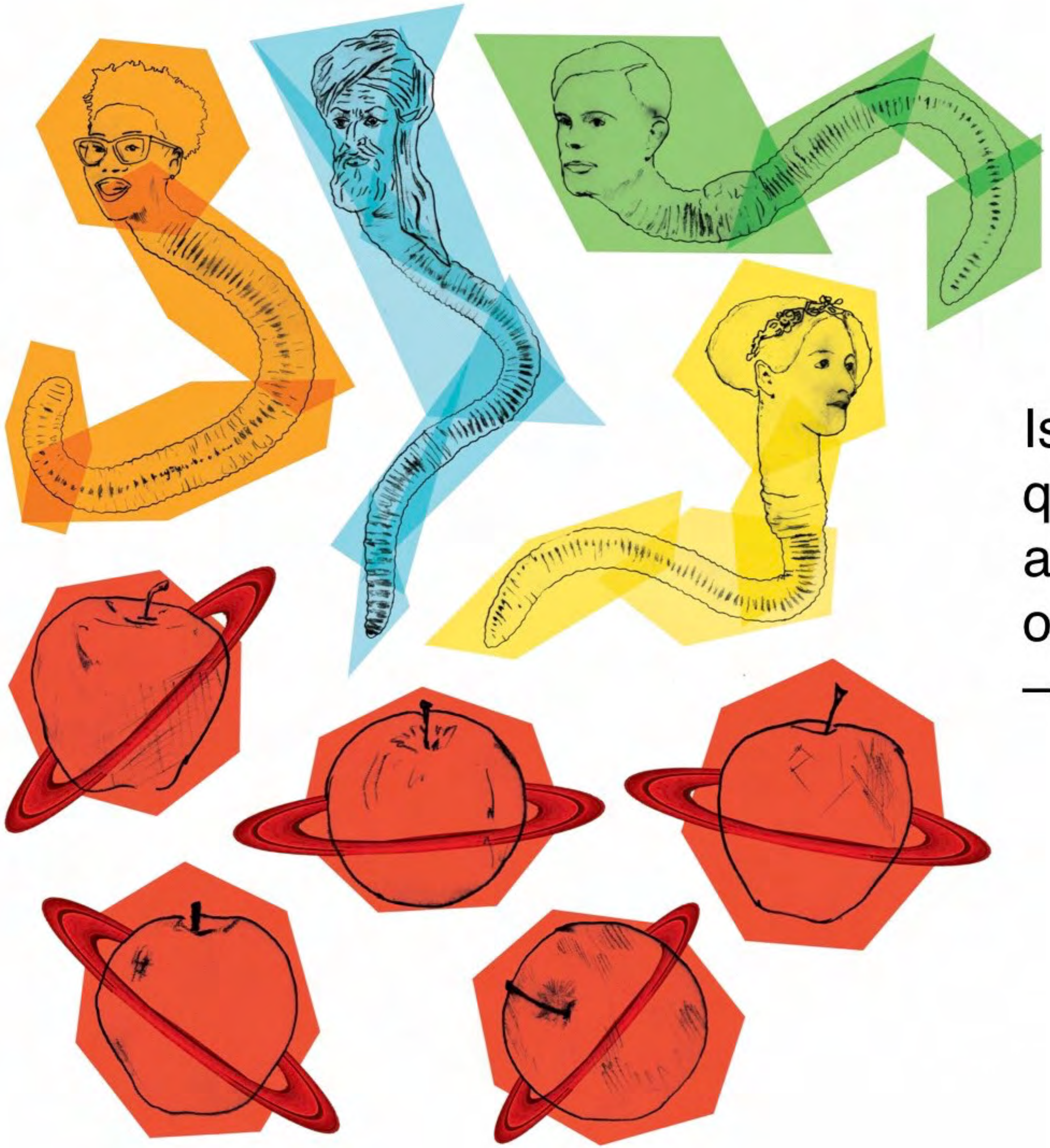
INFRA-STRUCTURES: mathematical choices and truth in data

or…

ALGORITHMS IN SPACE



Is Euclidean Geometry true? The question is nonsense. One might as well ask if the metric system is true and the old measures false…
– Henri Poincaré

ALGORITHMS IN SPACE

- Speculative + provocative + exploratory
- 1st part – I present some ideas / perspectives
- Intermission (surprise)
- 2nd part – we discuss implications and examples

Motivation

- (For everyone): show that behind mathematical assumptions are human choices (values)
    - More tools to challenge ADM (but unreasonable burden?)
- (Esp. for algorithm designers): urge that they acknowledge this, make their choices explicit, state how align with broader goals
    - Advantage of engagement using respected/familiar language

Which algorithms, what mathematics?

- Any ADM systems with human impact
- Decisions involve ranking/ordering, identifying, classifying (individuals, groups, behaviours, trends); optimising
- At level of making comparisons (greater than, close to) rather than procedure or logic (another story, perhaps?)

Topics

- Distance between points
- Size of numbers
- Shape
- (off) Units of measurement – 2019 S.I. redefinition
- (off) Fairness

Aside: values in mathematics

- As a discipline, practice, profession – certainly not value-free
- As a profession: majority of graduates to security agencies, which have a political culture
- Ideally, mathematical objects, arguments, proofs have ‘elegance’
    - Not just correctness, but simplicity, explanatory power, implications for other areas of maths
- Ethics in Maths (Chiodo & Bursill-Hall, forthcoming?)

Mathematical choices

- Canonical, natural

$$ax^2 + bx + c = 0$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Mathematical choices

- Canonical, natural (sure?)

$$ax^2 + bx + c = 0$$

or perhaps

$$x = \frac{2c}{-b \pm \sqrt{b^2 - 4ac}}$$

Mathematical choices

- Canonical, natural
- Conventional, convenient (for whom?)
- Efficient (for what purpose?)
- CONTINGENT

Claims… demand a declaration

- When mangling human data, some – many? – of the mathematical choices we make are contingent
- Choices informed (perhaps unconsciously) by world-view, social/political/cultural beliefs, ~~prejudices, biases~~
- Possibly, no available Archimedean point / god's eye view
- Choices shape outcomes for people
- Algorithm designers: make your low-level choices explicit and own them

1. Distance

- Formally: define distance (metric) on non-empty set of points A by specifying a function d that has to satisfy some conditions.
- Very general and abstract!

