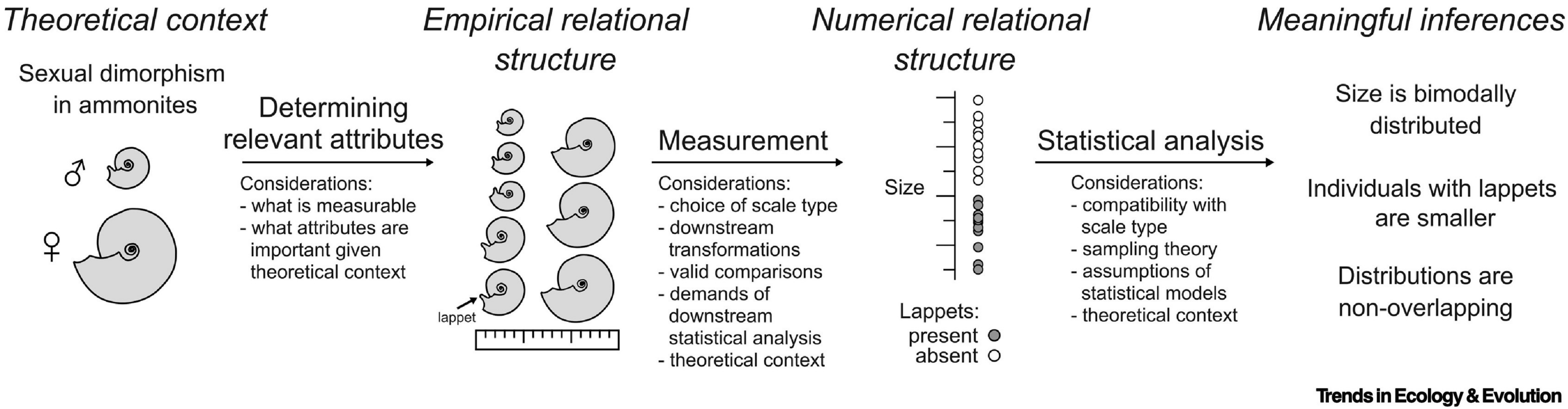
Measurement and meaning in statistics and biology

BIO708, February 13th

Ian Dworkin (dworkin “at” mcmaster.ca)



Voje et al. Trends in Ecology & Evolution 2023 381165-1176 DOI: (10.1016/j.tree.2023.08.005)

Goals for this week

- Making estimates ***meaningful***: What you measure, and how it impacts how you model data, and derive meaning from model estimates.

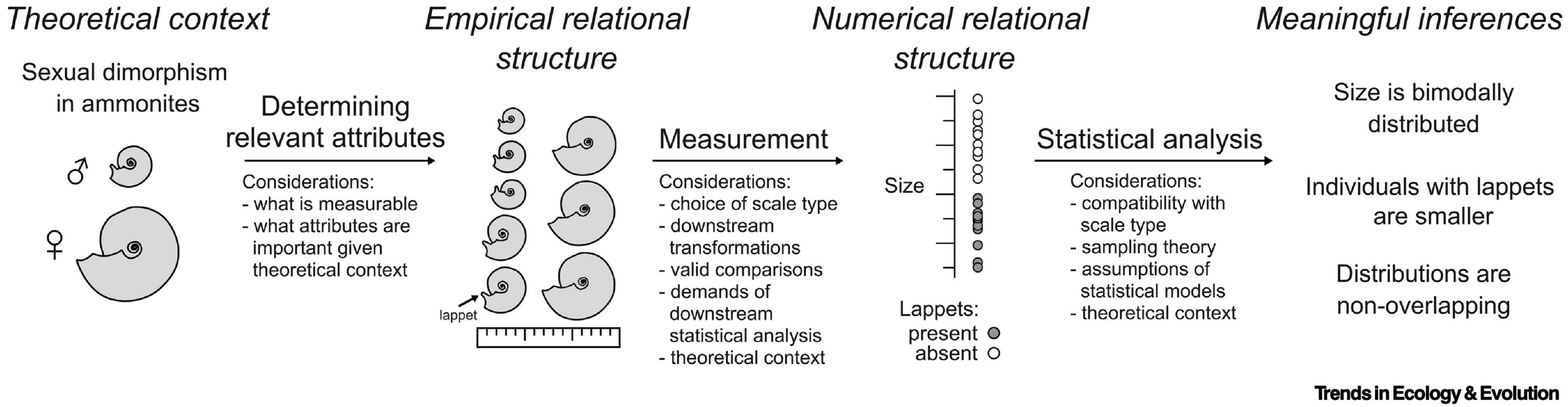
- Making estimates ***interpretable***: How simple “transformations” like mean-centring and standardization (*z*-transformation) predictor variables can simplify

- Making estimates ***comparable***: How to make your model estimates (coefficients) both within your experiment, and (hopefully) to enable meaningful comparisons with the broader literature.

Meaningful, interpretable and comparable estimates

- What we are discussing applies equally to "simple" models like a comparison between two groups, a simple linear regression or a complex linear model with multiple predictors with interactions.

- The work you do BEFORE you do any data modeling to help you focus on meaningful, interpretable and comparable estimates will likely help you a great deal in terms of developing the appropriate model as well.



