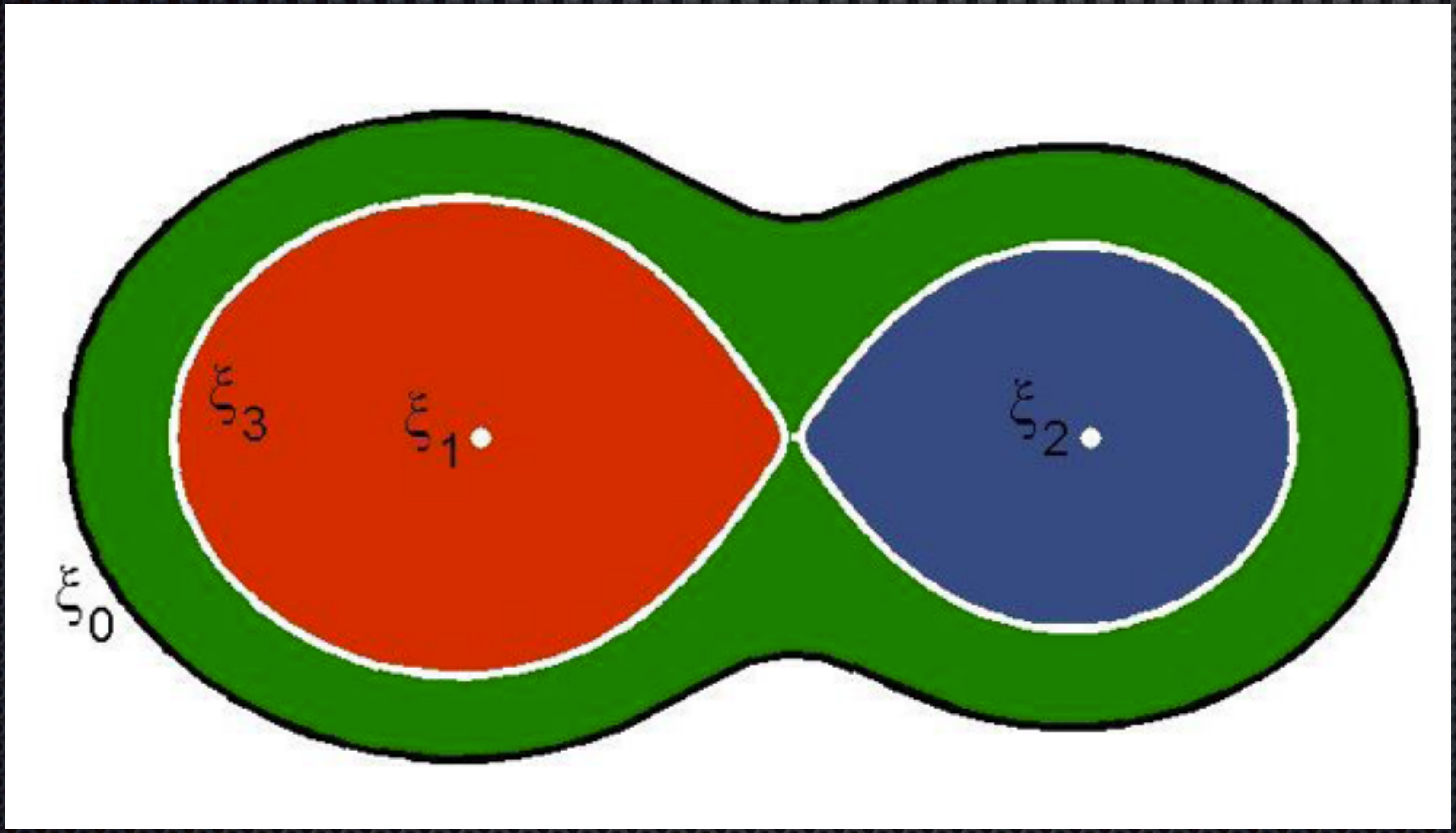
The function $f(x, y, z) = -x^2 - y^2 + z^2(z-1)$ has critical points $(0,0,0)$ and $(0,0,2/3)$. These determine a partition of the domain into three dimensional cells.

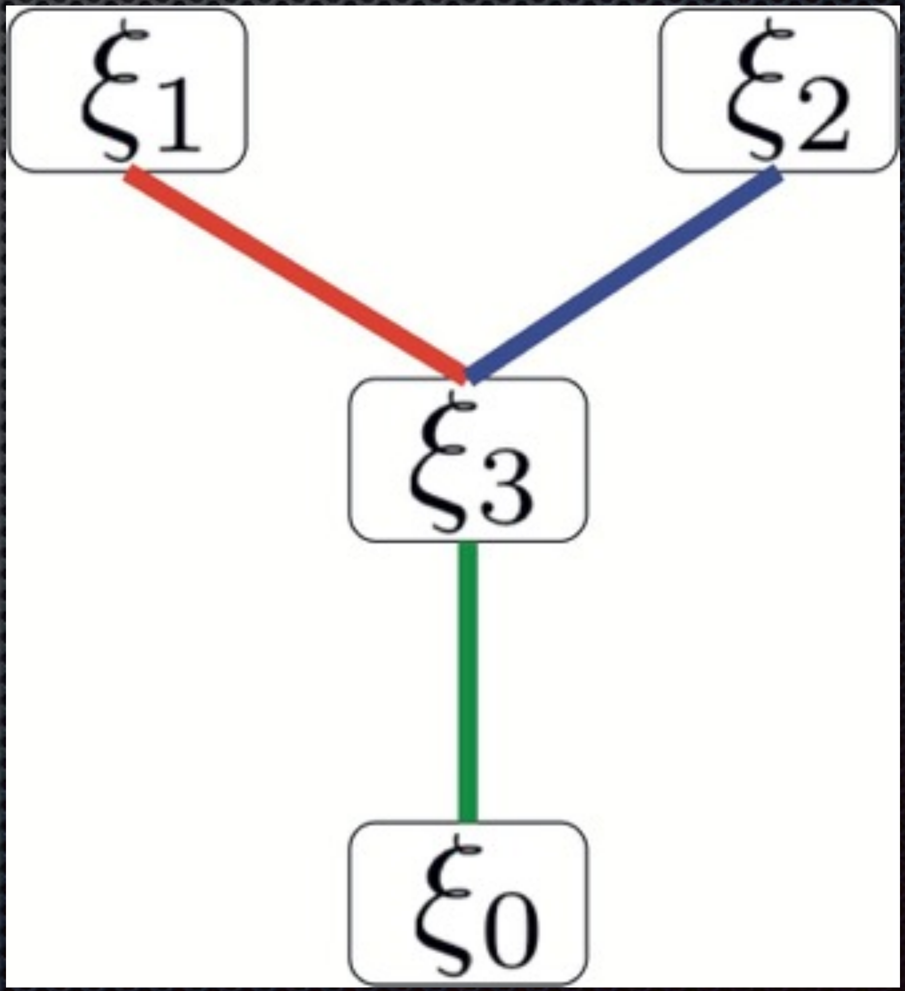
The topology-induced partition



The information theory of scalar fields and the topology-induced partition

The *topology-induced partition* of $f : X \to \mathbb{R}$





Motion Primitives for Reconnaissance of Random Scalar Fields

$b^{iso}(\mathbf{r}_0)$ = level set contour passing through r_0

$b^{grad}(\mathbf{r}_0)$ = gradient contour passing through r_0

$B_k = \{b_1, \ldots, b_k\}$ = motion program sequence sequence where $b_i \in \{b^{iso}, b^{grad}\}$