$$d : A \times A \longrightarrow \mathbb{R}_{\geqslant 0}$$

$$\forall p, q \in A$$

$$(\text{M1}) \quad d(p,q) \geqslant 0$$

$$d(p,q) = 0 \iff p = q$$

$$(\text{M2}) \quad d(p,q) = d(q,p)$$

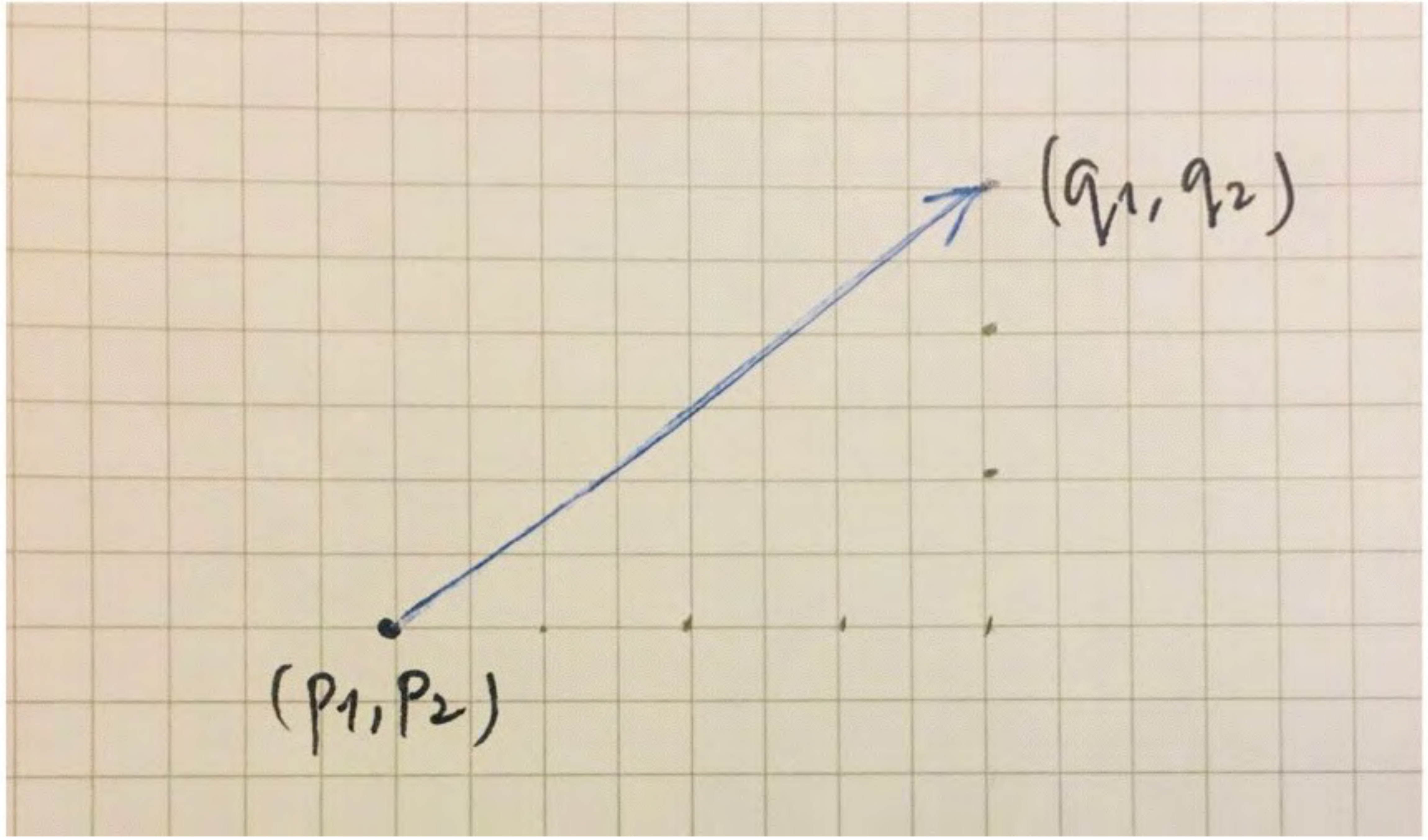
$$(\text{M3}) \quad d(p,r) \leqslant d(p,q) + d(q,r)$$

1. Distance

- The familiar distance function in the plane (Euclidean 2-space) is

$$d : \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}$$

$$d_2(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

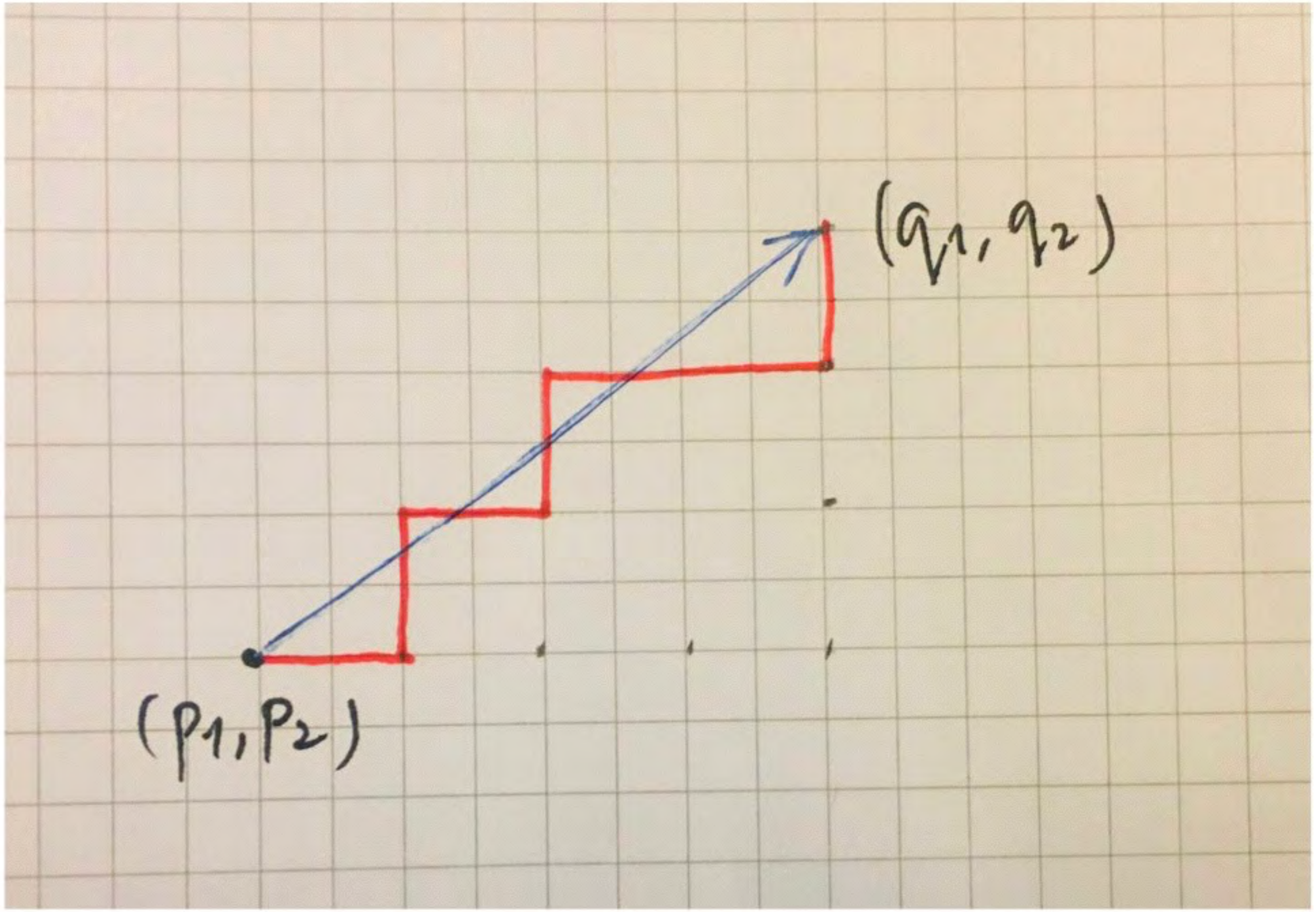


1. Distance

- Here's another equally valid distance function ('Manhattan metric')