Voje et al. Trends in Ecology & Evolution 2023 381165-1176
DOI: (10.1016/j.tree.2023.08.005)

What’s wrong with this statement?

- It was 20℃ yesterday, it will be 40℃ tomorrow, therefore it will be twice as hot.

What's wrong with this statement?

- It was 20℃ yesterday, it will be 40℃ tomorrow, therefore it will be twice as hot.

- Not true in Fahrenheit or Kelvins. That is, the statement is only true in C, but not in F or K.

- ***If it is not true for all these different scales of measurement for temperature, does the statement have any meaning at all?***

What do you think of Tim Horton’s coffee on a scale of 0-5?

- 0: ”Are you sure this isn’t toxic?”
- 1: “unpalatable”
- 2: “meh”
- 3: “it’s coffee”
- 4: “pretty good”
- 5: “As good as the finest Ethiopian Expresso”

*Statistics themselves are just 'machines' for producing results **– they have no understanding of the meaning of a measurement such as '5'** and whether this measurement is a category labeled '5' (nominal scale), whether it denotes a larger measurement than measurements labeled '4' and a smaller measurement than those labeled '6' (ordinal scale), or whether it means the measurement is one unit larger and one unit smaller than 4 and 6, respectively (ratio scale).*

Thus, the scale type to which the data belongs can have monumental consequences for the type of statistical ~~tests~~ (modeling) appropriate to perform and whether the results are meaningful.

Voje et al. 2023 (my small revision in blue)

We want inferences from models to reflect aspects of reality. *Do they?*

- We may be able to do statistics on any set of numbers.
  - (i.e. we can feed them into R, and can get a result).
- However, if we do not understand the source of those numbers, the results from the statistical analysis may be meaningless.
  - (The computer does not know the meaning of the numbers you input)

Measurement theory and statistics

- This has little to do with what "valid" statistical methods and inference can be applied.
- Instead, it provides a framework to make sure that what we include in our model reflects reality and can be interpreted.
- If we are interested in arbitrary scales for their own sake (the food critic example) then measurement theory does not matter so much.
- If we want to relate this scale to taste.... Then it matters a great deal.

From Cumming & Calin-Jageman 2017: *Introduction to the new statistics*

Table 2.1 The Four Levels of Measurement, NOIR

Level	Feature	Allowable operations	Examples
Nominal	Same/different	Count frequencies	Gender, ice cream flavors
Ordinal	Order	As nominal, plus: arrange in order	Ranking of colleges
Interval	Distance	As ordinal, plus: add and subtract, calculate mean	Birth year, longitude
Ratio	True zero	As interval, plus: calculate ratio and percentage	Height, mass, time to complete a task

The premise of representational measurement theory is that measurement consists of the assignment of numbers to empirical entities