$S(B_k) = \{\xi_1, \ldots, \xi_k\}$ = contours corresponding to B_k

The data-induced partition of random scalar fields

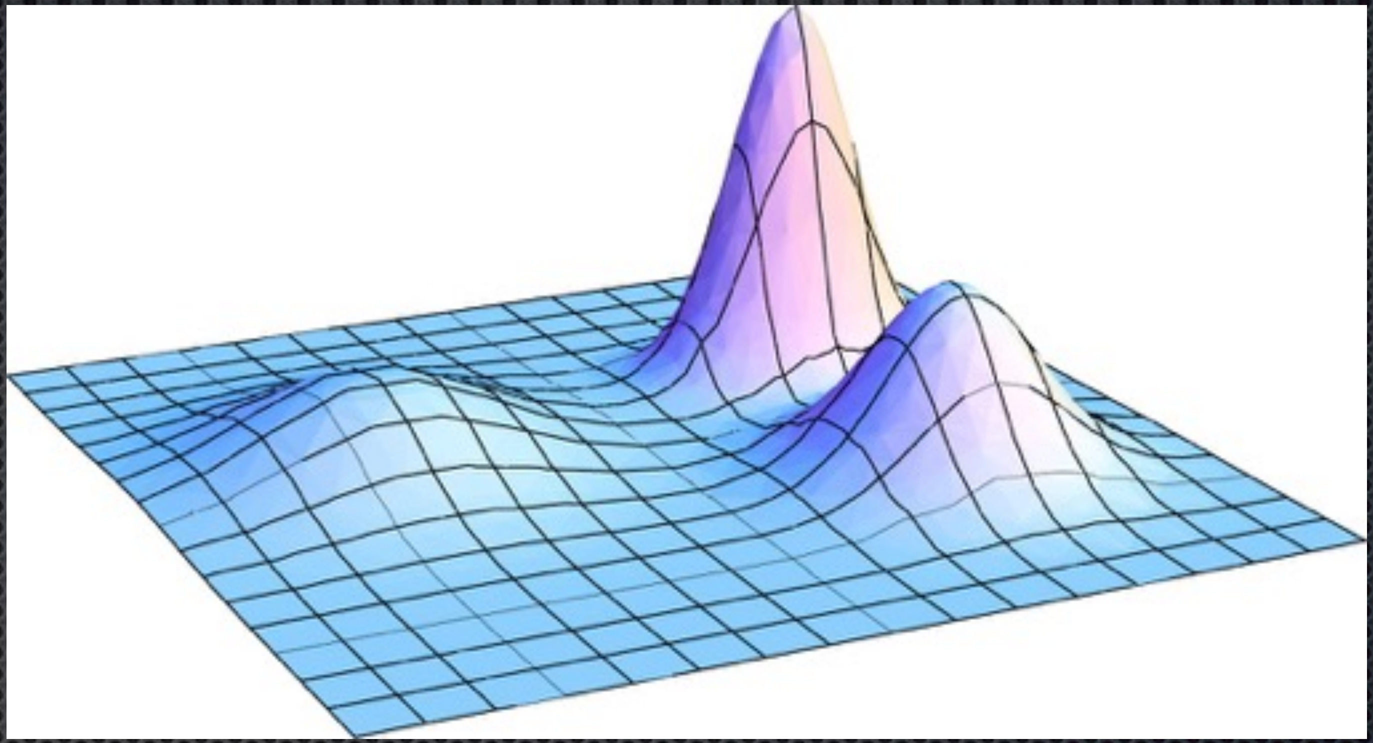
$$B_k = \{b_1, \ldots, b_k\} = \text{motion program sequence}$$
$$\text{sequence where } b_i \in \{b^{iso}, b^{grad}\}$$

$$S(B_k) = \{\xi_1, \ldots, \xi_k\} = \text{contours corresponding to } B_k$$

The data-induced partition is a proxy for the topology-induced partition:

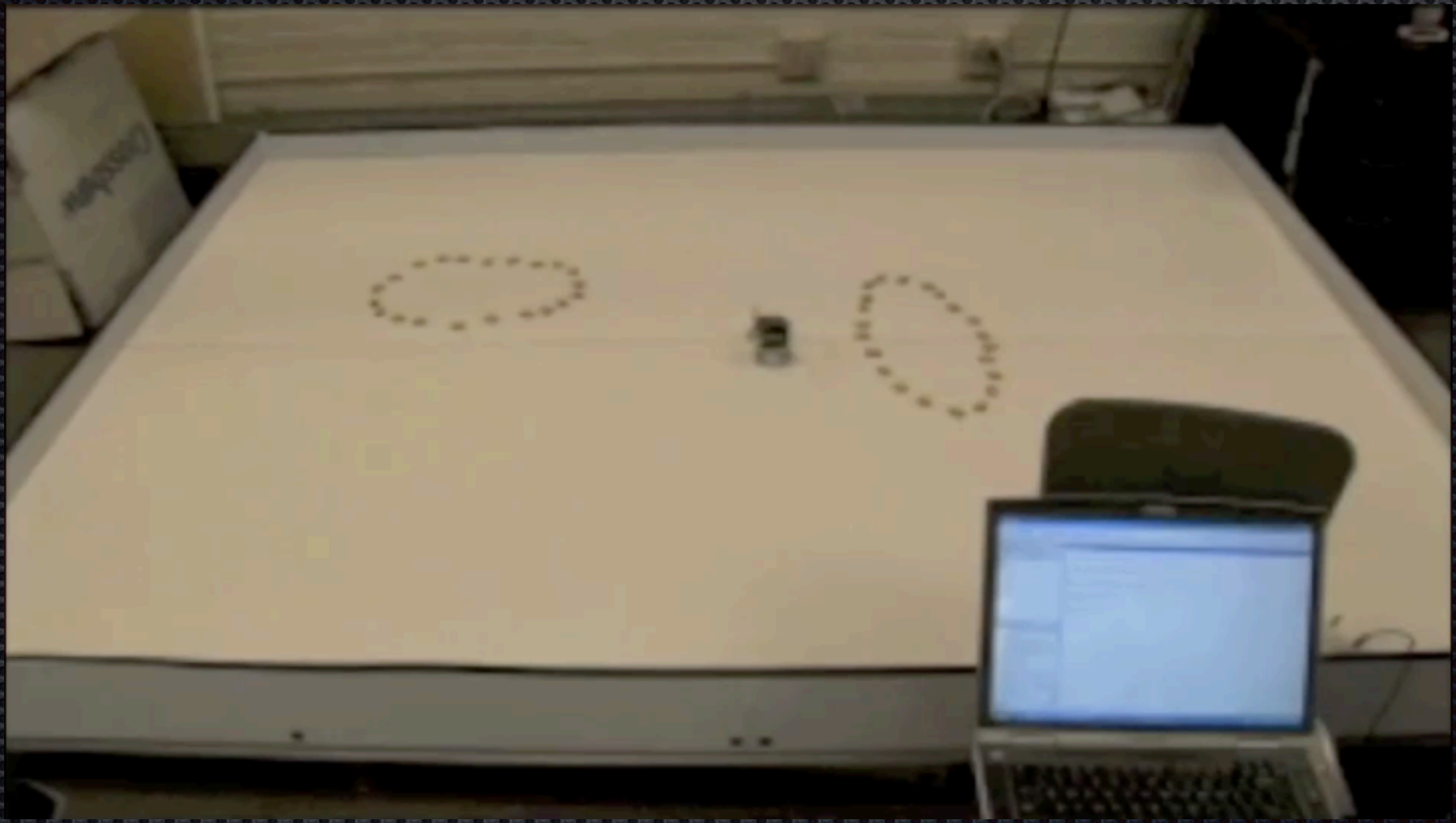
$$\begin{aligned} \mathcal{V}_n &:= \mathcal{V}(S(B_n)) \\ &= cc(X \backslash S(B_n)) \end{aligned}$$

Reconnaissance of Potential Fields Defined on 2-dimensional Domains



- Map level sets
- Map steepest ascent/descent curves

Robotic Search of an Unknown Magnetic Field



The Conditional Entropy of the TIP Given the Data Induced Partition

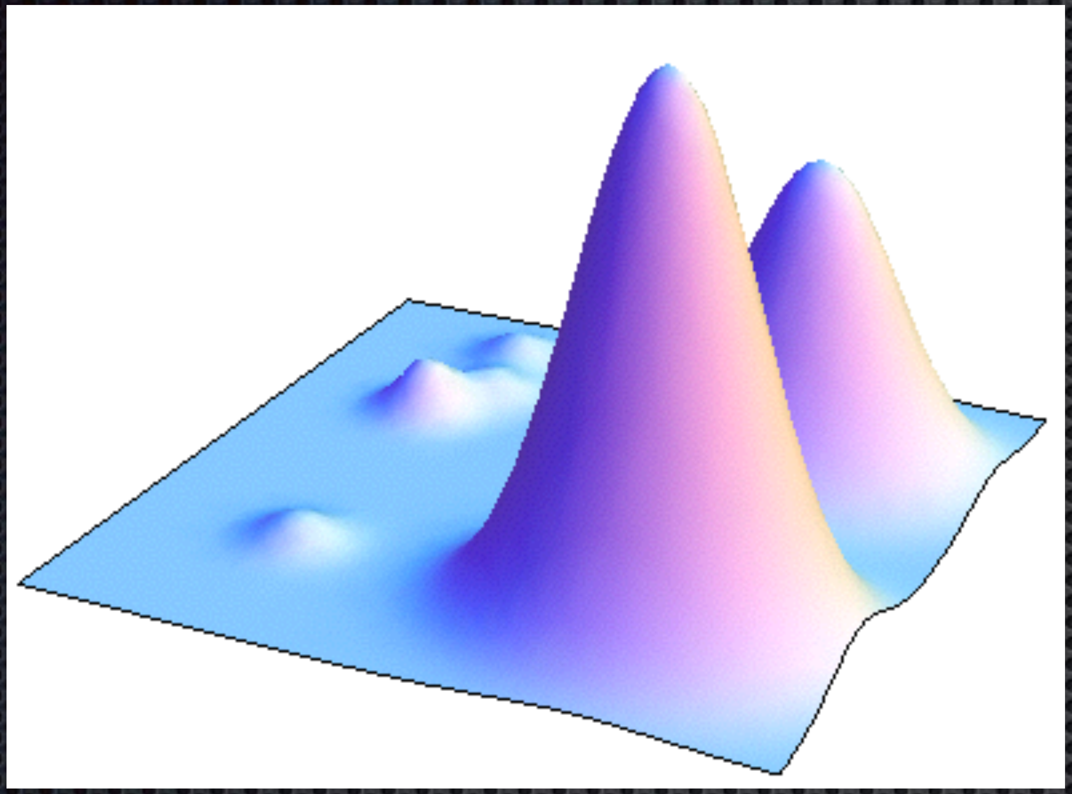
$$H(\mathcal{M}|\mathcal{V}_n) := -\sum_{M_i \in \mathcal{M}} \sum_{V_n^j \in \mathcal{V}_n} \frac{\mu(M_i \cap V_n^j)}{\mu(X)} \log_2 \frac{\mu(M_i \cap V_n^j)}{\mu(V_n^j)}$$

Theorem: Given the topology induced partition $\mathcal{M}$ and the data induced partition $\mathcal{V}_n$,

$0 \leq H(\mathcal{M}|\mathcal{V}_n) \leq H(\mathcal{M})$.

$H(\mathcal{M}|\mathcal{V}_n) = 0$ if and only if $\mathcal{V}_n$ is a refinement of $\mathcal{M}$.

Random fields as information channels - diversity and noise



Some peaks are more significant than others.

