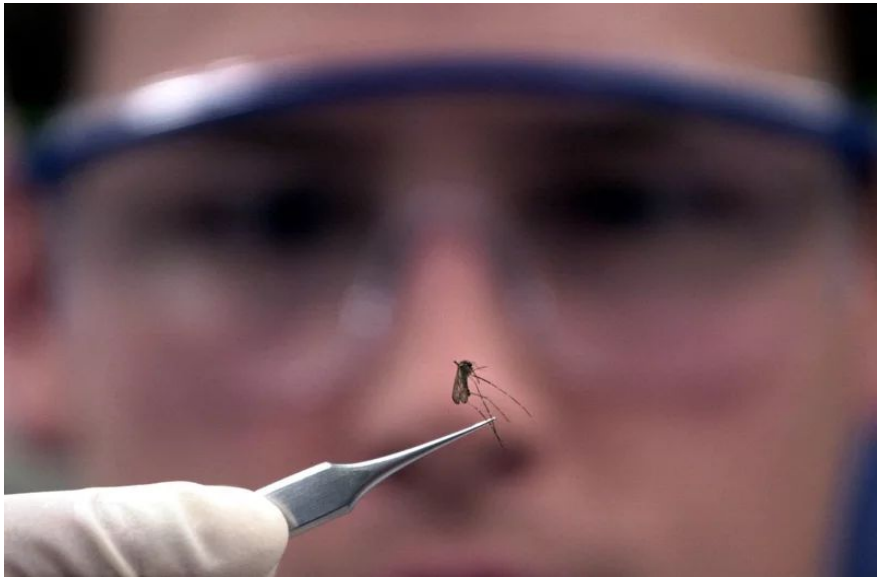
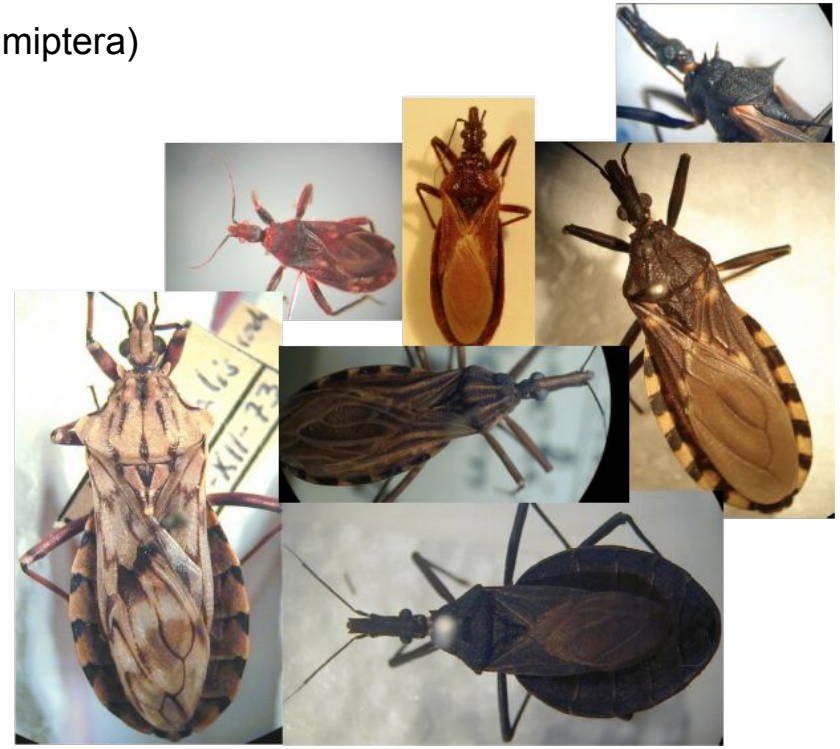
The recognition of forms
and
the taxonomists



http://time.com/5144257/fewer-scientists-studying-insects-entomology/ Fewer Scientists Are Studying Insects. Here's Why That's So Dangerous

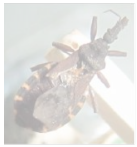
"**If you are studying a vector, or any other kind of organism, first try to recognize it**"
(A. Fain, Institute of Tropical Medicine, Belgium. 1980).

Triatominae (Heteroptera, Hemiptera)
Vectors of Chagas disease



“If you are studying a vector, or any other kind of organism, first try to recognize it”
(A. Fain, Institute of Tropical Medicine, Belgium).

Mexico
Guatemale
Honduras
Nicaragua
Panama
Colombia
Guianas
French Guyana
Venezuela
Ecuador
Perou
Brazil →
Uruguay
Argentine
Chile



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz
Centro de Pesquisas Aggeu Magalhães

“An alarming decline in the representation of taxonomy in university curricula and faculty was evident by the end of the **1980s** “… (Wheeler, 2014. Are reports of the death of taxonomy an exaggeration? New Phytologist 201 370 371)

Taxonomists: an endangered species...

The "**old guards**" of taxonomy have either retired or are retiring, many others have turned their backs on taxonomic research because of obvious and justifiable reasons, and unfortunately **new talents are nowhere in sight** mainly due to **lack of employability of taxonomists** (Jairajpuri, **1996**)

Melissa Mert. **2002** Taxonomy in Danger of Extinction. News from Science. http://www.sciencemag.org/news/2002/05/taxonomy-danger-extinction

"...Since 1992, f**unding for systematic biology** at major research institutions in the United Kingdom **has dropped** between 15% and 25% …"

Are We Losing the Science of Taxonomy?: As need grows, numbers and training are failing to keep up.

Lisa W. Drew

BioScience, Volume 61, Issue 12, 1 December **2011**, Pages 942–946,

"...During one recent visit to a **museum** that Mares will not name—"but it's one everyone on this planet has heard of," he says—he found that every **one of roughly 50 specimens**, representing seven species and three genera, was **mislabeled**."

The decline of Medical entomology...

About American and Canadian medical school curricula, a few years ago: **only 11 of 120 institutions include course content about arthropods**.

(**http://time.com/5144257/fewer-scientists-studying-insects-entomology/ Fewer Scientists Are Studying Insects. Here's Why That's So Dangerous**)

Sibling species

(Ernst Mayr in 1942,)

Two species that cannot be distinguished on morphological ground. (without necessarily any reference to the phylogenetic context).

Sibling species

(Ernst Mayr in 1942,)

Two species that cannot be distinguished on morphological ground.

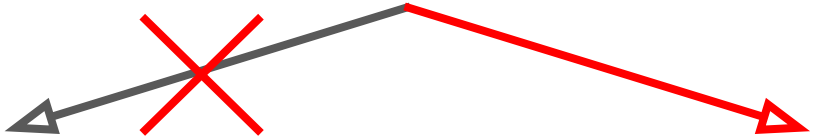
Anopheles gambiae,
~1950

- reproductive isolation between some laboratory strains
- cytogenetics could distinguish them
- Isoenzyme electrophoresis could etc.
- RFLP, RAPD, could etc.
- DNA probes could etc.

Sibling species, an ambiguous concept

(Ernst Mayr in 1942,)

“...cannot be distinguished…”



“ are identical (isomorphic) “

The characters currently used do not show difference.