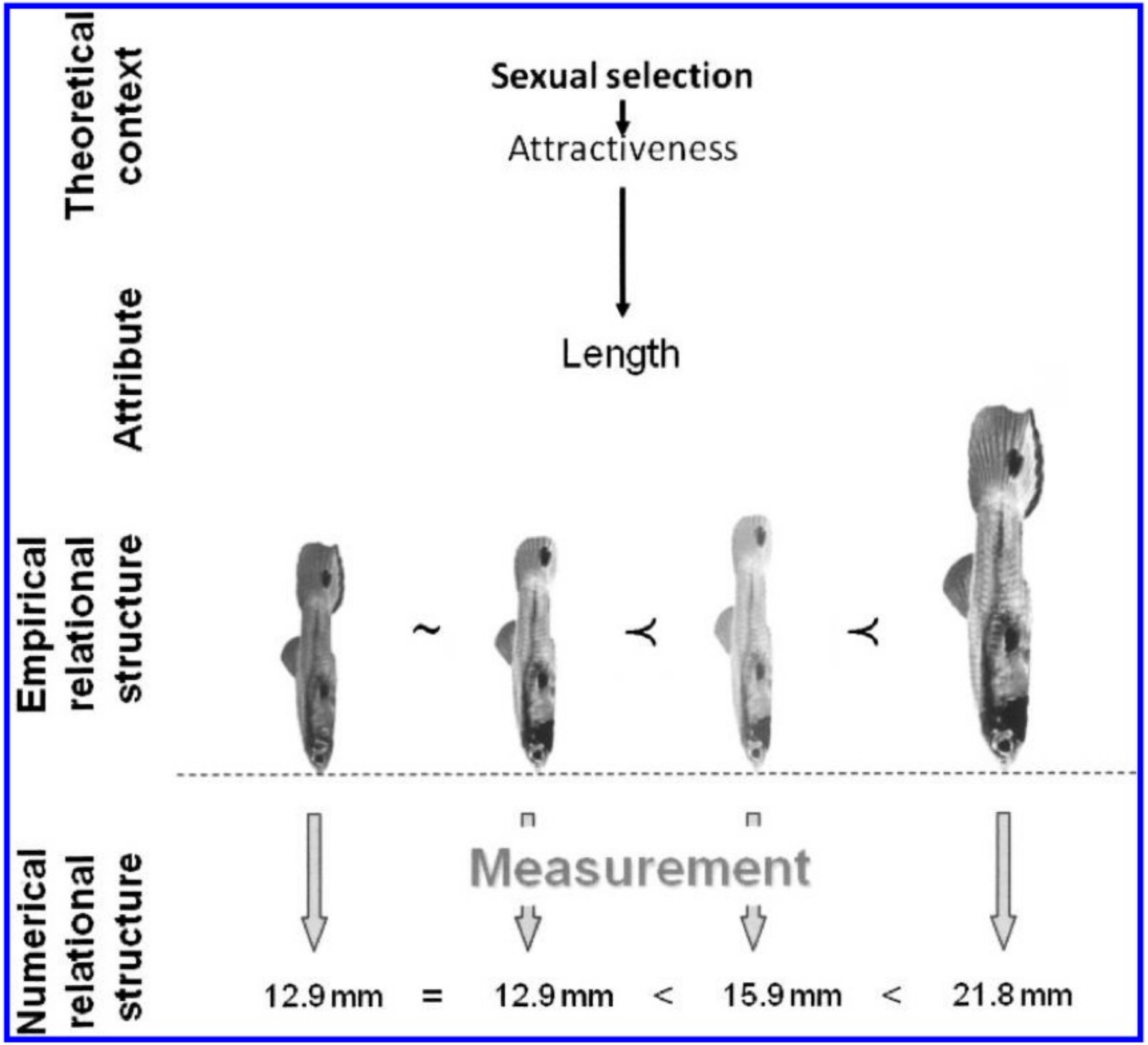
Houle et al. 2011

What is the relationship between what we measure and reality?

- Measurement = the assignment of numbers to attributes of the real world.
- The key principle of measurement theory is that theoretical context, the rationale for collecting measurements, is essential to defining ***appropriate measurements*** and interpreting their values.
- Its goal is ensuring that inferences about measurements reflect the underlying reality we intend to represent.

If *E* represents the empirical entities (say a sample of fish), then a scale can be viewed as a function from *E* to the real numbers.

The fundamental requirement of representational measurement is that ***the process of assigning numbers to elements of E should preserve the empirical relations on E***.