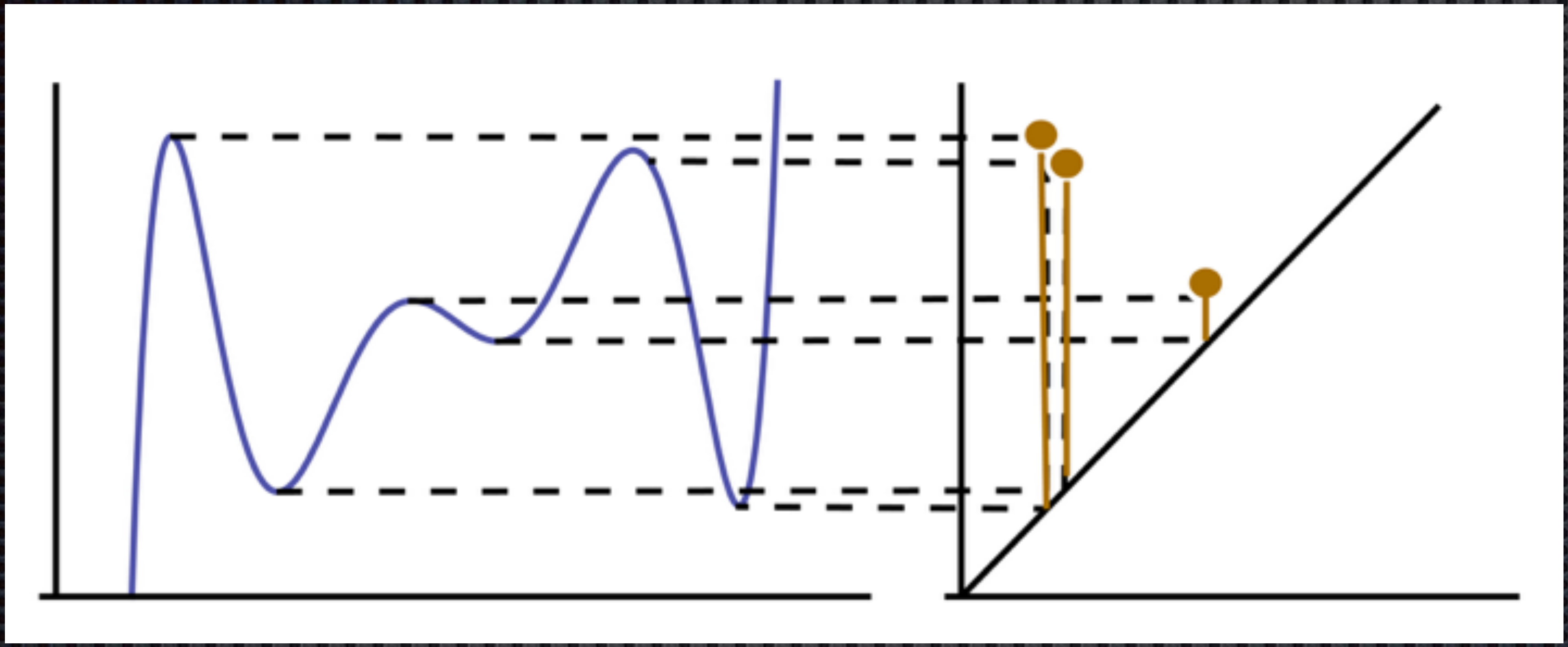
How can we quantify feature *significance*?

How do we respect diversity and reject noise?

1. Height.
2. *Topological persistence*.
3. *Information utility*.

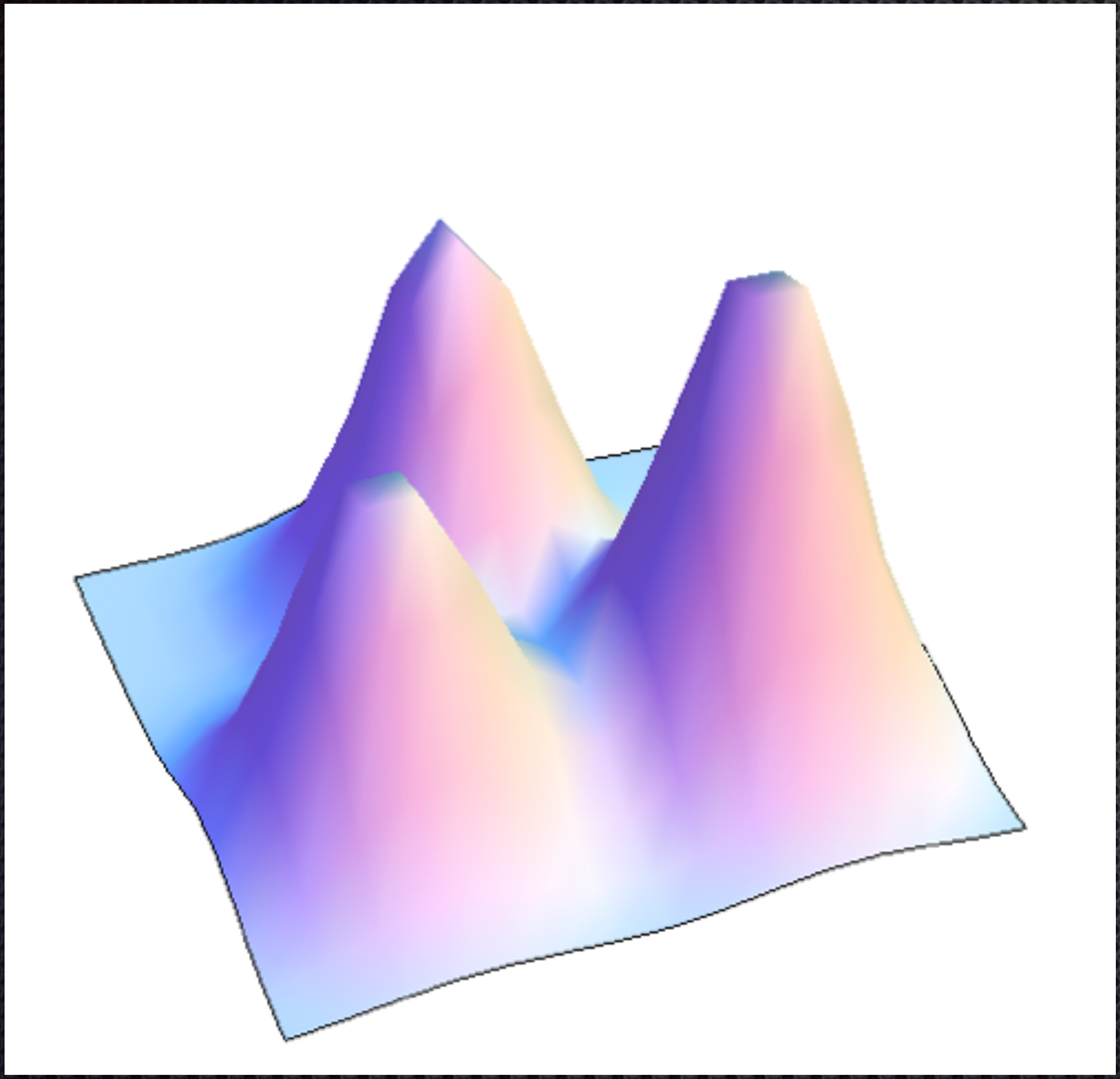
Random fields as information channels - diversity and noise

How can we quantify feature *significance*?