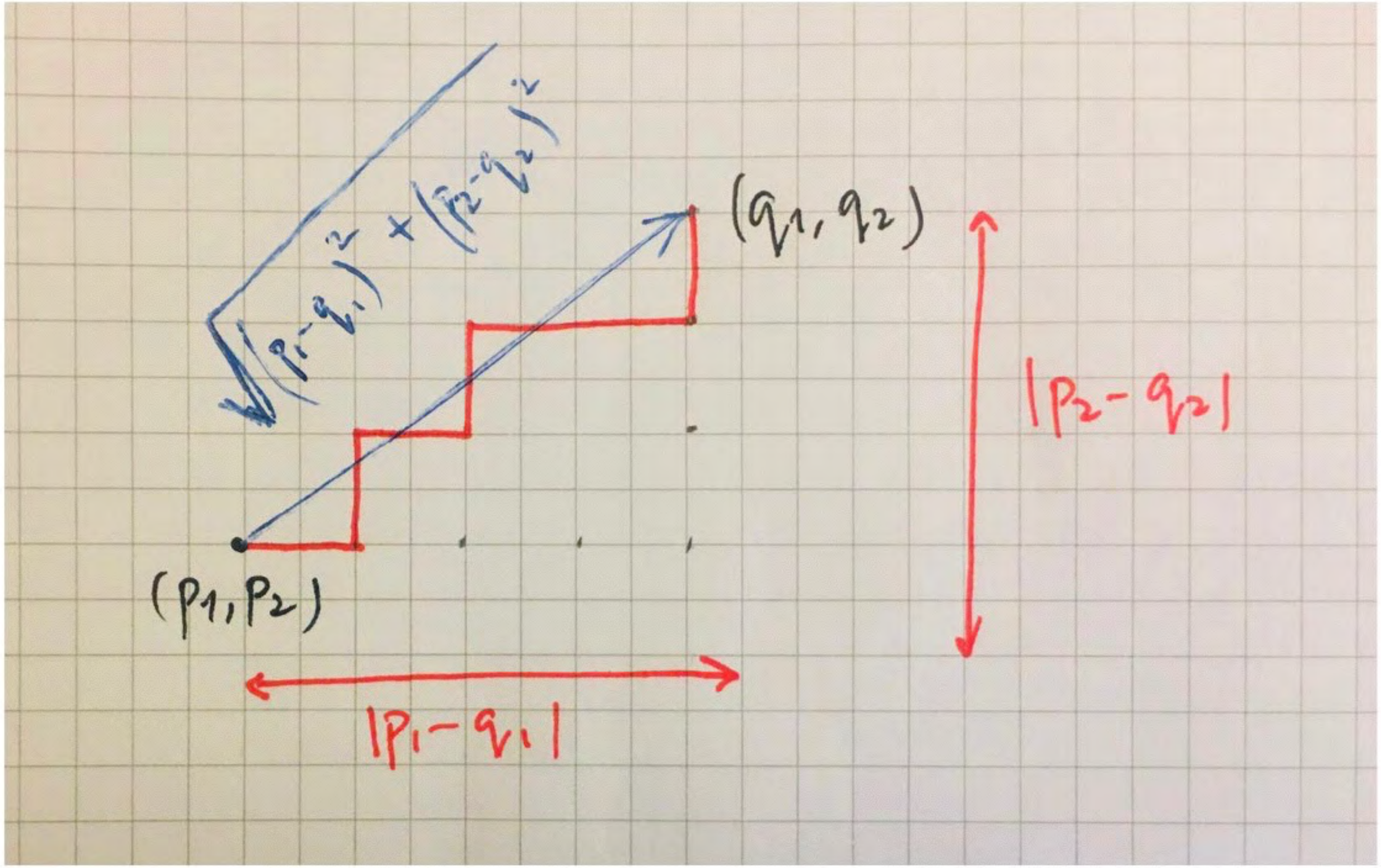
$$d : \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}$$

$$d_1(p, q) = |p_1 - q_1| + |p_2 - q_2|$$



1. Distance

- Points that are equidistant under one metric may not be under another



1.1 Distance in habitat planning

- A better metric – travel time?
    - for whom?

1.1 Distance in habitat planning

- Cyclists and pedestrians sensitive to hills
    - Manhattan (for x-y) and Euclidean (z)?
- Wheelchair and baby buggy users sensitive to pavement surface / width, road works, standing traffic, street furniture, kerbs and ramps
    - Weight for pavement (and air) quality?
- Transit users may be price sensitive (bus vs. tube in London)
    - Travel time modulo cost?

1.1 Distance in habitat planning

- London bus fares are approx. 50% of equivalent tube fare (journey time, from 100–300% of tube)
    - Is even local government fully appreciative of the stark choices facing the daily commuter on a low income?
    - Should you expect a well-paid programmer working under an external contract to be?
- Choice of metric has major implications on geographical distribution and access (to jobs, schools, hospitals…)

1.2 Hamming distance

- Distance between two (bit)strings of identical length
  - 1000 and 1001 – hamming distance 1
  - FAIR and FOUL – hamming distance 3
  - oversight (noun form of ‘overseeing’)
    oversight (noun form of ‘overlooking’)
    have opposite meanings, but hamming distance 0

1.3 Composite distance

- How ‘far apart’ are two people?
- Data: monthly disposable income, last 200 audio tracks streamed, language fluency, # family members attending religious services, …
- Is this ‘solved’, in any sense of ‘solved’ that can be disentangled from a specific instrumental goal?
    - If so, what’s the goal, please?

2. Size (numbers)

2. Size (numbers)

- ‘Modulus’, ‘norm’, ‘magnitude’…
- Formally: absolute value on a field K; function returns a nonnegative real number
- Loosely – ‘what it says on the box’. Apart from sign, nothing more to say…?

$$|\cdot| : K \longrightarrow \mathbb{R}_{\geqslant 0}$$

$$\forall x \in K$$

$$|x| := \begin{cases} x \text{ if } x \geqslant 0 \\ -x \text{ if } x < 0 \end{cases}$$

2.1 Wait, what's a number?

- Counting (natural) numbers

$$\mathbb{N}_0 = \{0, 1, 2, ...\}$$

- Integers

$$\mathbb{Z} = \{0, \pm 1, \pm 2, ...\}$$

- Fractions (rational numbers, field of fractions of integers)

$$\mathbb{Q} = \{(a, b) : a \in \mathbb{Z}, b \in \mathbb{Z} \setminus 0\}$$

(modulo some equivalence relation)

2.1 Wait, what's a number?

- Counting (natural) numbers $\mathbb{N}_0 = \{0, 1, 2, ...\}$
- Integers $\mathbb{Z} = \{0, \pm 1, \pm 2, ...\}$
- Fractions (rational numbers, field of fractions of integers) $\mathbb{Q} \ni \frac{1}{3}, 0.5, ...$
- Real numbers $\mathbb{R} \ni \sqrt{2}, \pi, ...$ $\qquad \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$
- Complex, hypercomplex, transfinite, surreal… 🙊

2.1 Wait, what's a number?

- Pythagoreans: numbers exist in the world as commensurable lengths, i.e. ratios (= fractions) are all there is
- Someone said "but $\sqrt{2}$" and got drowned (apparently)
- Question: what kind of number are we dealing with when we measure (say, with a ruler) and write down a length? (rational approx. to a real + an error term?)

2.2 How computers represent numbers

- Can write an integer thousands of digits long
- Some real numbers (irrational numbers like $\sqrt{2}$) have non-terminating, non-repeating decimal expansions
    - actually, ‘most’ real numbers – Cantor
- Real computers are finite (have finite storage) and therefore limited precision
- Problem?

2.2 How computers represent numbers

- For special applications (research maths, number theory), computers can represent numbers to arbitrary precision (e.g. ’exact real arithmetic’, symbolic algebra systems)

- Everywhere else – numbers are represented to some fixed maximum level of precision (which is reduced during arithmetical operations)

2.2 How computers represent numbers

- IEEE standard double-precision binary floats
  (64 bits: 52 for mantissa, 11 exponent, 1 sign)

  - max. # that can be represented is approx. 1.8 x 10^{308}
  - min. (positive) # approx. 2.2 x 10^{-308}
  - total # of different #s that can be represented =
    2 (signs) x 2046 exponents x 2^{52} mantissas, + 2 signed zeros + 2 infinities + NaN etc.
    which is big, but finite
  - precision 16 sig. figs. (decimal)

2.2 Aside: precision

- Double precision floating point numbers seem to have plenty of sig. figs. for ‘real world applications’
- But precision is reduced (often alarmingly) during mathematical operations
- Many pitfalls for naive programmers:
  - $1 + (10^{17} - 10^{17}) = 1$
  - $(1 + 10^{17}) - 10^{17} = 0$

2.2 Aside: precision

- Errors can be easily and directly fatal:
  - One-tenth = 0.1 is a repeating fraction in binary 0.0001100110011… (like one-third in decimal)
  - 25 Feb 1991 – rounding errors in US Patriot missile system (whose clock ticks every 0.1s) built up to 0.34s which caused it to miss incoming Scud and led to 25 fatalities
    [US General Accounting Office. Patriot Missile Defence: Software Problem Led to System Failure at Dhahran, Saudi Arabia. IMTEC-92-26. Feb 4, 1992]

2.2 Aside: precision

- Errors can be easily and directly fatal:
  - Recall two versions of formula for roots of quadratic equation
  - Iraqi targeters used version that gave imprecise value for smaller root and sent missiles into civilian area
    [J. Mestel, ICL ref. to follow]
- Controlling precision is not a trivial problem

2.2 How computers represent numbers

- Claim: computers represent only rational numbers
- Counter: computers represent all (computable) irrational numbers, as rational numbers plus an error term
    - Warrants deeper discussion. All algebraic numbers, π, e (i.e. all relevant real numbers?) are computable. The problem of irrational numbers that are not computable, remains.