The concept of sibling species is a subjective, or an ambiguous one

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

The concept of sibling species is a subjective, or an ambiguous one

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

The “**old guards**” of taxonomy have either retired or are retiring, … (Jairajpuri, **1996**)

The concept of sibling species, an ambiguous concept

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

- were the characters quantitatively defined?

The concept of sibling species, an ambiguous concept

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

- were the characters quantitatively defined?

SNEATH PH, SOKAL R. (1973). Numerical Taxonomy: The principles and practice of numerical classification. Ed. W. H. Freeman & Co., San Francisco 535pp

The concept of sibling species, an ambiguous concept

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

- were the characters quantitatively defined?

SNEATH PH, SOKAL R. (1973). Numerical Taxonomy: The principles and practice of numerical classification. Ed. W. H. Freeman & Co., San Francisco 535pp

To be possible, numerical taxonomy requires each morphological trait to have the same weight.

The concept of sibling species, an ambiguous concept

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

- were the characters quantitatively defined?
  (same weight for each one)
- were the quantitative characters redundant ones ?

The concept of sibling species, an ambiguous concept

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

- were the characters quantitatively defined?
  (same weight for each one)

- were the quantitative characters redundant ones ?

COOLEY AL, LOHNES P.(1971). Multivariate data analysis. Wiley, New York. 346pp.
CHATFIELD C, COLUNGS AJ. (1980). Introduction to multivariate analysis. Chapman & Hall, New York. 246pp.

The concept of sibling species, an ambiguous concept

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

- were the characters quantitatively defined?
  (same weight for each one)

- were the quantitative characters redundant ones ?

LANE RP, READY PD. 1985. Multivariate discrimination between Lutzomya wellcomei, a vector of mucocutaneous leishmaniasis and Lu. complexus (Diptera: Phlebotominae) Ann. Trop. Med. & Parasitol. 79:46947

MS. McNAMEE C. DYTHAM 1993 Morphometric discrimination of the sibling species Drosophila melanogaster (Meigen) and D. simulans (Sturtevant) (Diptera: Drosophilidae). Systematic Entomology Volume 18, Issue 3

Etc.

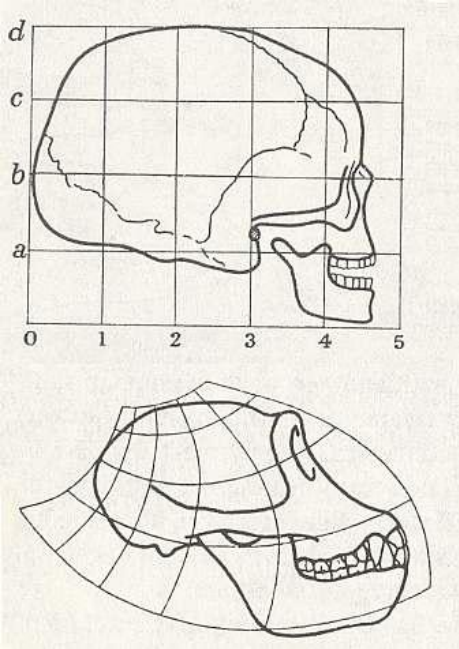
The concept of sibling species, an ambiguous concept

Species recognition in insects or in other organisms may rely on characters other than the ones traditionally used by taxonomists

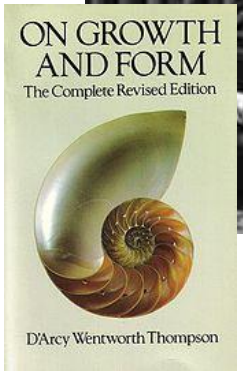
- were the characters quantitatively defined?
  (same weight for each one)
- were the quantitative characters redundant ones ?
  (multivariate analyses)
- were the quantitative characters completely defined?
  (shape ?)

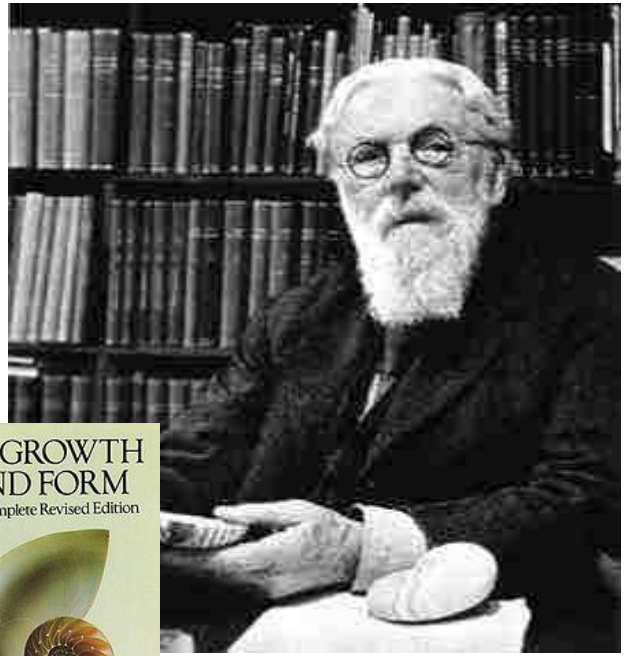
were the quantitative characters completely defined?

Size and **shape !**

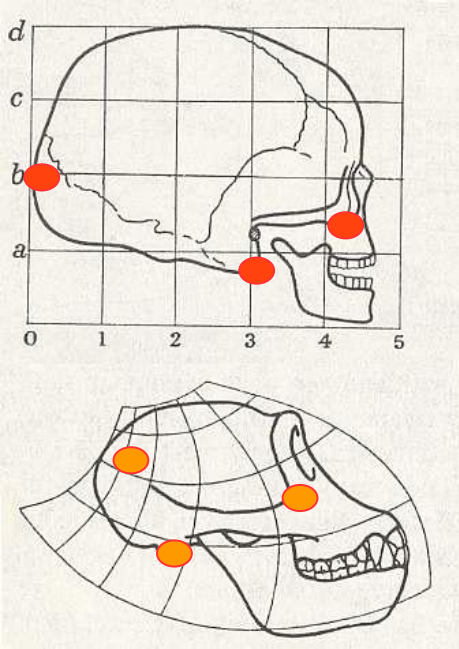


« landmark-based »