1. Height of critical values.
2. *Topological persistence*.
3. *Information utility*.

Random fields as information channels - diversity and noise

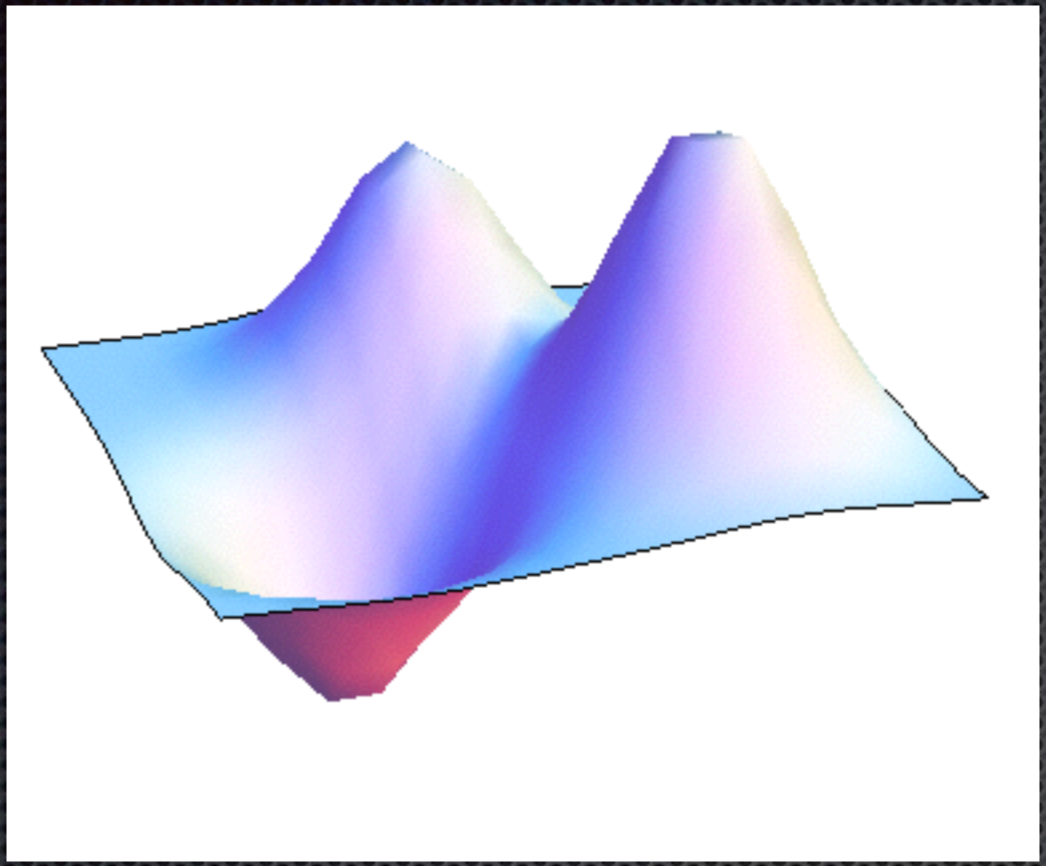


As the dimension of X increases, the topology of the sublevel sets

$$\mathbb{R}_t = f^{-1}(-\infty, t]$$

becomes more complex.