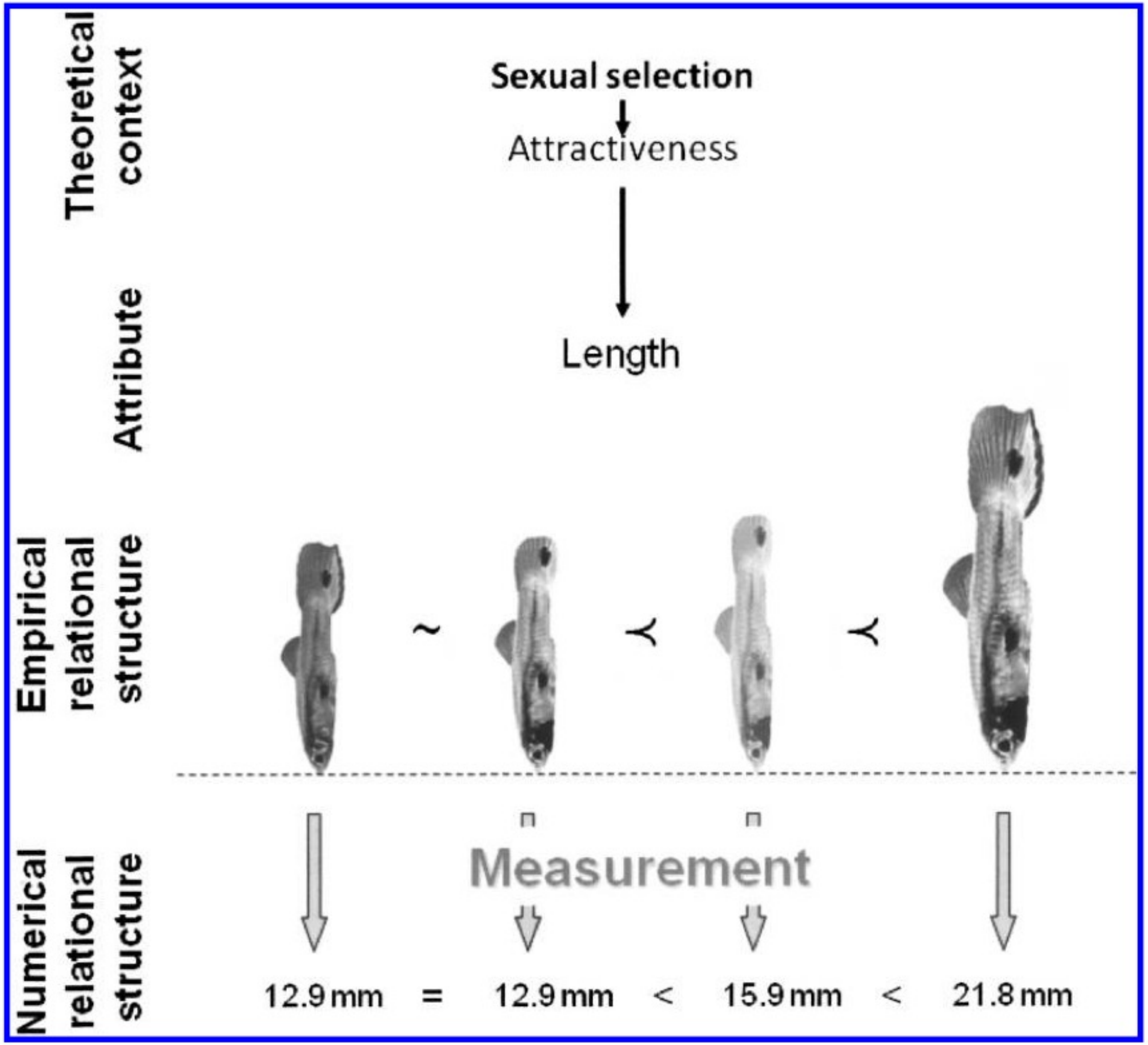


Houle et al. 2011

What relations should we consider? *Order*

- Can order of attributes be established? Elephant is bigger than a cat which is bigger than a mouse.

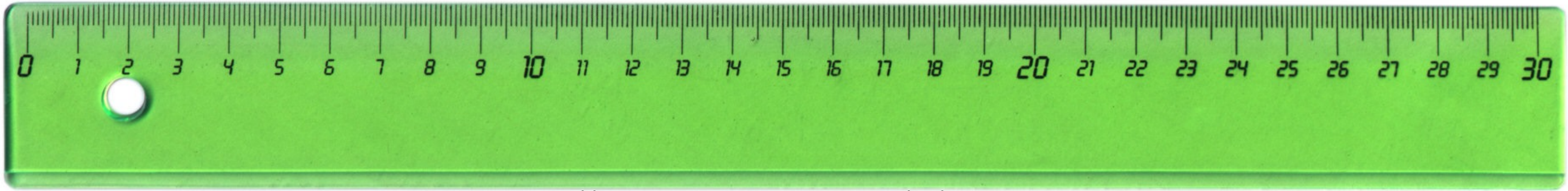
- Wild-type is more functional than a hypomorph which is more functional than a null morph in terms of alleles.



Houle et al. 2011

What relations should we consider? Distance

- Equally spaced, ordered units.



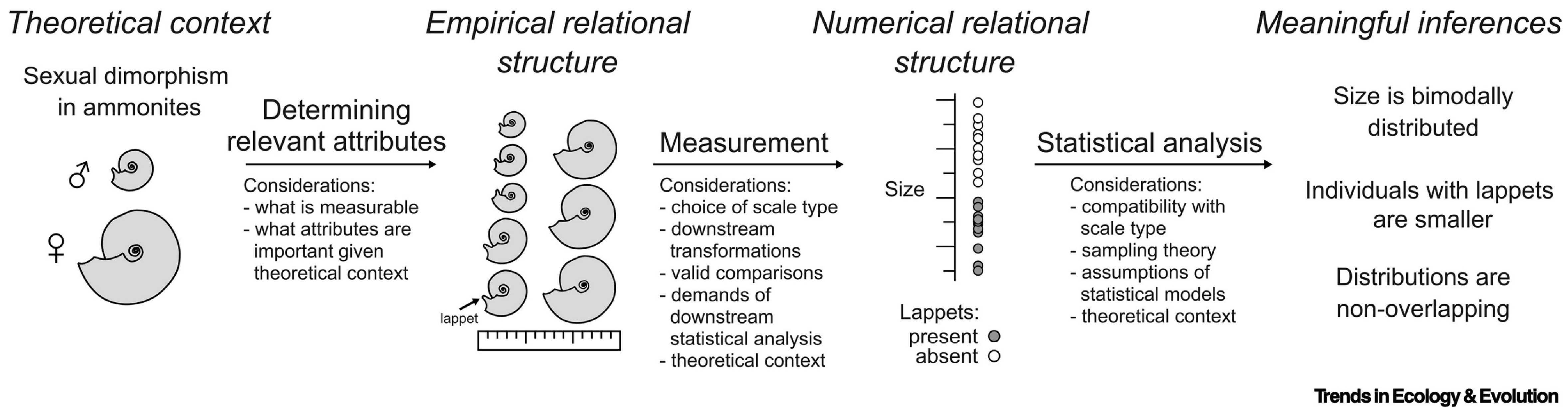
The distance between 1-2 on the ruler is the same as the distance between 20-21

What operations and relations should we consider? *Concatenation*

- Are the combined lengths of fish A and B (when lined up head to fin) longer than fish C?
- When concatenation operations are possible, then we can construct a ***standard sequence*** for measurement (such as a ruler in this case).

Relation structures depending on paired comparisons:

- Can you say that the attractiveness of Timothée Chalamet + Tom Holland is greater than the attractiveness of Michael B. Jordan?
- No natural concatenation is possible.
- Under controlled conditions ordering ***may*** be possible for this.