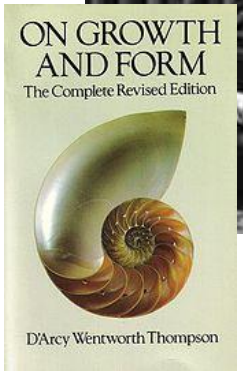


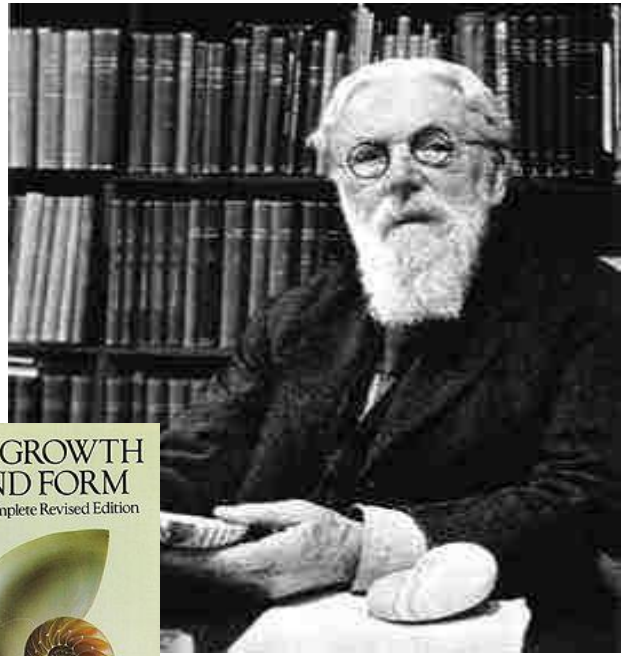


Born
2 May 1860
Edinburgh
Died 1948 (aged 88)
St Andrews
Occupation
Mathematical biologist

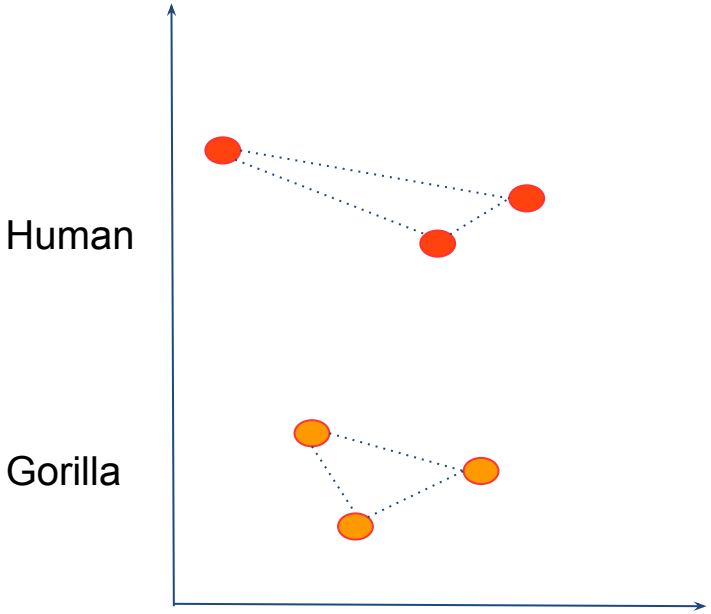


« landmark-based »

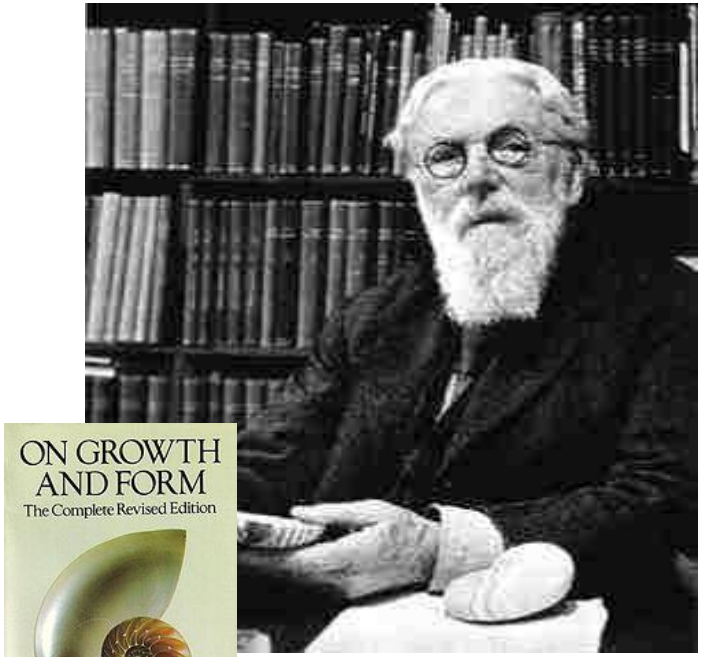




Born
2 May 1860
Edinburgh
Died 1948 (aged 88)
St Andrews
Occupation
Mathematical biologist



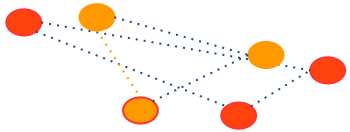
« **landmark-based** »



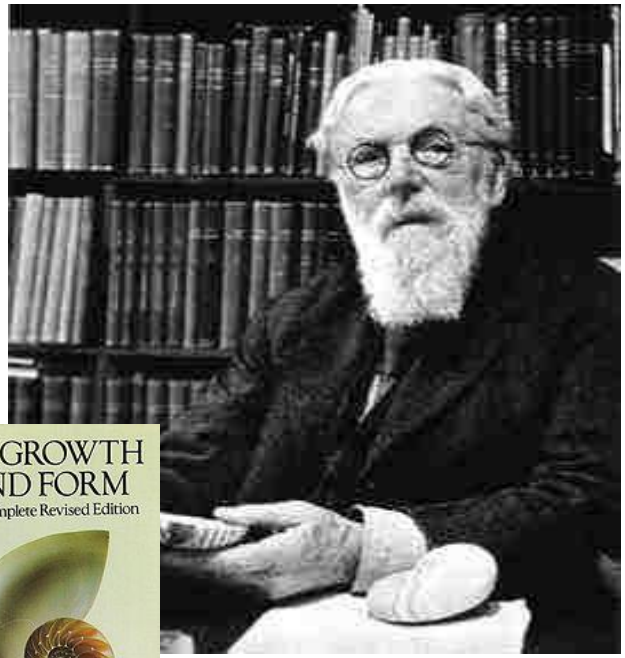
Born
2 May 1860
Edinburgh
Died 1948 (aged 88)
St Andrews
Occupation
Mathematical biologist

Human

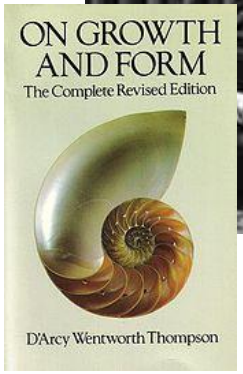
Gorilla



« **landmark-based** »



Born
2 May 1860
Edinburgh
Died 1948 (aged 88)
St Andrews
Occupation
Mathematical biologist



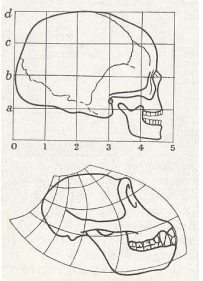
Kendall D. G. 1984. Shape manifolds, Procrustean metrics and complex projective spaces. Bulletin of the London Mathematical Society 16: 81-121

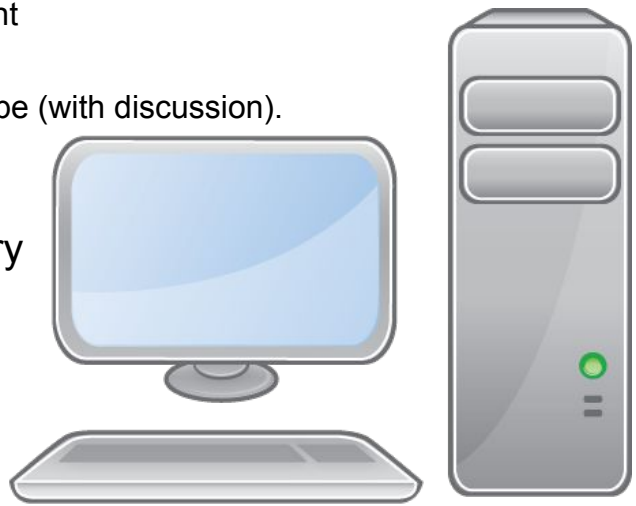
Bookstein F. L. 1984. A statistical method for biological shape comparisons. Journal of Theoretical Biology 107: 475–520.