The Differential Topology of Scalar Fields



Let regular values x_j

$$a = x_0 < x_1 < \cdots < x_m = b.$$

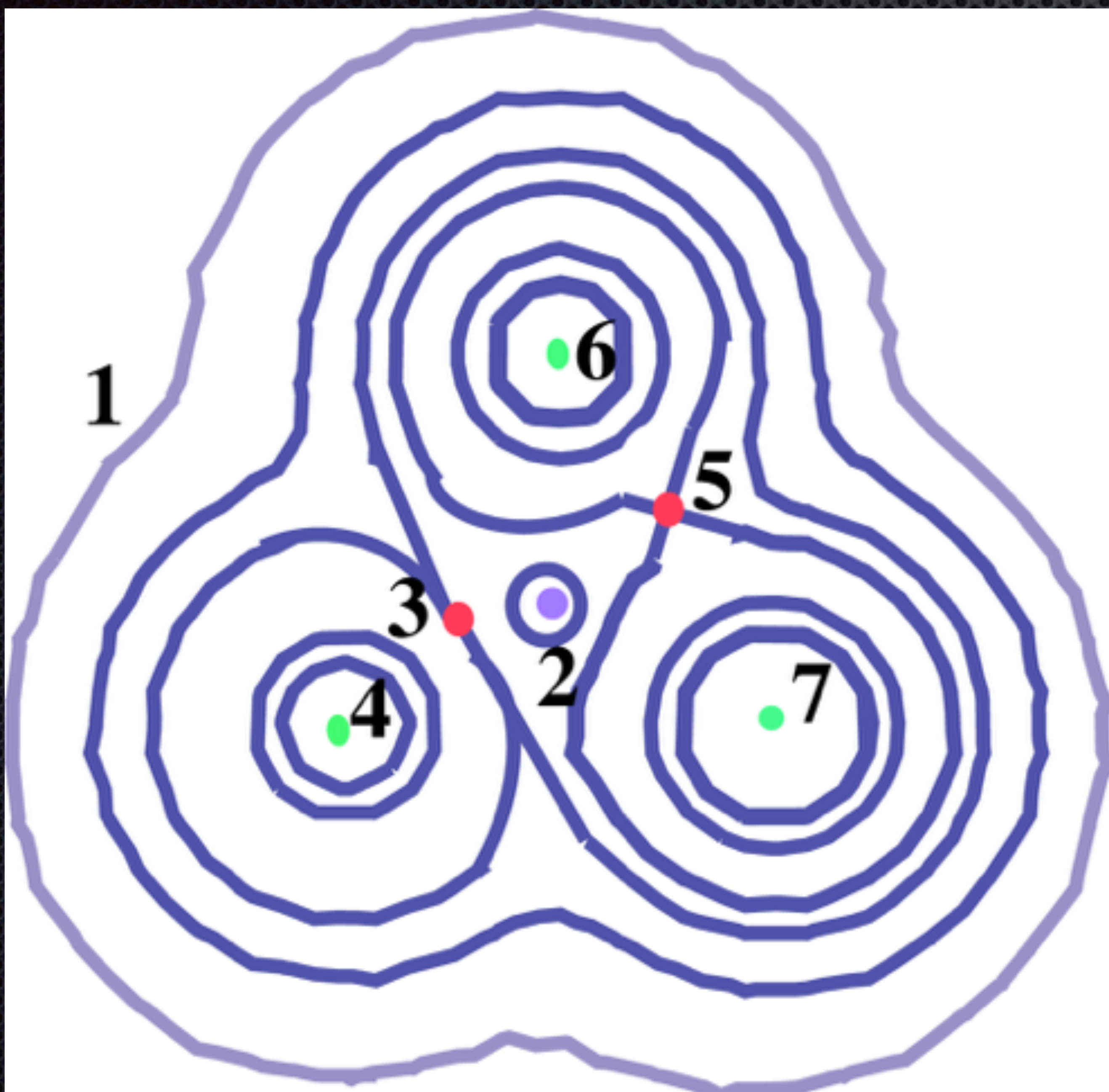
bracket the *m* ciritical values of

$$f : X \to \mathbb{R}.$$

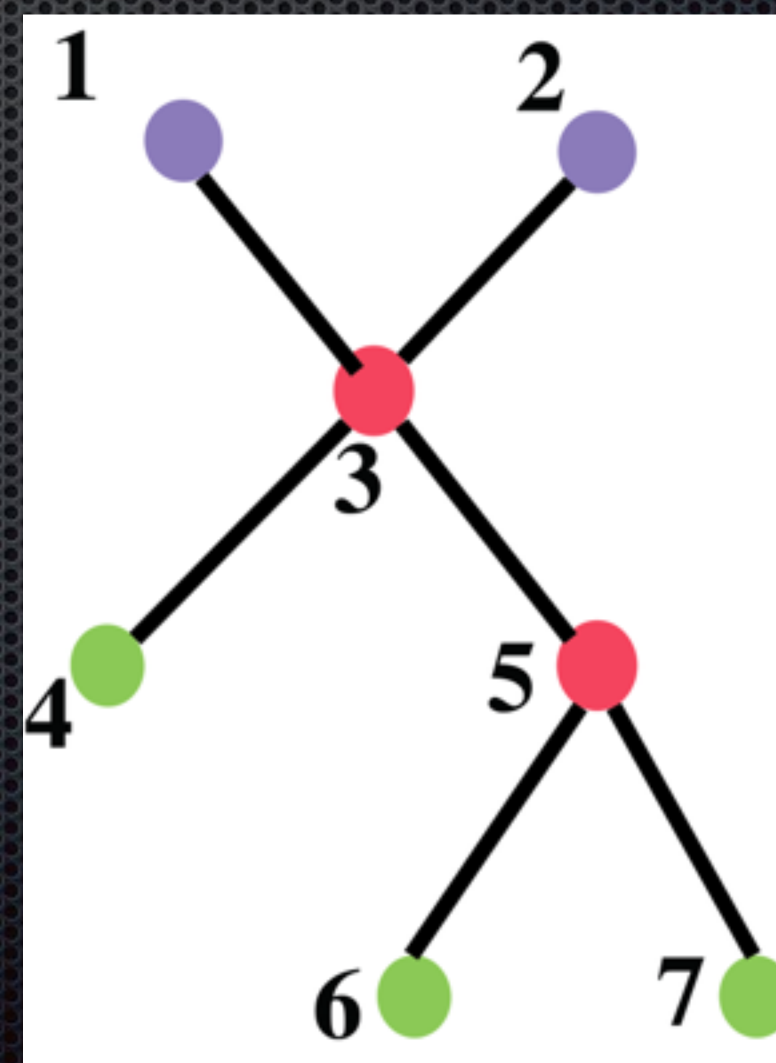
Case dim X=1, persistance tracks $\beta_0(\mathbb{R}_{x_j}) = rank\ \mathcal{H}_0.$

Case dim X>1, persistance tracks $\beta_p(\mathbb{R}_{x_j}) = rank\ \mathcal{H}_p.$

Random fields as information channels - diversity and noise



Topological persistence looks at the topology of sublevel sets.



How the Topology Induced Partition is Related to Topological Entropy

Theorem: (Baronov, 2010) Let $\mathcal{V}_N$ be the domain partition of $f : X \to [0,1]$ corresponding a uniform partition of the range into subintervals of length $1/N$. Let $\mathcal{M}$ be the topology induced partition of the same function. Define:

$$\delta_j = \frac{\sup_{\mathbf{R}\in M_j} f(\mathbf{R}) - \inf_{\mathbf{R}\in M_j} f(\mathbf{R})}{\sup_{\mathbf{R}\in X} f(\mathbf{R}) - \inf_{\mathbf{R}\in X} f(\mathbf{R})}.$$

Then

$$\lim_{N\to\infty} \left(H(f, \mathcal{V}_N) - \log_2 N\right) \leq H(\mathcal{M}) + \sum_{i=1}^{n} \frac{\mu(M_i)}{\mu(X)} \log_2 \delta_i.$$

Information and the topology of unknown fields

- One interpretation of Baronov's theorem is that the critical sets of an unknown field encode the essential information that can be obtained through exploration and mapping.

- This raises the question, will humans engaged in reconnaissance focus on discovering these topological characteristics?

- Another question is whether it is possible to design reconnaissance strategies aimed at discovering the topological characteristic of an unknown field.

Goal: Design algorithms for sequentially refining the topology induced partions in order to climb and information gradient aimed at learning the topology induced partition.

1. Initialize $\mathcal{V}_0 = X$ (hence $H(\mathcal{M}|\mathcal{V}_0) = H(\mathcal{M})$),
2. Refine $\mathcal{V}_k$ at the k-th step such that as $k \to \infty$, $H(\mathcal{M}|\mathcal{V}_k) \to 0$.

Baronov's second theorem and search heuristics

Theorem: (Baronov, 2010) Suppose that

$$\mathrm{Cr}^{even}(f, X) \subset S(B_n),$$

and define the set

$$\mathcal{V}_n' = \{V_n^i \in \mathcal{V}_n : \chi(V_n^i) \leq -1\}.$$

Then $H(\mathcal{M}|\mathcal{V}_n) = H(\mathcal{M}|\mathcal{V}_n')$.

$\Rightarrow$ Initiate motion programs b^{iso} in regions with $\chi < 0$.