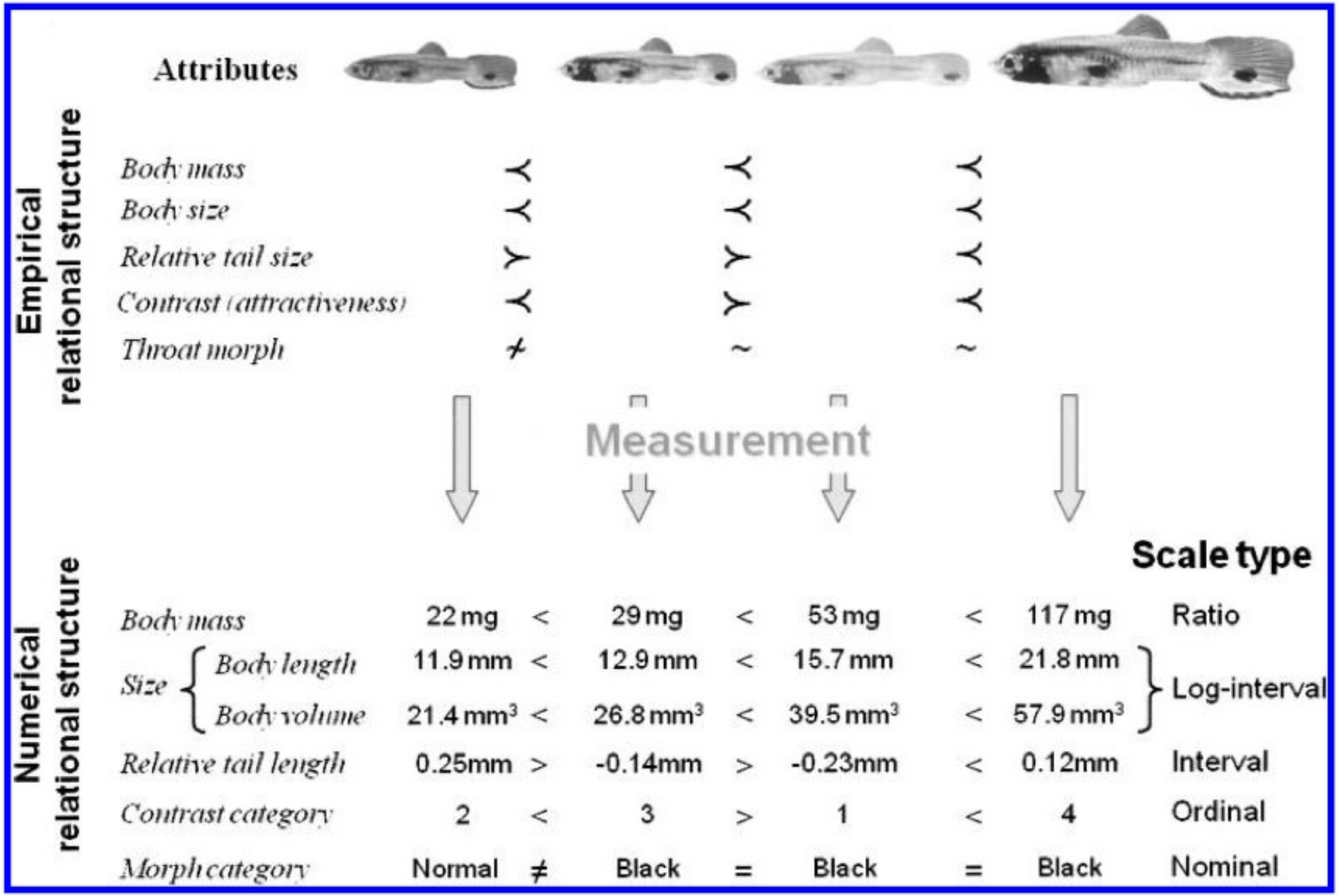
Voje et al. Trends in Ecology & Evolution 2023 381165-1176
DOI: (10.1016/j.tree.2023.08.005)

What relations should we consider? A meaningful zero.

- Does zero mean the absence of whatever is being measured?

