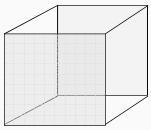
Experimental Research

D.G. Bouwhuis
HTI – Human Technology Interaction
Eindhoven University of Technology TU/e
d.g.bouwhuis@tue.nl

3 principles of science according to Guy J. Consolmagno

1. there is an real external objective world
2. within this world all events and relations follow a system of logical and consistent rules
3. the pursuit of science must be meaningful and worthwhile

objective			subjective
-1	-1	1	
1	-1	1	
-1	1	1	
1	1	1	
-1	-1	-1	
1	-1	-1	
-1	1	-1	
1	1	-1	



- What is your research question?
- can you state in a single sentence what you want to know, or what you are interested in

- What is your research question?
- can you state in a single sentence what you want to know, or what you are interested in
- what is the scientific added value of the research
  - confirmatory
  - exploratory
  - methodologically
  - theoretically
- who, or what might benefit from the research

Newtonian Physics — blocks, spheres, mass, force, friction. . .

3TU School Human Technology Interaction

The Sydney Opera House roofs consist of sphere segments, made of reinforced concrete. The size, thickness, strength and curvature were empirically determined on site.

Expansion and contraction of the shells during the day as a function of temperature had strong and unexpected effects

3TU School Human Technology Interaction

Newtonian physics is typically directed at a single object, or two objects with some mechanical relation.

Our solar system is at first approximation Newtonian, (elliptic orbits, mutual attraction proportional to squared distance..) but on a longer term chaotic.

1.Computational problems become unsolvable with increasing number objects
2.There are too many "unobservables"

3TU School Human Technology Interaction

Newtonian Physics — blocks, spheres, mass, force, friction. . .
molecular and atomic physics — quantum theory

3TU School Human Technology Interaction

Newtonian Physics — blocks, spheres, mass, force, friction. . .
molecular and atomic physics — quantum theory
particle physics, high energy physics — the Standard Model

3TU School Human Technology Interaction

large complex hybrid systems — engineering problems,
Newtonian Physics — blocks, spheres, mass, force, friction. . .
molecular and atomic physics — quantum theory
particle physics, high energy physics — the Standard Model

3TU School Human Technology Interaction

social systems — economy, health ↭ human behaviour
large complex hybrid systems — engineering problems,
Newtonian Physics — blocks, spheres, mass, force, friction. . .
molecular and atomic physics — quantum theory
particle physics, high energy physics — the Standard Model

←low scalability→

Quantitative Research
('exact' sciences, behavioural research)
- representative sampling
- independent of investigator
- measurement consists of quantitative and symbolic/categorical assessment
  - any researcher should find the same analysis results of the data
  - any researcher should be able to replicate the experimental study,
  - and find the same experimental results

Qualitative Research
(social/behavioural research)
- purposive, not random sampling
- key role of investigator in the research process
- measurement consists of coding
  - participation in the setting,
  - direct observation,
  - in depth interviews,
  - analysis of documents and materials

- N=1 research (or N=few)
  need not be problematic as long as there is a sufficient number of measurements.
  - it is actually difficult to have close to 1 or 2 measurement points
  - for low numbers of observations the Fisher Exact probability test is very well suited.

  (data matrix k x l; $k, l \geq 2$)

- N=1 research (or N=few)
  need not be problematic as long as there is a sufficient number of measurements.
  - it is actually difficult to have close to 1 or 2 measurement points
  - for low numbers of observations the Fisher Exact probability test is very well suited.

  (data matrix k x l; $k, l \geq 2$)
- N=0 research (psychoanalysis, creationism)

Data levels
1. Nominal
2. Ordinal
   2.1 Ordered metric
3. Interval
4. Ratio

Data levels	**Allowed transformations**
1. Nominal	renaming
2. Ordinal	strictly monotonic
2.1 Ordered metric	strictly monotonic
3. Interval	$ax + b$
4. Ratio	ax

Nominal are all identification responses:
- I saw the word "`city`"
- the colour is 'red'
- this is beautiful
- that is literature
- I can only see it as political
- a theme from Madama Butterfly!
- I agree moderately (*on a 5-point scale*)

Nominal

identification, recognition and categorization

- nominal data handling
  - adding up the number (frequency) of the responses
  - Thurstonian scaling
  - measuring the time it takes to produce the response, slower is less identifiable
  - counting the errors (if applicable) in responding (an effectiveness measure)