2.2 How computers represent numbers

- Claim: computers represent only rational numbers
- How do measure the size of a number, again?

2.3 Absolute value

$$|\cdot| : K \longrightarrow \mathbb{R}_{\geqslant 0}$$

$$\forall x \in K$$

$$|x| := \begin{cases} x \text{ if } x \geqslant 0 \\ -x \text{ if } x < 0 \end{cases}$$

2.3 Absolute value

$$\forall x, y \in \mathbb{R}$$

$$|x| \geqslant 0$$

$$|x| = 0 \iff x = 0$$

$$|xy| = |x|\,|y|$$

$$|x + y| \leqslant |x| + |y|$$

2.3 Absolute value

- Earlier we chose among different metrics (distance functions) – the 'usual' Euclidean, Manhattan, Hamming,…
- We value numbers with the 'usual' absolute valuation $| \cdot |_\infty$
- But here too we have a choice…

2.3 Absolute value

- Theorem (Ostrowski, 1916):

  Every non-trivial absolute value on the field of rational numbers is (equivalent to) the usual absolute value or the p-adic absolute value (for some prime number p).

2.3 Absolute value

- The p-adic world is strange (but just as good as the usual one).
- ‘p’ in p-adic stands for a (particular) prime, and p-adic valuations tell us about divisibility by p. In the 3-adic world, for example, the size of a number tells us about its divisibility by 3.

2.3 Absolute value

$$|\cdot|_p : \mathbb{Q} \longrightarrow \mathbb{R}_{\geqslant 0}$$

$\forall a \in \mathbb{Q}, a \neq 0,$ write $a = p^m \dfrac{b}{c}$ where p is prime and $b, c \in \mathbb{Z}$ are coprime to p

then $|a|_p := p^{-m}$ and $|0|_p := 0$

2.3 Absolute value

$$36 = 3^2 \frac{4}{1} \quad \text{so} \quad |36|_3 = 3^{-2} = \frac{1}{9}$$

$$81 = 3^4 \frac{1}{1} \quad \text{so} \quad |81|_3 = 3^{-4} = \frac{1}{81}$$

$$\text{while } 80 = 3^0 \frac{80}{1} \quad \text{so} \quad |80|_3 = 3^0 = 1$$

2.4 Absolute provocation

- Computers represent rational numbers.
- (Ostrowski) The p-adic absolute value is mathematically as valid as the usual absolute value on rational numbers.
- The choice of absolute value has a dramatic effect on how large (and how close together) numbers are.
- There had better be a good (human) reason for choosing an absolute value on data that has human impact.
- In particular, when the data is highly abstract/composite (social credit?), the usual absolute value has no ‘natural’ precedence.

3. Space

- (Given earlier discussion of metrics, abs. vals.) does data have a ‘natural’ structure?
    - Does structure reflect causal relationships?
- Is there a ‘natural’ space in which to locate / visualise / interpret data points?
- Choice of space (like that of metric or abs. val.) can transform relationships between points

3. Space

- Consider ‘abstract’ data e.g. people’s political affiliation (not their heights)

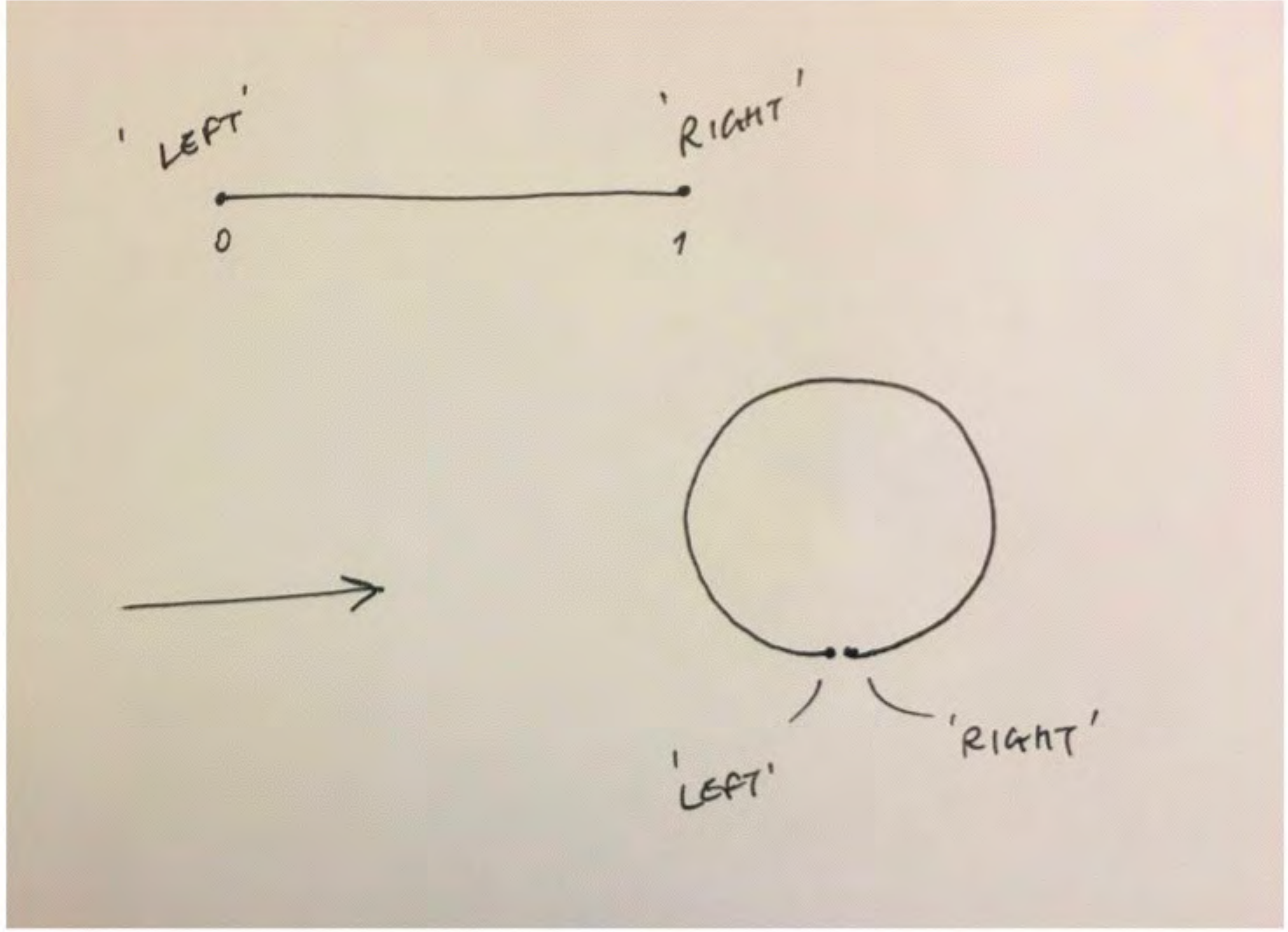
- Suppose it’s one-dimensional, i.e.

LEFT WING———————————————RIGHT WING

- Even if a single dimension is sufficient to capture the data, we can choose how to visualise / interpret.