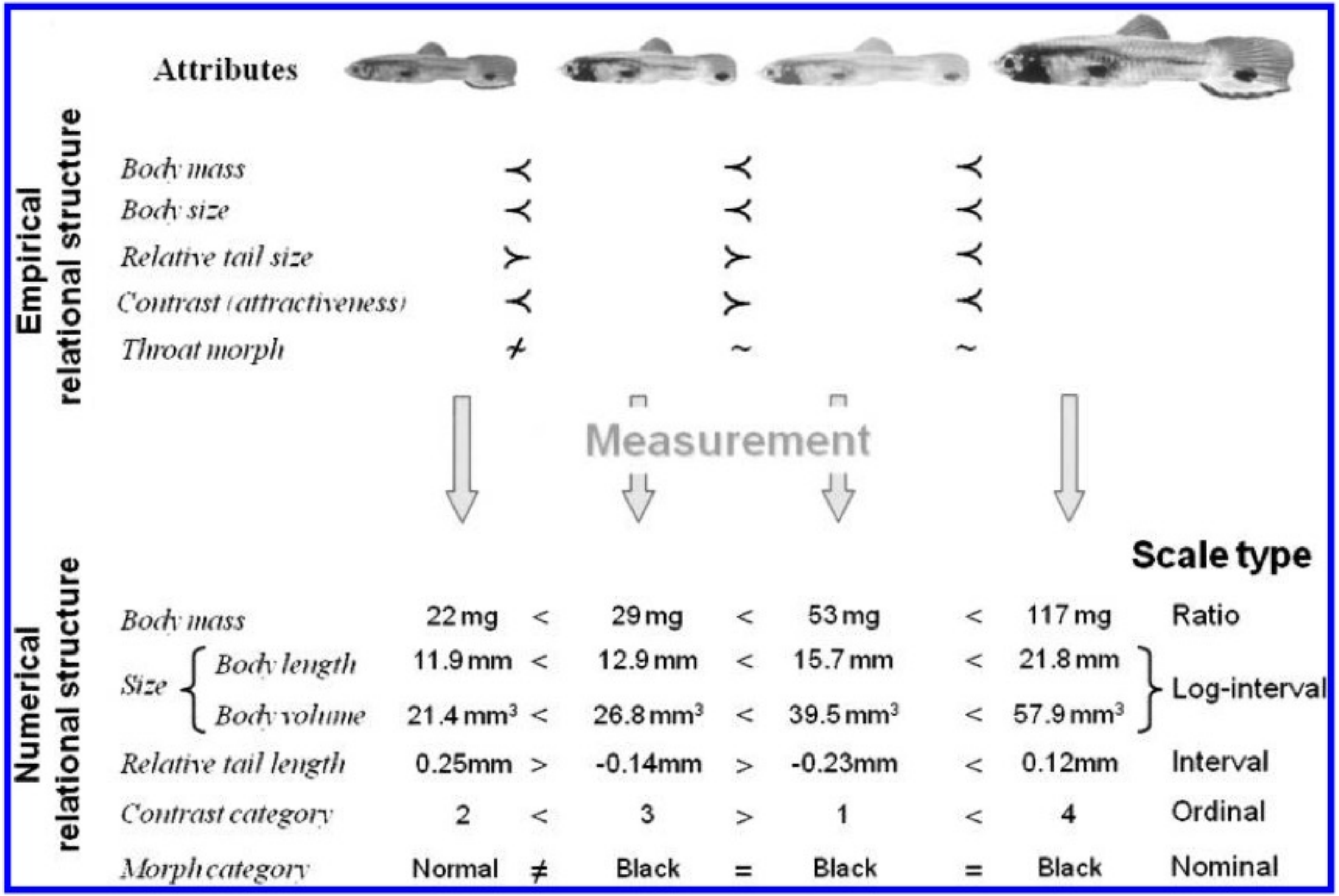
Nominal

- Things like species, genotype, flavour of ice cream (chocolate vs. strawberry), handedness (laterality).
- Two entities are assigned the same “value” or symbol if they have the same value of the attribute (examples).
- One-one transformations (equality) and many-one transformations are permissible. However the many-one loses information.
- Even if you code species A as “1” and species B as “2”, you can’t meaningfully add these!
- Permissible statistics (counting, frequency, mode, χ^2) (Stevens 1946, 1951)


Attributes
Empirical relational structure
Body mass ≺ ≺ ≺
Body size ≺ ≺ ≺
Relative tail size ≻ ≻ ≺
Contrast (attractiveness) ≺ ≻ ≺
Throat morph ≁ ~ ~
Measurement
Scale type
Numerical relational structure
Body mass 22 mg < 29 mg < 53 mg < 117 mg Ratio
Size
Body length 11.9 mm < 12.9 mm < 15.7 mm < 21.8 mm
Body volume 21.4 mm³ < 26.8 mm³ < 39.5 mm³ < 57.9 mm³
Log-interval
Relative tail length 0.25mm > -0.14mm > -0.23mm < 0.12mm Interval
Contrast category 2 < 3 > 1 < 4 Ordinal
Morph category Normal ≠ Black = Black = Black Nominal

Ordinal

- ***Has order and rank*** (wild type, weak perturbation, strong perturbation), but there is ***no meaning in the distance between ranks***.
  - i.e. we can not assume the same distance between wild type and a weak mutation, and between a weak mutation and a strong mutation.
  - Think about school grades, or winners of a competition coming in 1st, 2nd or 3rd
  - You can rank and order them; but cannot add or subtract them.
- Entities are assigned numbers such that the ranking/ordering of the numbers reflects the ordered relation for values of the attribute.
- If you have ”entitities” x & y with attribute values a(x) and a(y) with a(x) > a(y), then these are assigned measurement values m(x), m(y) such that m(x) > m(y).
- However concatenation (addition) may not be meaningful.
- Monotone increasing transformations are permitted (as long as order is maintained, any transformation is ok).
- Median and percentile are meaningful.
- ***mean, standard devaition are not meaningful***.



Attributes
Empirical relational structure
Body mass ≺ ≺ ≺
Body size ≺ ≺ ≺
Relative tail size ≻ ≻ ≺
Contrast (attractiveness) ≺ ≻ ≺
Throat morph ≁ ∼ ∼
Measurement
Numerical relational structure
Scale type
Body mass 22 mg < 29 mg < 53 mg < 117 mg Ratio
Size
Body length 11.9 mm < 12.9 mm < 15.7 mm < 21.8 mm
Body volume 21.4 mm3 < 26.8 mm3 < 39.5 mm3 < 57.9 mm3
Log-interval
Relative tail length 0.25mm > -0.14mm > -0.23mm < 0.12mm Interval
Contrast category 2 < 3 > 1 < 4 Ordinal
Morph category Normal ≠ Black = Black = Black Nominal