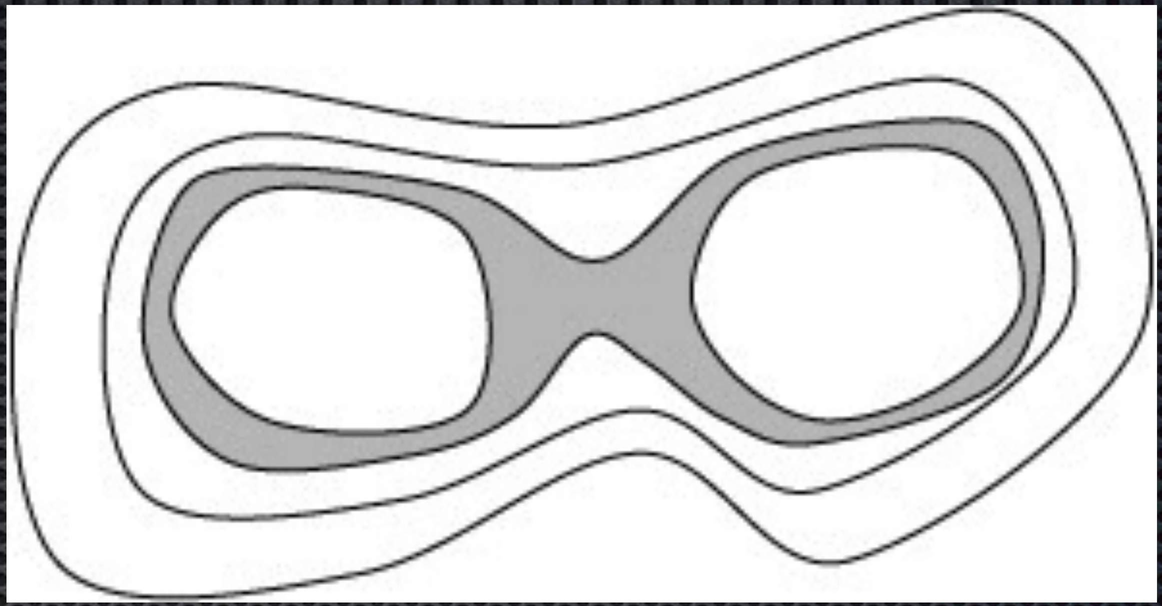
A Proxy for topological entropy conditioned on data:

- $\bar{H}(\mathcal{V}_k) = \sum_{V_k \in \mathcal{V}_k} \frac{\mu(V_k)}{\mu(X)} \log_2 |-2\chi(V_k) + 1|$,
- $H(\mathcal{M}|\mathcal{V}_k) \leq \bar{H}(\mathcal{V}_k)$,
- Mapping subdivides $V \in \mathcal{V}_{k-1}$ with $\chi(V) < 0 \Rightarrow \bar{H}(\mathcal{V}_k) < \bar{H}(\mathcal{V}_{k-1})$,
- $\lim_{k \to \infty} \bar{H}(\mathcal{V}_k) = 0$.

$\Rightarrow$ Initiate motion programs b^{iso} in regions with $\chi < 0$.

Implications of Baronov's second theorem

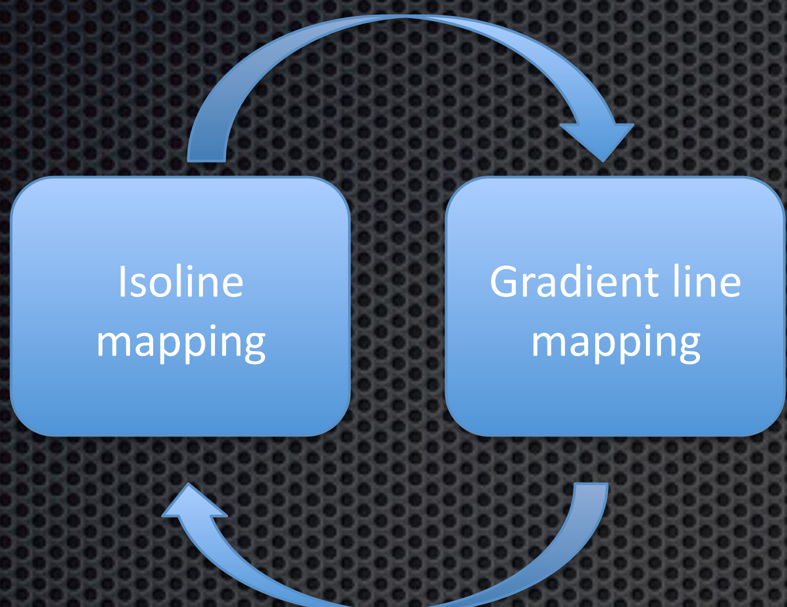
Map isolines only in cells that have Euler characteristic ≤ -1.

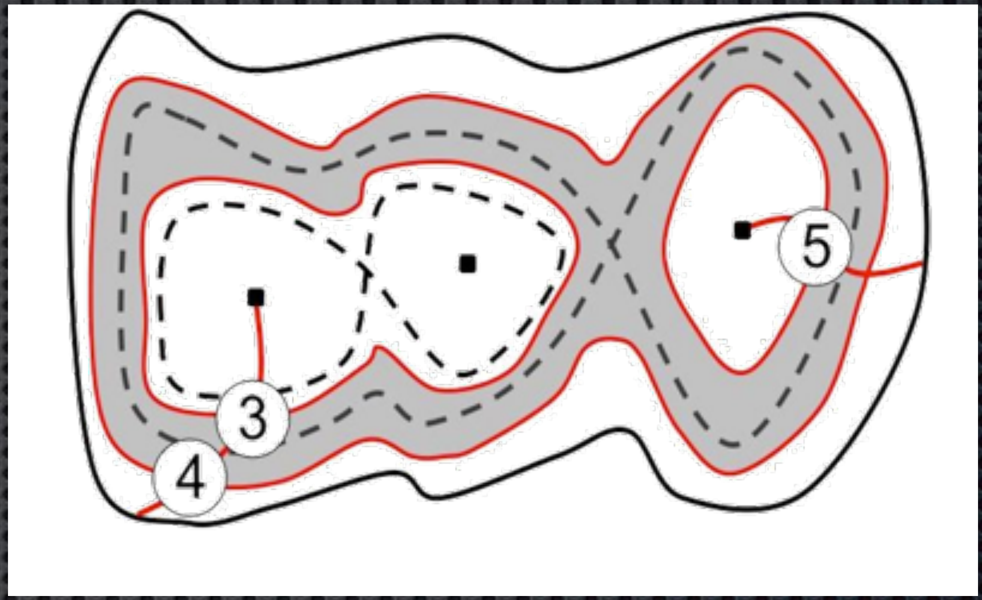


The result: The conditional entropy is strictly decreasing: $H\left(\mathcal{M}|\mathcal{V}_n\right) \rightarrow 0$

Reconnaissance tools

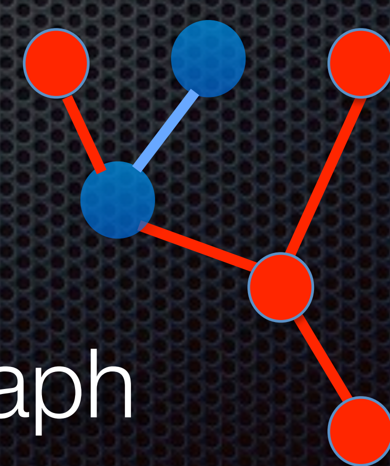
Map gradient lines starting from existing isolines.