3. Space

- Identifying end points of the interval (line segment) [0,1] gives us a circle



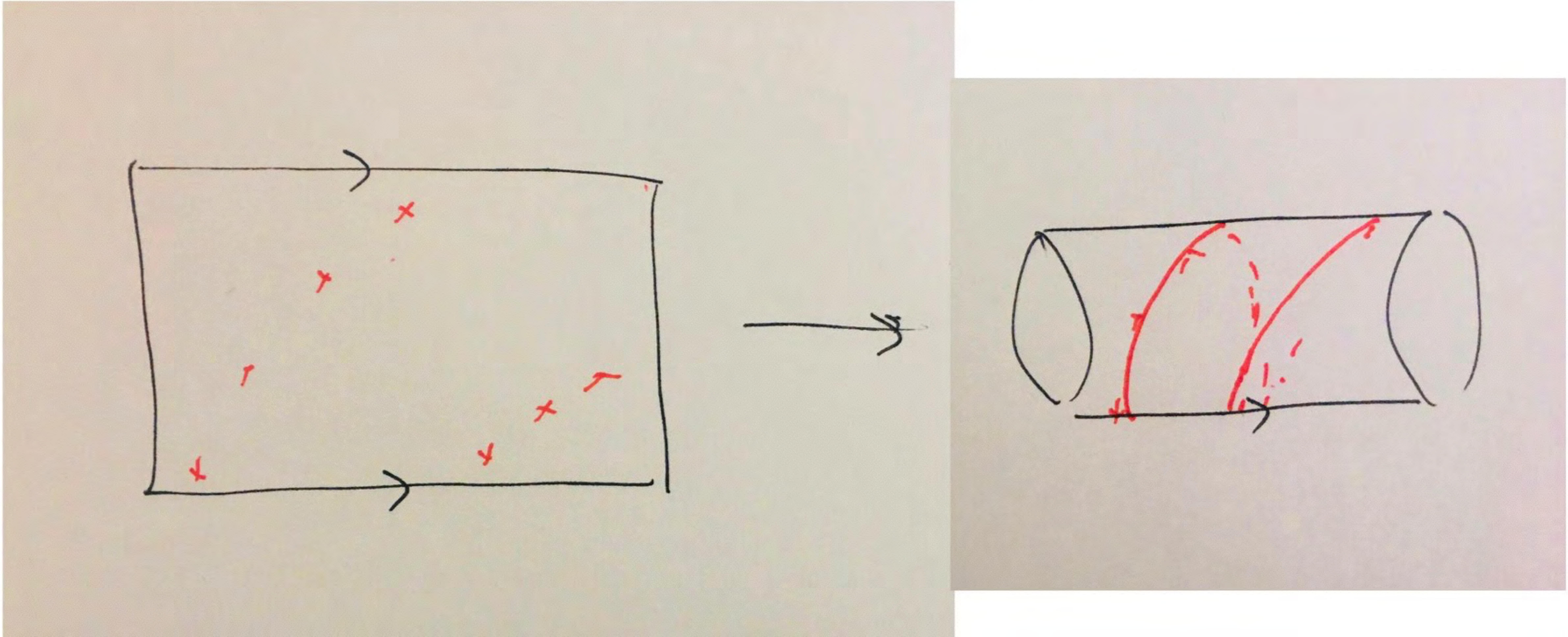
- But political affiliation is clearly more than one-dimensional…

3. Space

- Circular (as opposed to linear) statistics – support is unit circle, not interval. Example – bearings
    - 359° is very close to 1°
    - mean of 4 northerly bearings 350°, 355°, 5°, 10° is 180°