Bookstein F. L. 1986. Size and shape spaces for landmark data in two dimensions (with discussion). Statistical Science 1: 181–242.

Rohlf J. F. 1986. Relationships among eigenshape analysis, Fourier analysis, and analysis of coordinates. Mathematical Geology 18: 845–654.
Rohlf J. F. 1999. Shape statistics: Procrustes superimposition and tangent spaces. Journal of Classification 16: 197–223.

Goodall C. R. 1991. Procrustes methods in the statistical analysis of shape (with discussion). Journal of the Royal Statistical Society, Series B 53: 285-339

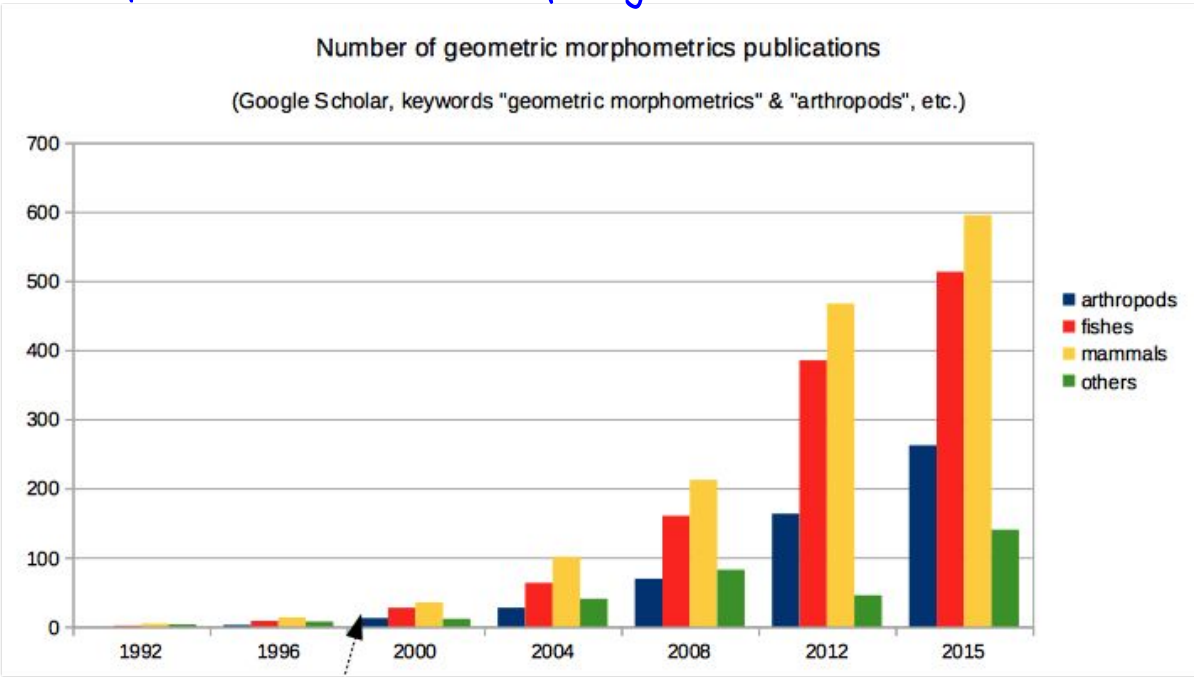




Other important names in developing techniques and theory since 1990

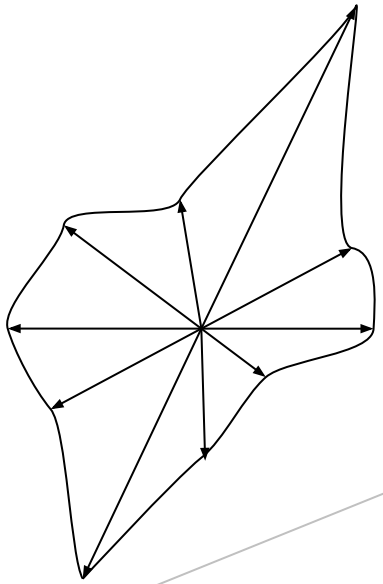
Ian **Dryden**,
Kanti **Mardia**,
Dennis **Slice,**
Dean **Adams**,
… Etc.

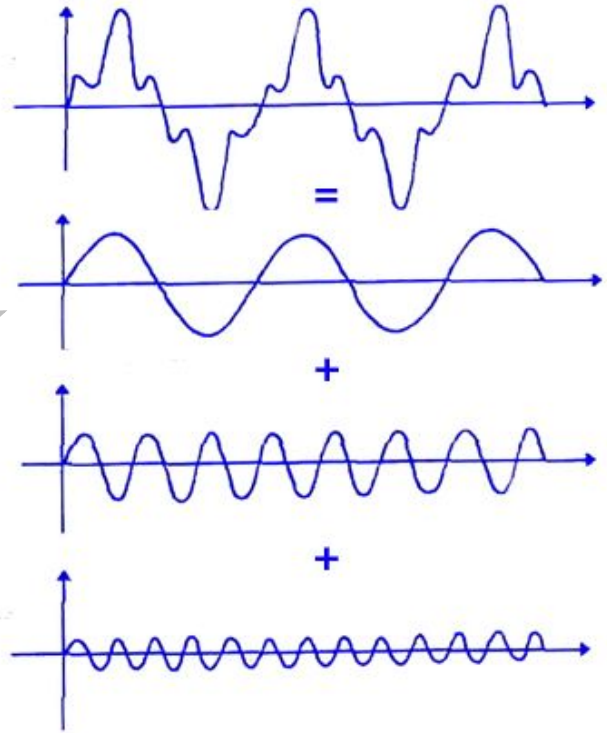
were the quantitative characters completely defined? Size and **shape !**



A revolution morphometrics. Rohlf, Marcus
Trends Ecol Evol. **1993** 8(4):129-32. doi: 10.1016/0169-5347(93)90024-J..

Geometric Morphometrics: Ten Years of Progress Following the 'Revolution'
Adams, Rohlf, Slice. Ital. J. Zool., **2004** 71:5-16.