Interval

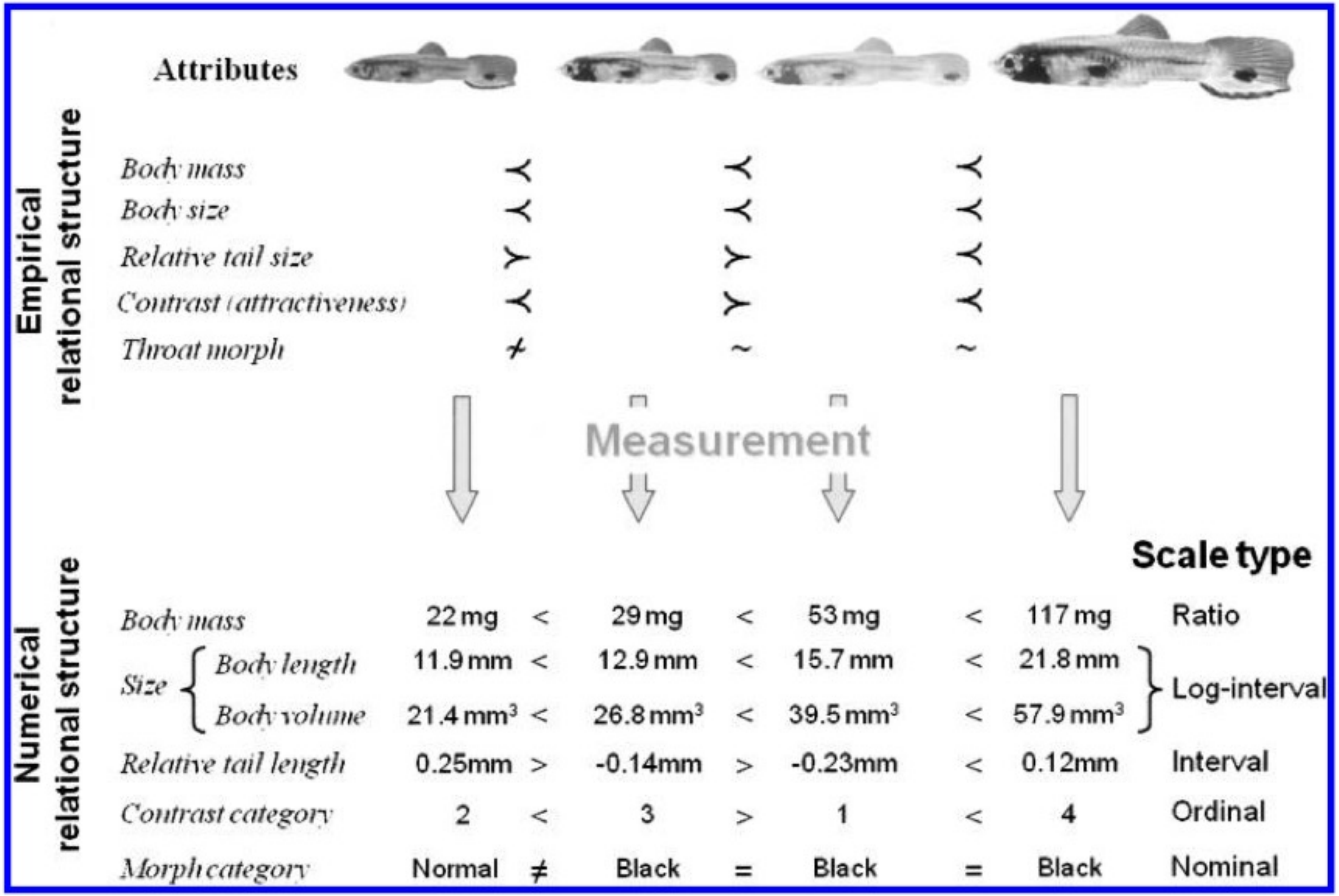
- Has order and distance. ***Each unit interval has the same distance (size) between it***.
- **Origin and unit of measurement are considered arbitrary**.
- Entities are assigned measurement values such that differences between the numbers reflect differences of the attribute. ***That is you could fit them along a line***.

- i.e. interval measures have no true zero (origin).
  - Using date as a measure of age as an example.
  - The difference in age between people born in 2006 and 1986 is the same (distance/size) as the difference between a person born in 1986 and 1966.
  - Longitude is another example.
- While differences (so adding and subtracting) have meaning, ***ratios do not*** (without a true zero. think about the example above).

- If m(x) – m(y) > m(u) – m(v) then a(x) – a(y) > a(u) – a(v)

- Permissible stats: mean, standard deviation, Pearson correlation, linear models (regression, ANOVA etc).
- Affine (including linear) transformations are possible

  t(m) = c*m + d   (with c, d constants)



Attributes
Empirical relational structure
Body mass
Body size
Relative tail size
Contrast (attractiveness)
Throat morph
Measurement
Numerical relational structure
Scale type
Body mass
22 mg < 29 mg < 53 mg < 117 mg
Ratio
Size
Body length
11.9 mm < 12.9 mm < 15.7 mm < 21.8 mm
Body volume
21.4 mm3 < 26.8 mm3 < 39.5 mm3 < 57.9 mm3
Log-interval
Relative tail length
0.25mm > -0.14mm > -0.23mm < 0.12mm
Interval
Contrast category
2 < 3 > 1 < 4
Ordinal
Morph category
Normal ≠ Black = Black = Black
Nominal

Ratio

- Ratio scaled variables have a natural zero, as well as meaningful distances.
- The measure represent absolute quantities of what is being measured. A zero means a complete absence.
- Ratio measures are the most informative ones.
- Entities are assigned numbers such that differences and ratios between the numbers reflect differences and ratios of the attribute.
- Linear transformations are permitted

  t(m) = c*m. Only the unit of measurement is arbitrary (not the origin).