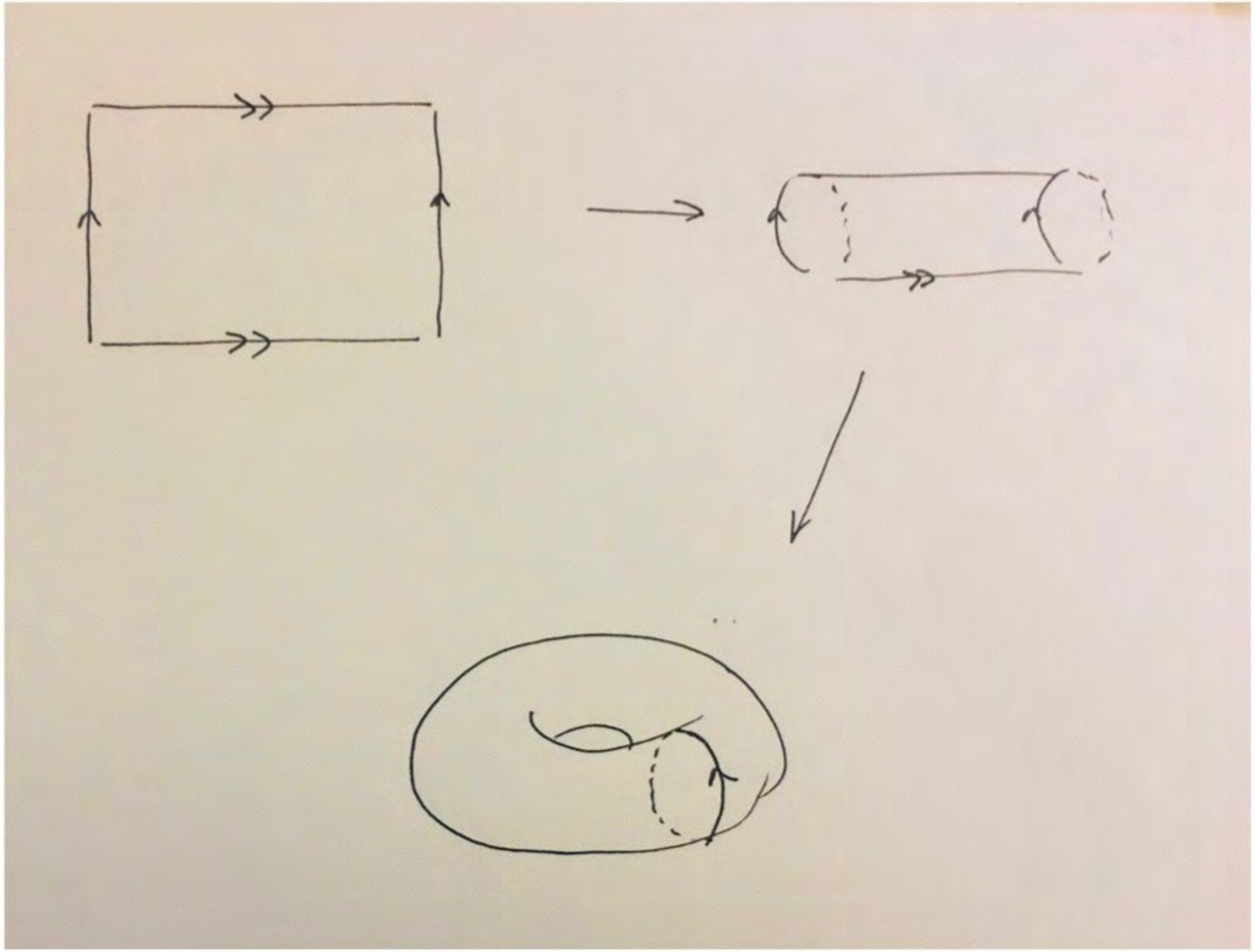
3. Space

- Periodic data – on a cylinder; two periodic variables on a torus



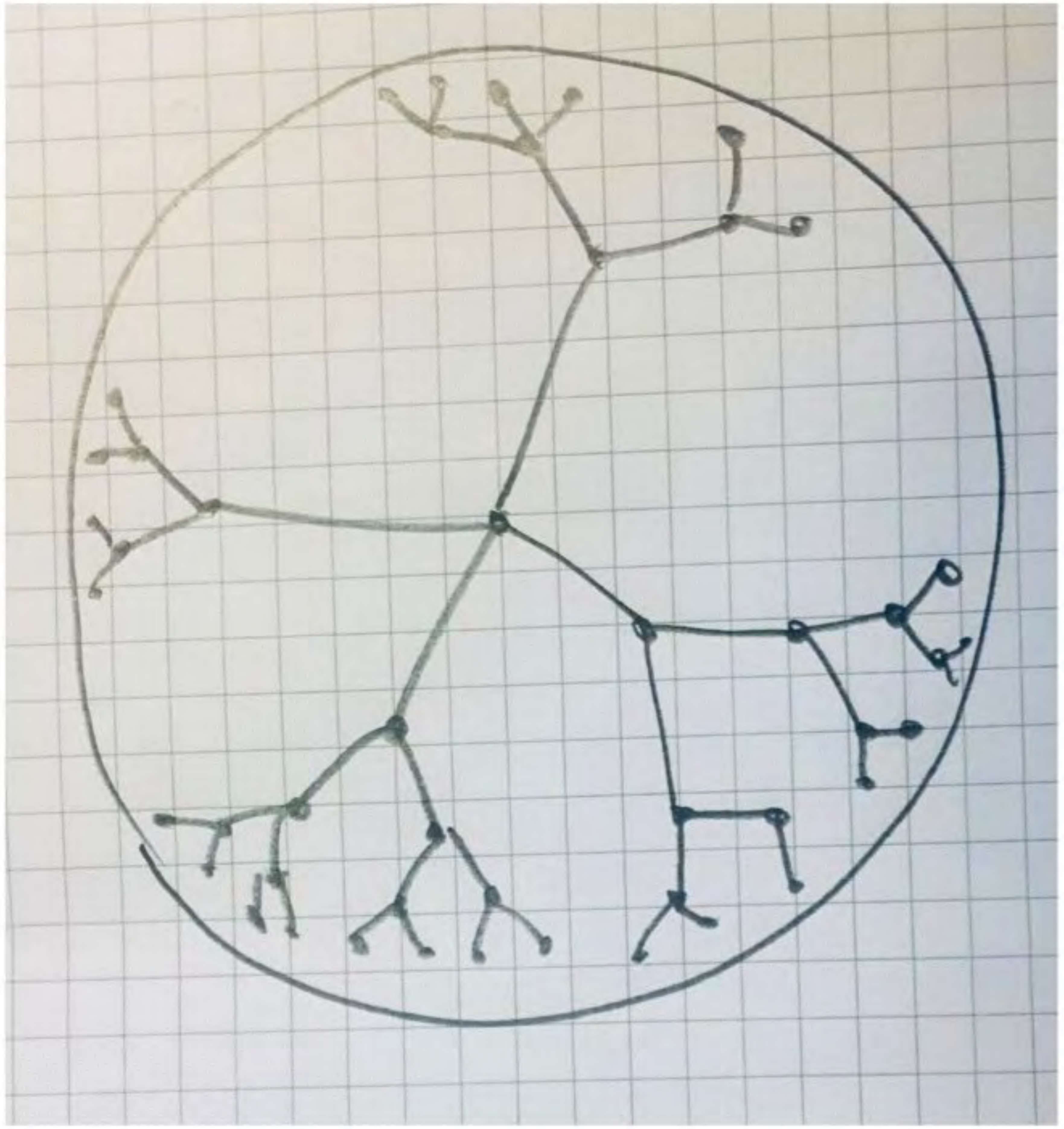
3. Space

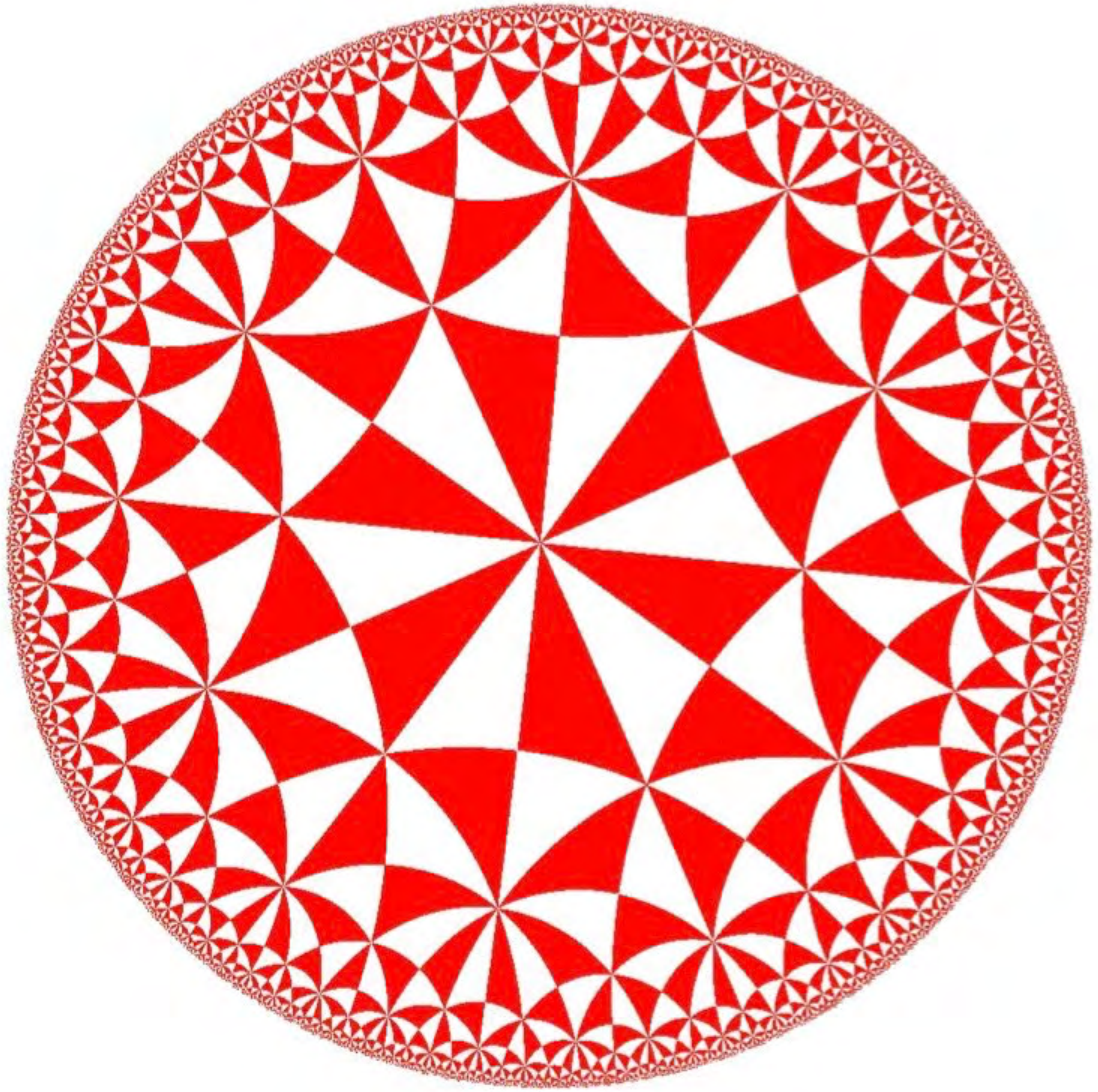
- Periodic data – two periodic variables on a torus