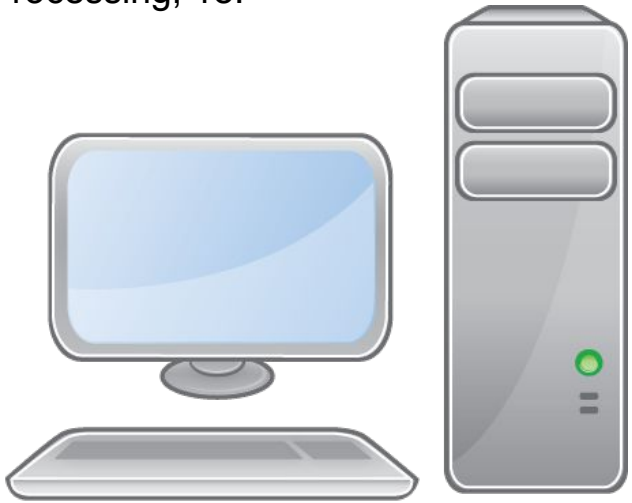
Harmonics

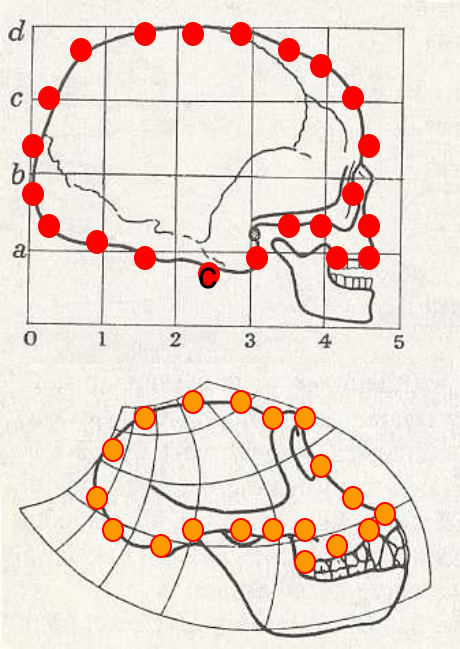
Jean Baptiste Joseph Fourier (1768 – 1830) was a French physicist and mathematician who is known for investigating the Fourier series and its application to problems of heat flow.

Kuhl FP and **Giardana CR**, **1982**. Elliptic Fourier features of closed contour. Computer Vision, Graphics and Image Processing, 18: 236-258.

Other names

Lestrel, 1987
Rohlf, 1990
Etc.

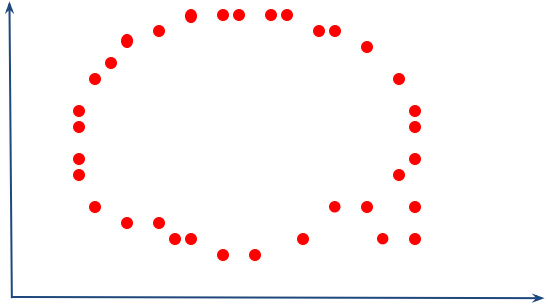


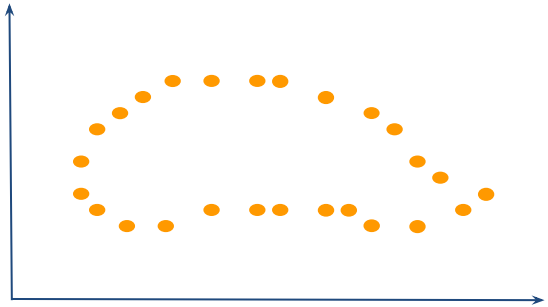


Harmonics

Jean Baptiste Joseph Fourier (1768 – 1830) was a French physicist and mathematician who is known for investigating the Fourier series and its application to problems of heat flow.

Kuhl FP and **Giardana CR**, **1982**. Elliptic Fourier features of closed contour. Computer Vision, Graphics and Image Processing, 18: 236-258.

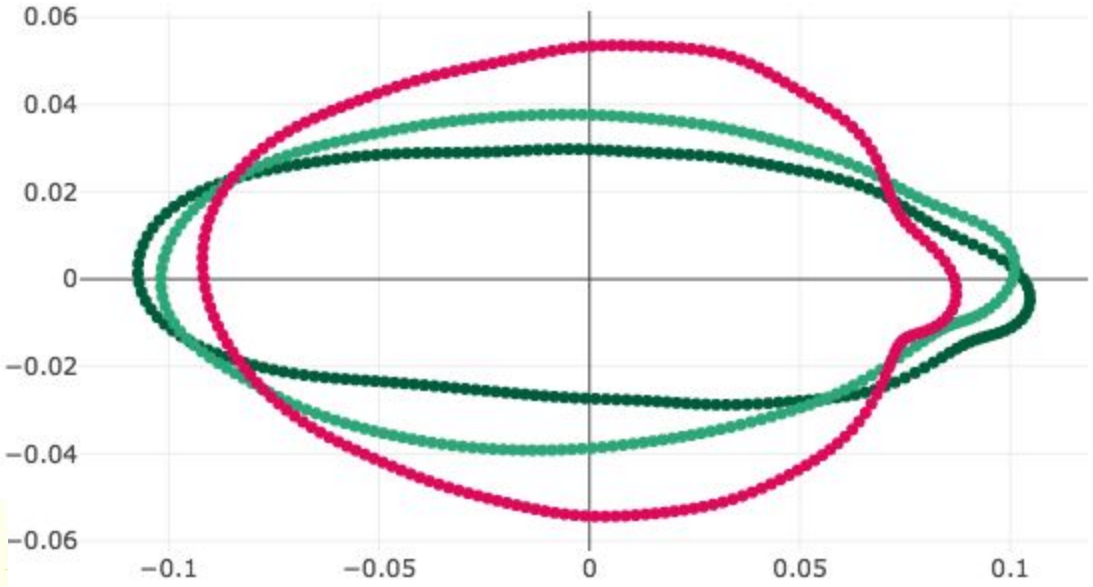




Harmonics

Jean Baptiste Joseph Fourier (1768 – 1830) was a French physicist and mathematician who is known for investigating the Fourier series and its application to problems of heat flow.

Kuhl FP and **Giardana CR**, **1982**. Elliptic Fourier features of closed contour. Computer Vision, Graphics and Image Processing, 18: 236-258.