All statistics are permitted for interval including ratios, arithmetic, geometic and harmonic mean, coefficient of variation and logarithms.

What are permissible transformations?

- Permissible/admissible transformations are transformations of a scale of measurement that ***preserve the relevant relationships*** of the measurement process (Sarle 1997).

- They transform one meaningful measurement scale into another meaningful measurement scale.
  - Celsius to Fahrenheit or Kelvins

- Only these transformations maintain the underlying biological meaning of the attributes from our entities (usually an organism).

Box 1. Scale types

Table I. Common scale types in paleobiology

Meaningful comparisons[a]	Scale type	How to identify[b]?	Examples in paleobiology	Allowed transformations[c]	Example transformation	Common meaningful statistics[d]	Meaningless statistics
Equivalence	Nominal	Labels	Species names, presence/absence of character	Any one-to-one mapping	{Presence, absence} = {0, 1}	Count and frequency of cases, mode	Distances
Equivalence, order	Ordinal	Ordered labels	Preservation categories, size categories, extent of trait development	Substitutions that maintain order	{Poor, medium, excellent} = {0,1,2}	Median, percentiles	Distances, arithmetic and geometric means, variance
Equivalence, order, differences	Interval	Equally spaced ordered units	Relative size, dates	Any linear transformation	Fractionation of oxygen isotopes between calcium carbonate and water to temperature in Celsius [70].	Mean, standard deviation	Geometric mean
Equivalence, order, differences, ratios	Ratio	Equally spaced ordered units with a meaningful zero	Length, width	Multiplication by a constant	100 g = 0.1 kg = 0.22 lb	Log transformation, geometric and arithmetic means	

Voje et al. Trends in Ecology & Evolution 2023 381165-1176. DOI: (10.1016/j.tree.2023.08.005)

Measurement Theory’s dos and don’ts

Box 3. Measurement theory’s dos and don’ts

Identifying the scale type of a measurement is imperative and is step 1 in any application of measurement theory. The following four questions categorize measurements into the scale types of nominal, ordinal, interval, and ratio, from lower to higher on the hierarchy (see Box 1), and should be asked in ascending order:

1. Could the values be replaced with characters/symbols and not lose any information?
   a. Yes (go to 2), the data belong to a nominal or ordinal scale.
   b. No (go to 3).
2. Does the order of the characters carry any meaning?
   a. Yes, the data belong to an ordinal scale.
   b. No, the data belong to a nominal scale.
3. Do the measurements have units in which increments are placed at equal distance from one another? If yes, the data belong to either an interval or ratio scale (go to 4).
4. Does a value of 0 mean there is ‘nothing’ of the entity in question?
   a. Yes, the data belong to a ratio scale type.
   b. No, the data belong to an interval scale.

Voje et al. 2023

Measurement Theory's dos and don'ts – part II

To avoid violating basic measurement theoretical principles, remember these points (the points are on a nominal scale):

- Means and variances of measures on a nominal or ordinal scale type should never be calculated. Calculating means and variances assumes distances between numbers carry meaning, a property that is not part of measures on a nominal and ordinal scale.
- Medians and modes can be calculated on ordinal scale variables because these are measures of central tendency in an ordered set of measures and do not assume meaningful distances between values.
- Numbers and frequencies of nominal and ordinal measures in a sample can be meaningful (e.g., the number of females in a population was 35, which represents 60% of the population).
- Logarithmic transformation should be applied only to measures on a ratio scale. Logarithmic transformation assumes that 0 means 'nothing of the entity measured.' Statements such as 'twice as large' and '20% less' are only meaningful if there is a true 0.
- Nonlinear transformations of data are often warranted (given the theoretical context of a study) but have consequences. Any nonlinear transformation of measurements (e.g., arcsine, logarithmic transformation, square root) changes the original link between the measurements and the reality they (initially) represented. Fitting a model to data before and after a nonlinear transformation means investigating different hypotheses.
- Applying the same statistical technique on variables belonging to different scale types assumes all the variables have the same properties. This is the same as making unjustified assumptions about the data.

Voje et al. 2023

Checklist from Houle et al 2011

- Keep theoretical context in mind.
- Honour your family of hypotheses.
- Make Meaningful definitions.
- Know what the numbers mean.
- Remember where the numbers come from.
- Respect scale type.
- Know the limits of your model.
- ***Never substitute a test for an estimate***.
- ***Clothe estimates in the modest raiment of uncertainty***.
- Never separate a number from its unit.

References

- Voje et al. 2023. Measurement Theory and Paleobiology.
- Houle et al. 2011. Measurement and Meaning in Biology.
- Wolman, A. 2006. Measurement and meaningfulness in conservation science.

- Hand, D.J. 1996. Statistics and the theory of measurement.
- Stevens, S.S. 1946. On the theory of Scales of Measurement.