3. Space

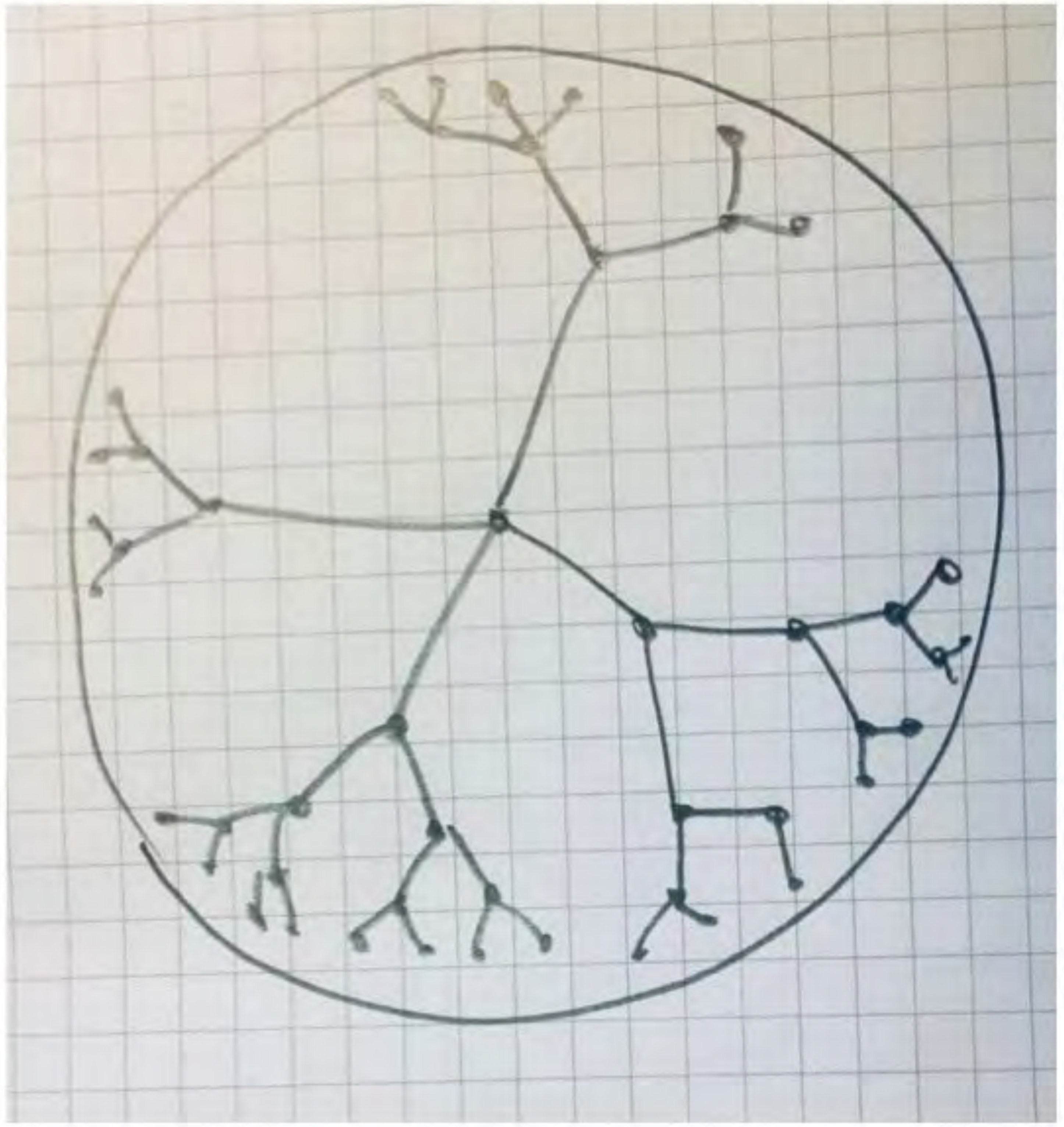
- Hierarchical data (trees) in hyperbolic space

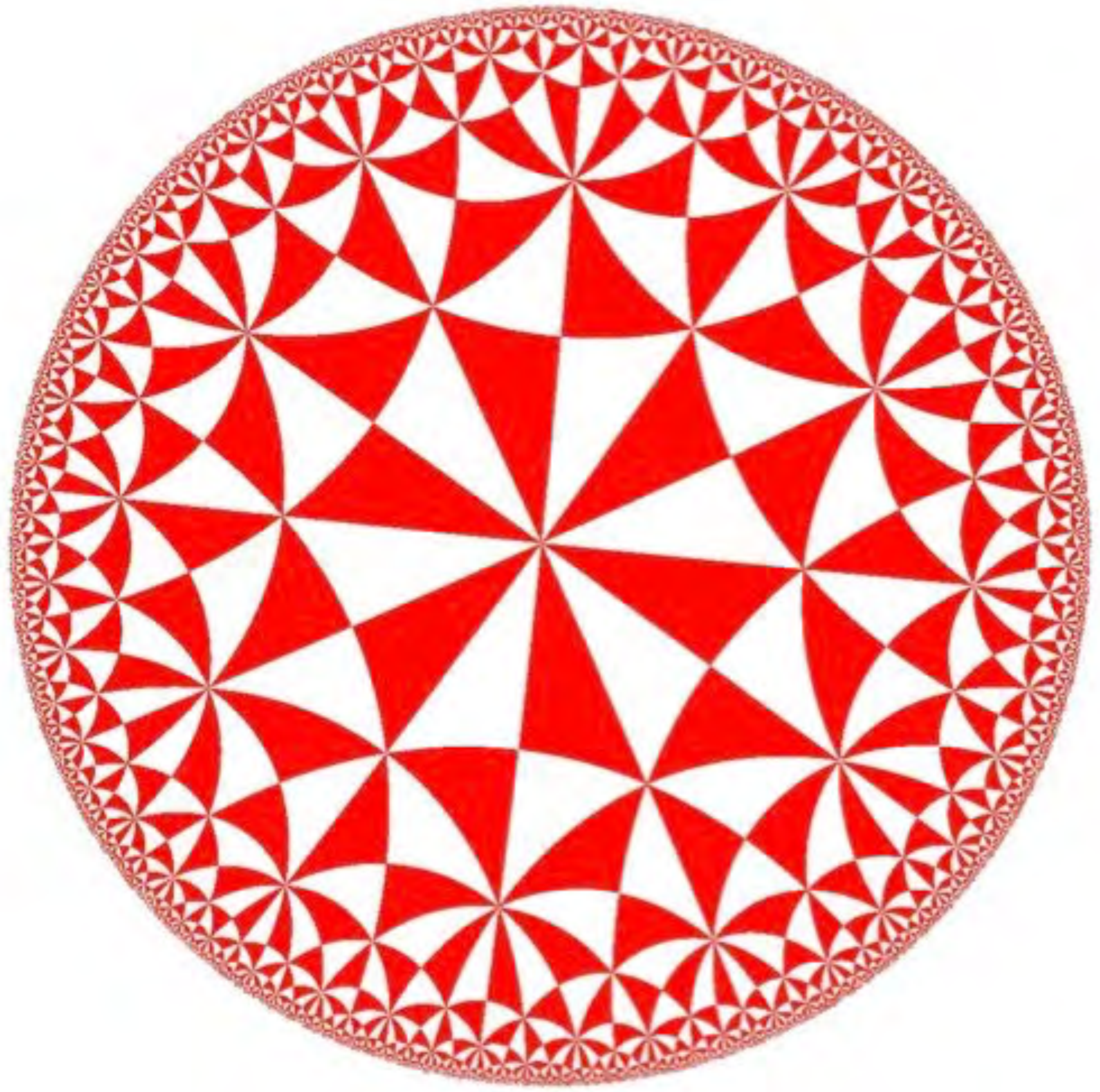




3. Space

- Hyperbolic space is a continuous analogue of hierarchical tree structure, so data with such structure should embed well and predictive/generalisation power should be high.