17 F. gigantica

9 F. hepatica

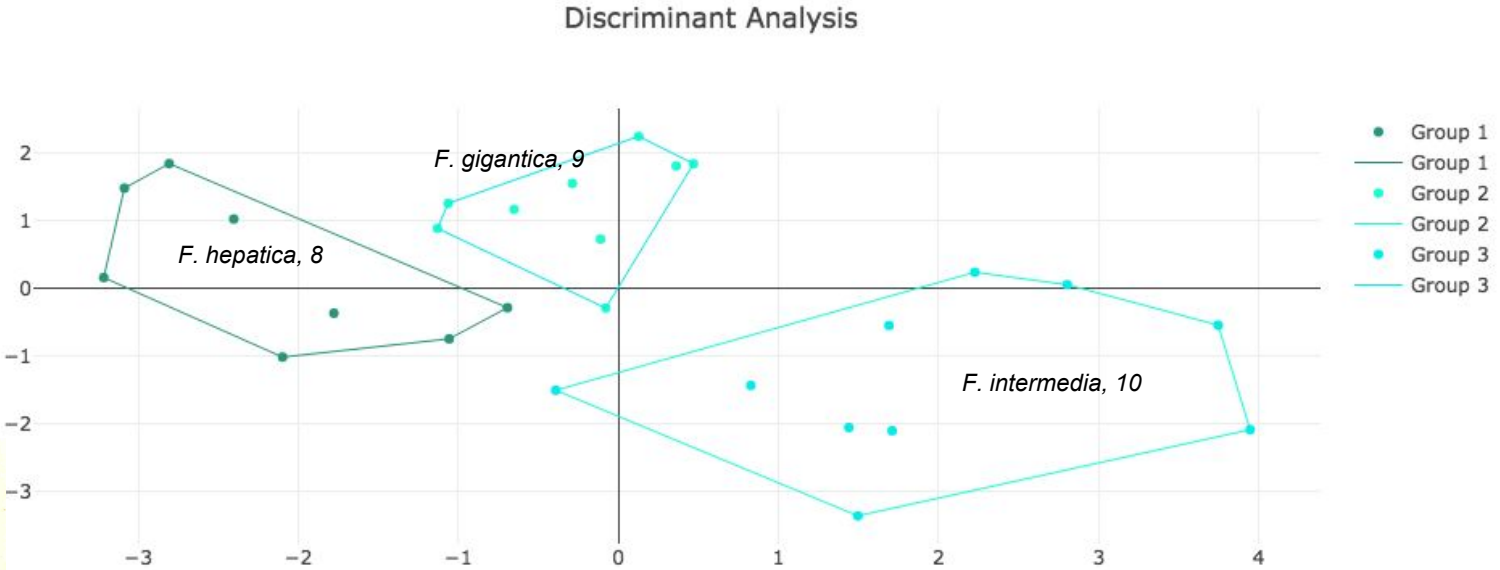
10 F. intermedia

Jean Baptiste Joseph Fourier (1768 – 1830) was a French physicist and mathematician who is known for investigating the Fourier series and its application to problems of heat flow.

Fasciola sp.

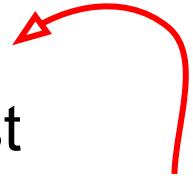
Graphical outputs

PCA GRAPHIC | DA GRAPHIC | HC GRAPHIC



Analysis Report of 7/15/2018, 6:15:57 PM

The “new” taxonomist,
as opposed to the “old” taxonomist,
does not need to be a hyperspecialist
of the organism under study.

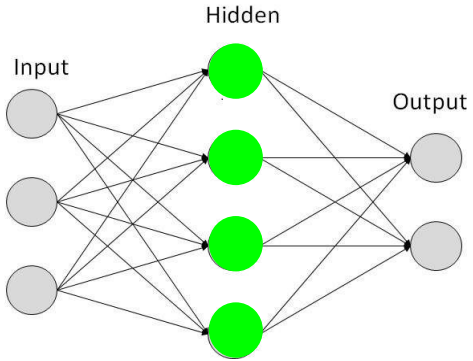


The “**old guards**” of taxonomy have either retired or are retiring, ... **lack of employability of taxonomists** (Jairajpuri, **1996**)

MACHINE LEARNING

-> SUPERVISED LEARNING

-> Support Vector Machine
-> Neural Network
-> etc. etc.



MACHINE LEARNING

Known data

species A, species B, etc.

MACHINE LEARNING

Known data

species A, species B, etc.

Quantitative data

(Morphometric data)
- size
- log-shape ratios
- coordinates after geometric processing
- etc.

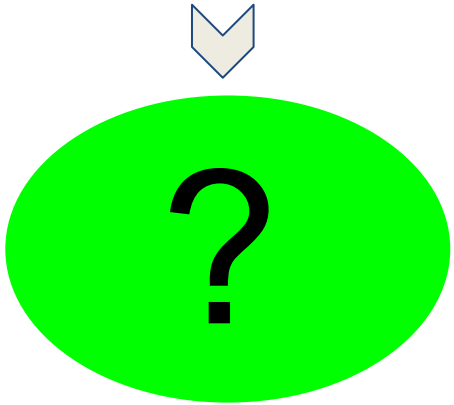
(Other kind of data)
- GPS coordinates
- Etc.

Quantitatively coded data
- present/absent
- frequencies

MACHINE LEARNING

Known data = Reference data

species A, species B, etc.



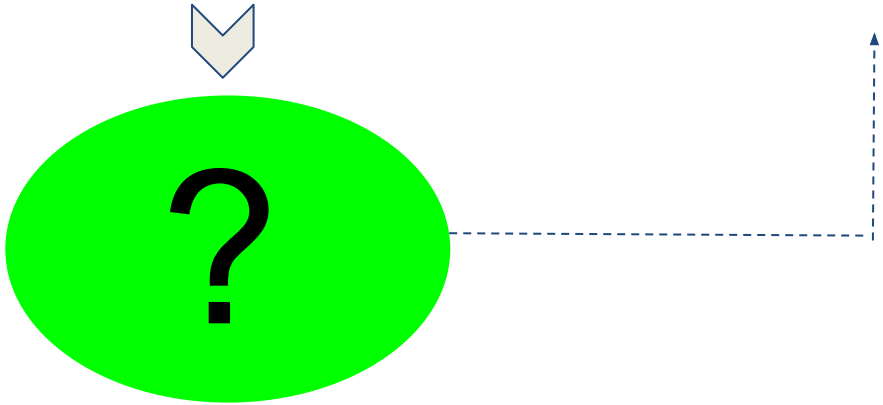
MACHINE LEARNING

Known data = Reference data

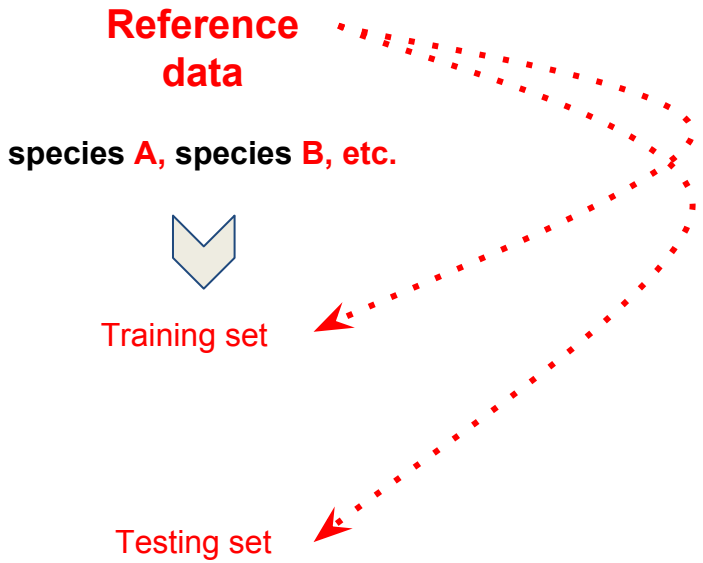
species A, species B, etc.

Unknown data

species ??