Map isolines starting from existing gradient lines.

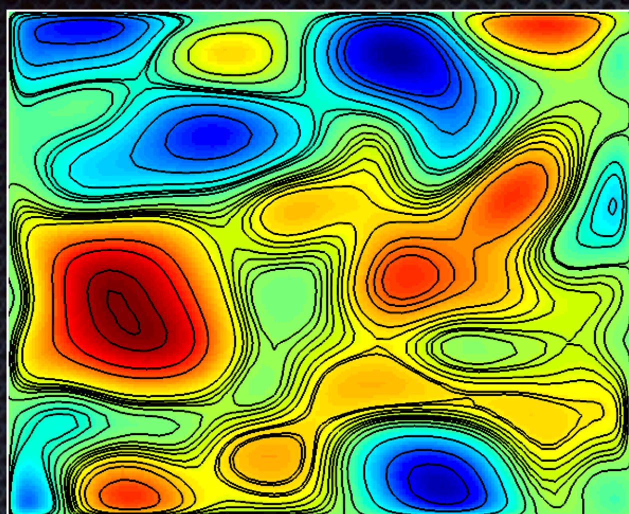
Reeb graph



Nodes = critical level sets

Edges = areas bounded by cobordant level sets

The role of topology in human reconnaissance decisions



Mapped isolines

Black regions have
Euler char. ≤-1

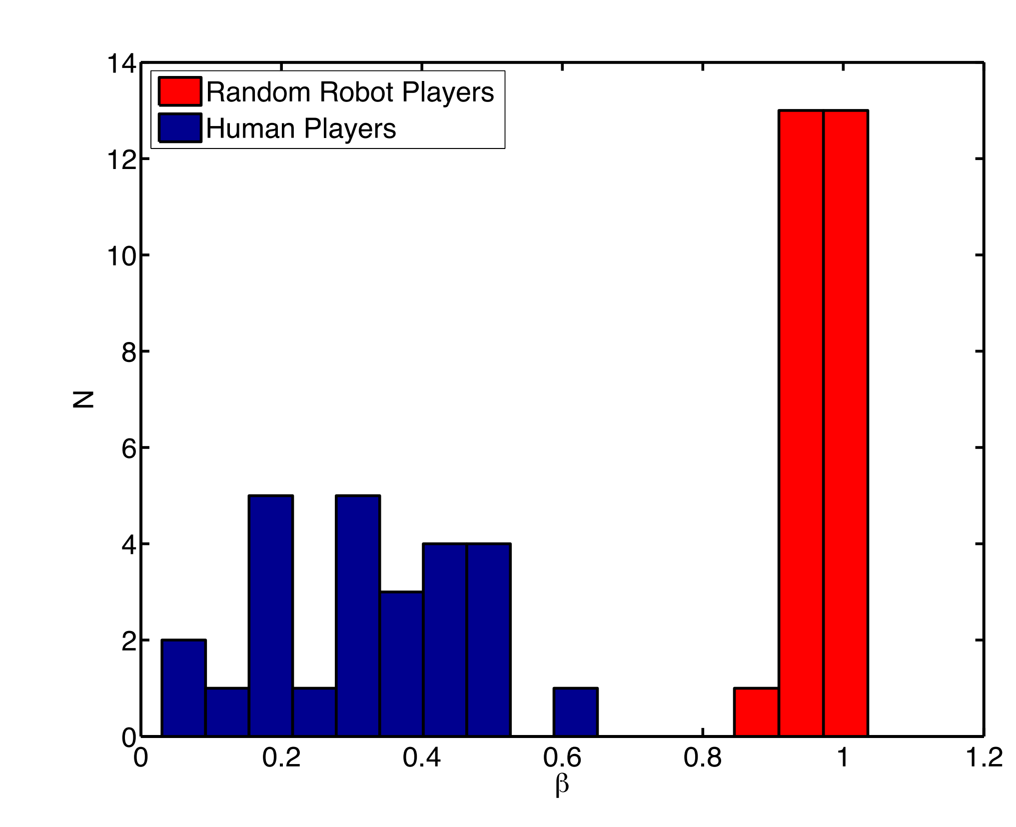
$\mathcal{V}'_n$ Sets with Euler char. less than 0

Hypothesis	Implications	ML estimator
$P^\beta(\mathbf{r}_i) = \left(\frac{\mu(\mathcal{V}'_n)}{\mu(X)}\right)^\beta$	$\beta = 1$ - random $\beta < 1$ - topology based feedback	$\beta = \arg\max_{\beta>0} P(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m)$ $P(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m) = \prod_{i=1}^{m} \begin{cases} P^\beta(\mathbf{r}_i) & \text{if } \mathbf{r}_i \in \mathcal{V}'_{n_i} \\ 1 - P^\beta(\mathbf{r}_i) & \text{else} \end{cases}$

A Game of Reconnaissance

Humans want to discover topological characteristics

- Mean beta for human players: 0.32
- 95% confidence 0.26-0.38
- Human beta is normally distributed with p=0.90 (Kolmogorov-Smirnov test)
- The human players and the random robotic players are statistically distinguished with p=0.00 (Kolmogorov-Smirnov test)



Histogram of the beta characteristics for two groups:
- Human players, and
- Robotic players that randomly map the same number of isolines in the same potential fields

What have we learned from the games - and the theories developed behind them?

- When people play with awareness of others playing simultaneously, *competitiveness* emerges.
- Knowledge of others' reward affects play style more than knowledge of others' play strategies.
- "People do not learn play strategy effectively in competitive situations."
- People seek to discover topological characteristics when acquiring spatial information.
- Some subjects' play styles tend toward *exploitation* while others tend toward *exploration*.

The interplay between topology, geometry and information theory redux

- Biological motion control is guided by perception — not mere reaction to features

- Features registered on the visual cortex are ephemeral (No asymptotic stability!)

- There are simple geometric relationships between an animal's motion through the environment and the neurological replication of that motion on the visual cortex

- Topological persistence is useful in identifying significant features in data sets.

- A similarly useful notion of information saliency also identifies significant features