3. Space

- M. Nickel & D. Kiela (2017). ‘Poincaré Embeddings for Learning Hierarchical Representations.’ https://papers.nips.cc/paper/7213-poincare-embeddings-for-learning-hierarchical-representations.pdf

- Embed the data in different spaces an then reconstruct it, and compare the errors…

3. Space

- Hierarchical data might embed better in hyperbolic space but is not structured uniformly. Better results in combined (product) spaces? A. Gu, F. Sala, B. Gunel, C. Ré (2019). 'Learning mixed-curvature representations in products of model spaces.' Conference paper https://openreview.net/pdf?id=HJxeWnCcF7

4. New philosophy of measurement

- 2018: Redefinition of S.I. base units in terms of physical constants
- Theory: universal physical constants (eg. c) fixed for all time and places.
  Actual numerical values depends on definition of units.
    - Historically: units defined with reference to a standard unit (e.g. metre: convenient fraction of the Earth's circumference; later, with ref. to physical artefact)
    - Numerical values for physical units were then determinedly experimentally.
    - But 'standards' (IPK International Prototype Kilogram, etc.) drifted…
- In the 1970s, instruments to measure c became accurate to less than 1 m s^{-1}. The metre was redefined in terms of an explicit constant in 1983, when c was fixed to have a value of 299 792 458 m s^{-1}.

4. New philosophy of measurement

- Now: explicitly fix the numerical value of the universal physical constants

- Define the unit in terms of this invariant (implicit definition).

TIME: The second (s) is defined by fixing $\Delta\nu_{Cs}$, the unperturbed ground-state hyperfine transition frequency of the caesium-133 atom, to be 9 192 631 770 Hz (i.e. 9 192 631 770 s^{-1}).

LENGTH: The metre (m) is defined by fixing c, the speed of light in vacuum, to be 299 792 458 m s^{-1}, where the second is defined in terms of $\Delta\nu_{Cs}$,.

4. New philosophy of measurement

- $\Delta\nu_{Cs}$ never changes. It doesn't really matter what numerical value we assign to it, because the size of the unit we measure it in is arbitrary. However, we do have an existing, convenient unit – the second – that we wish to refine, so we assign a value that makes the newly defined second as close as possible to the second by the old definition. Instead of having a fixed definition of the second, and measuring $\Delta\nu_{Cs}$ relative to it, we fix a convenient number for $\Delta\nu_{Cs}$ and define the second relative to that.

5. Fairness

- Another angle one could pursue: define fairness mathematically?
  - Bi-Lipschitz map between two metric spaces (control space and decision space) for some ‘well-chosen’ epsilon and delta Topological argument understood by approx. 3.5 people (Maurice Chiodo, Ethics in Maths / Univ. Cambridge has details)

5. Fairness

- Can we even decide fairly between competing conceptions in a voting process?
  - Arrow's Impossibility Theorem: Given at least 3 candidates, then the 6 axioms of any reasonable voting system are inconsistent.
    V. short proof by Terence Tao at https://www.math.ucla.edu/~tao/arrow.pdf
  - Continuing relevance? Peter Eckersley. 'Impossibility and Uncertainty Theorems in AI Value Alignment, or why your AGI shouldn't have a utility function.' Partnership on AI & EFF. https://arxiv.org/abs/1901.00064
  - Link to computability: H. Reiju Mihara (1997). 'Arrow's Theorem and Turing Compatibility'. Economic Theory 10, 257-276.

    Another angle:
    J. Kleinberg, S. Mullainathan, M. Raghavan (2016). 'Inherent Trade-Offs in the Fair Determination of Risk Scores.' https://arxiv.org/abs/1609.05807

To summarise

- Fundamental mathematical choices (how to measure distance, size; what space to inhabit) may be contingent, esp. where data is abstract
- Data may not have ‘natural’ structure reflective of causal relationships
- Where choices are contingent, should
    - (1) be made with cognisance of potential impact of algorithm output
    - (2) be declared explicitly
- Discussion – in particular, seeking data sets to stretch across different metrics / evaluate with different abs. value / plot in different spaces

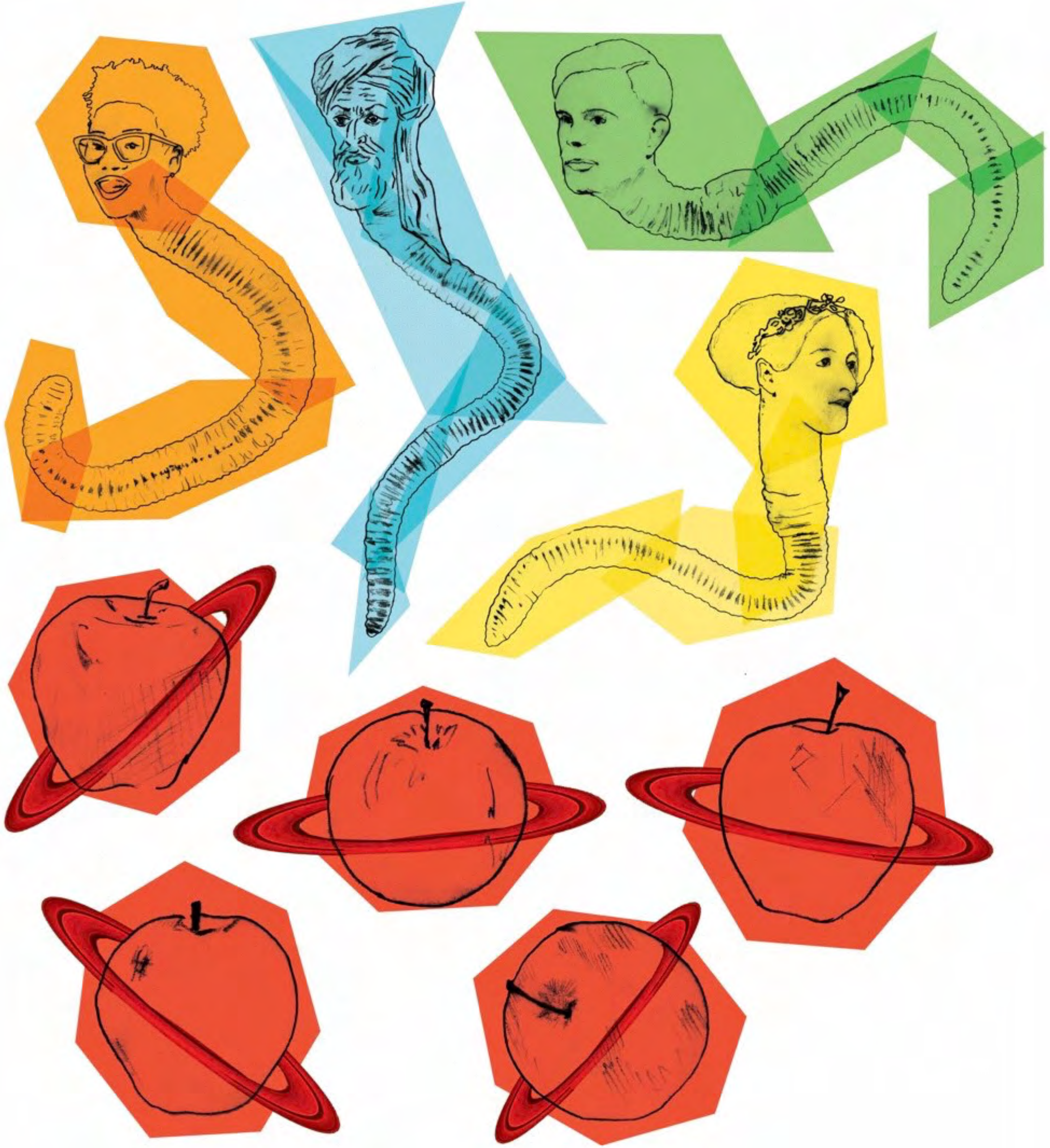
Where now?

- Gut instinct – technical-philosophical arguments could work well with CS/ algorithm designers

- But even different versions of absolute value too complex to bring up in policy contexts.
  (UK Govt. House of Lords, #bishops = 6 ≫ #mathematicians = 2)

- Wider public requires something more immediately graspable.

- Different metrics, embeddings of data more tangible. In a media space that rejects extended/detailed argument, better strategy to be silent and show. Visual, physical models of data in different space could be compelling.



INTERMISSION

Session Tote Bag!

(and puzzle: connect each AI-worm
to each apple with a direct line without crossings).

A clue…