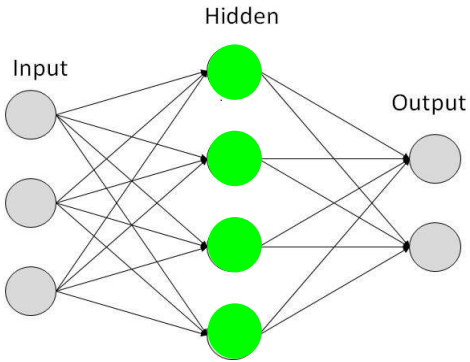
MACHINE LEARNING



MACHINE LEARNING / NEURAL NETWORK

Reference data

species A, species B, etc.



Neural network

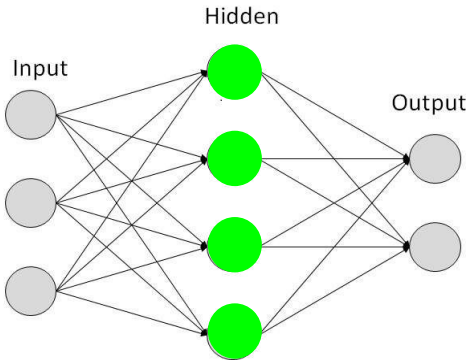
MACHINE LEARNING / NEURAL NETWORK

Reference data

species A, species B, etc.

Unknown data

species ??



"weights"

Neural network

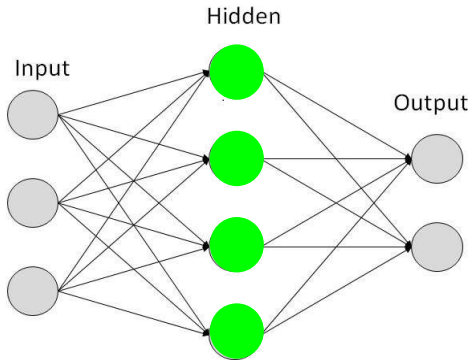
MACHINE LEARNING / NEURAL NETWORK

Quantitative data
Or
Quantitatively coded

“weights = 1”

Reference data

species A, species B, etc.



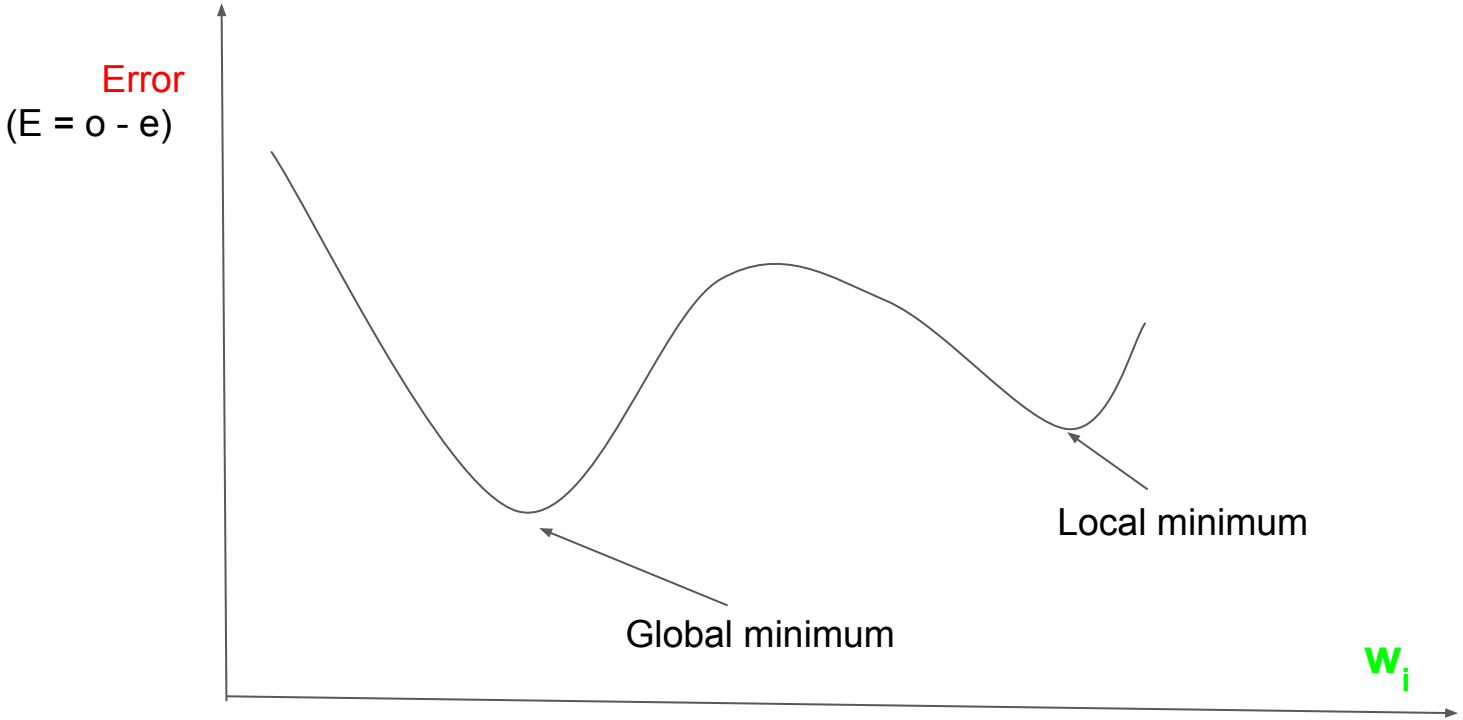
Neural network

Unknown data

species ??

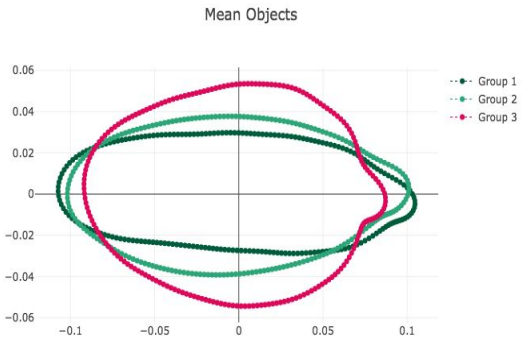
“weights”

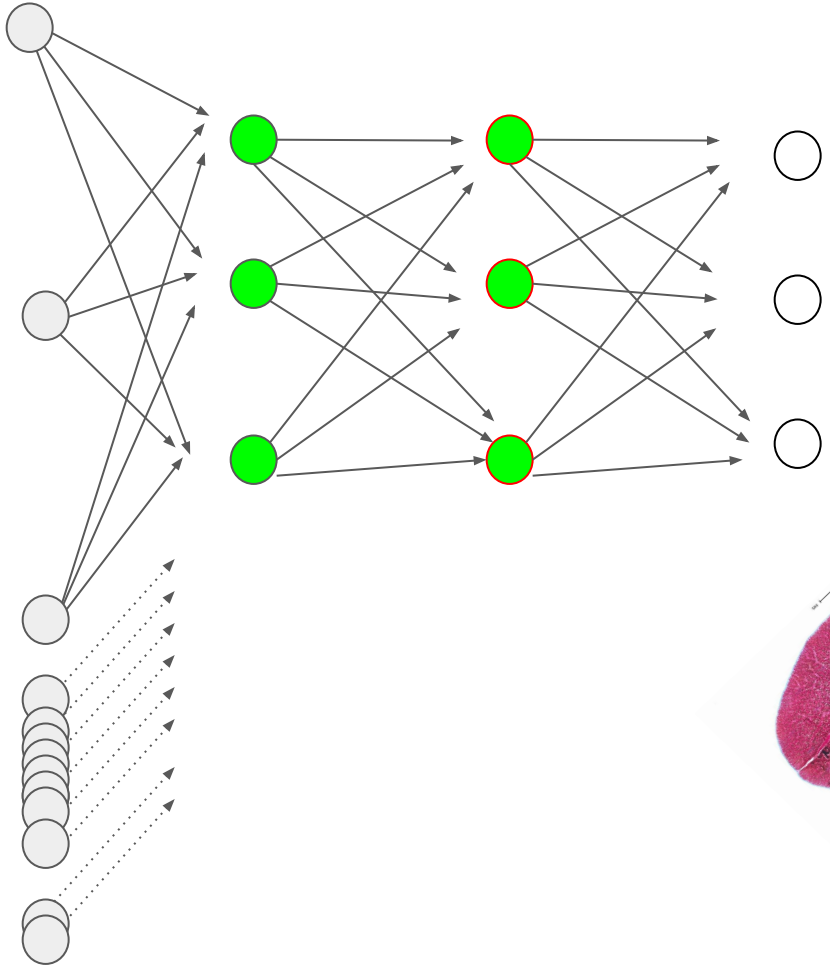
A graph of the training error as a function of the input parameters



The weight update for a given node has the following (simple) form:

$$\Delta w_i = -\alpha \frac{\partial E}{\partial w_i}$$





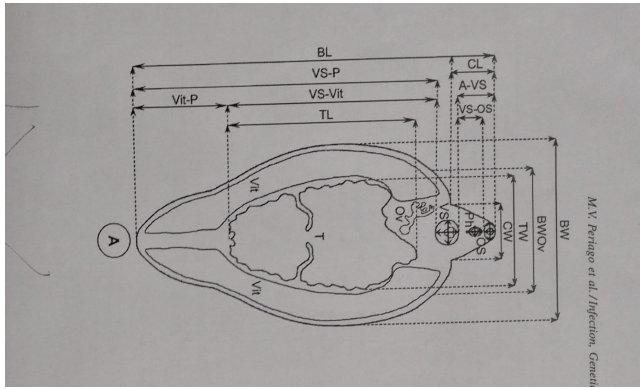
17 F. gigantica

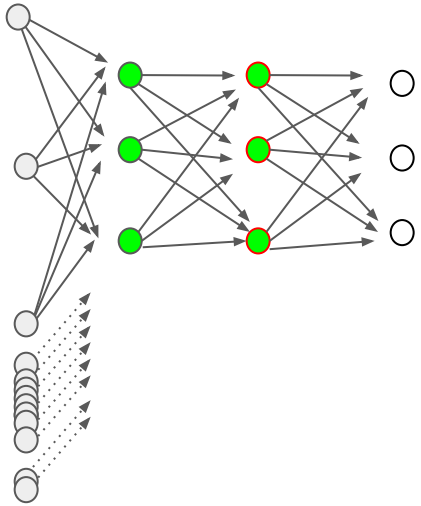
9 F. hepatica

10 F. intermedia

Scores of validated classification using as input for the multilayer perceptron log-transformed **linear measurements** (LN)

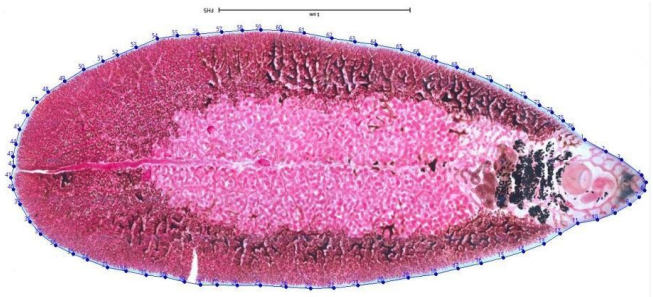
	LN
4 LN	64% (13/36)
6 LN	78% (8/36
8 LN	81% (7/36
10 LN	87% (5/36
12 LN	89% (4/36)

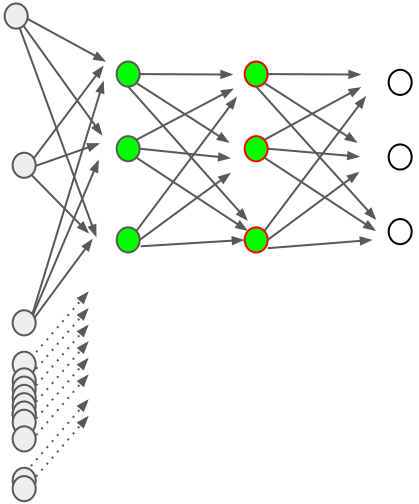




Scores of validated classification using as input for the multilayer perceptron OUTLINE DATA.

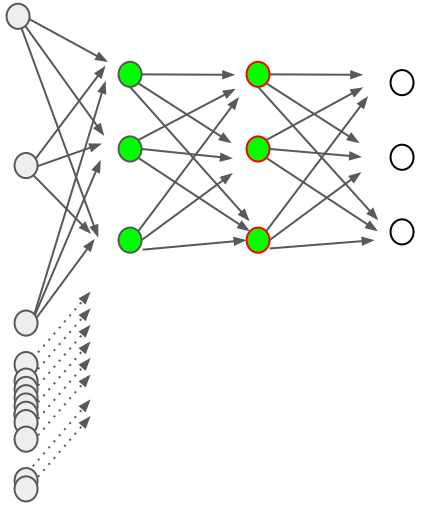
	LN	
4 LN	64%	
6 LN	78%	
8 LN	81%	
10 LN	87%	
12 LN	89% (4/36)	89% (4/36)

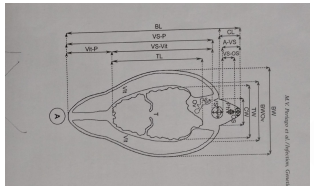




Scores of validated classification using as input for the multilayer perceptron
either log-transformed linear measurements (LN) alone,
or **a combination of LN with outline data**.

	LN	LN + outline data
4 LN	64%	95% = 2 errors (out of 36 identifications)
6 LN	78%	98% = 1 error (out of 36 identifications)
8 LN	81%	95%
10 LN	87%	98%
12 LN	89% (4 errors /36)	**100% = 0 error** (out of 36 identifications)





SUMMARY

Traditional taxonomy was based on qualitative morphological characters <- Human EYES

It was too subjective of an approach, and has been put into difficulty by a subjective concept, the “sibling species” concept.

Modern taxonomy makes use of quantitative morphological characters <- Computer EYES

- It removes the requirement to be an expert of the group of insects under study.
- It removes subjectivity of the taxonomist (89%, 95%, etc.).

CONCLUSION

The new taxonomist is expected **to be able to use computerized methods** for taxonomic data collection, analyses and classification

Computers perform billions of accurate calculations in less than a second…

Do we conclude that we do not need mathematicians anymore?