Ordinal

expressing an order relation between two (or more) stimuli

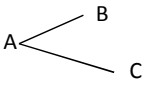
A p B : A has less of property X than B

A f B f C: A,B and C can be ordered this way

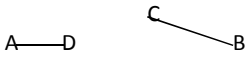
transitivity implies A f C

Ordered metric

A is closer to B than C (conjoint pairs)



A is closer to D than B is to C (disjoint pairs)



Ordered metric scale

when the number of stimuli increases the number of order relations allows to construct an 'almost' metric (interval) scale in one, or more dimensions

order relations = n(n-1)/2

n = 5	10	10
n = 10	45	120
n = 15	105	455
n = 20	190	1140

additional constraints rely in the triangle inequality (numbers in gray).

Interval scale

In actual practice always derived from scaling methods like Thurstone scaling , or normal deviates (standard scores)

or from scores on a test, like intelligence, personality, field dependence, etc.

Important: there is no natural origin, or zero.

Ratio scale data

- ratio scaling: "this stimulus is 2.5x as strong as the standard"
- direct estimation: "this is about 87% of the standard"
- response times can be considered to be ratio scale data, but mostly behave statistically like interval data

« ratio scale data are restricted to real positive numbers
« Logarithmic transformation of ratio scale data makes them indistinguishable from interval data.
« there are no special statistical tests for ratio scale data

3TU School Human Technology Interaction

Behaviour measurement types

quantitative measures

1. counting (phenomena, choices made)
2. speed, response times
3. accuracy (identification, recognition, solution)

subjective measures

4. preferential judgment
5. ratings (with respect to a concept or variable)
6. ratings (with respect to similarity or distance)

3TU School Human Technology Interaction

Behavioural measurement types

physical measures

6. behavioural measures (gestures, gaze, comments. .)

biophysical measures

6. physiological measures (heart rate, heart rate variability, galvanic skin response, pupil size. . .)
7. neurological measures (neural imaging, EEG, fMRI, CATscan, TMR. .)

3TU School Human Technology Interaction

One example: Ratings

totally disagree | disagree | neutral | agree | totally agree

an example of categorical scaling, in which categories are separated by boundaries b_k

3TU School Human Technology Interaction

b_0 b_1 b_2 b_3 b_4

The score x_i (impression or attitude) of subjects is a stochastic variable x the probability of which is:

$$P(x_i \geq b_j) = ND(x_i - b_j) \text{ or } LD(x_i - b_j)$$

The number of parameters is n $(x_1\text{-}x_n)$ + 4 $(b_1\text{-}b_4)$

A considerable reduction is obtained by assuming the category widths to be equal:

$$P(x_i \geq b_j) = ND(x_i - jb) \text{ or } LD(x_i - jb)$$

This is the basic form of Thurstonian scaling

3TU School Human Technology Interaction

This model produces an interval scale value for each variable x_i ; these usually derived by linear least squares estimation.

advantages:

- not all cells have to be filled
- the scaling procedure allows to include more variables as additive, or linear combinations, allowing to test a wide range of more complex and comprehensive models.
- standard errors of estimations are available.

3TU School Human Technology Interaction

Experimental example
(Swerts, Bouwhuis & Collier)

judgment of spoken sentence finality (end of sentence)
independent variables:

- register
- range
- slope of pitch descent

Finality judgements can be predicted as an additive weighted combination of all three variables , with an interaction between two

3TU School Human Technology Interaction

There is **no** statistical test that can produce this result with these data.

A model of the data, based on perceptual and cognitive concepts, is far more sensitive to the process properties reflected in the data than any statistical test

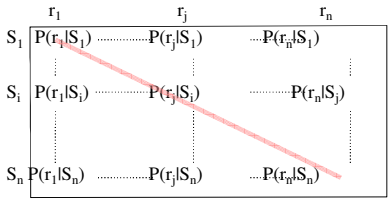
By modelling it is easier to obtain and evaluate convergent evidence.

Aim first at modelling, and include only statistics when journal editors complain about the absence of an ANOVA. Or refuse.

3TU School Human Technology Interaction

The two data models — 1. Identification

diagonal represents correct responses



3TU School Human Technology Interaction

Three typical data models — 1. Identification

Identification data do not imply 'correctness'

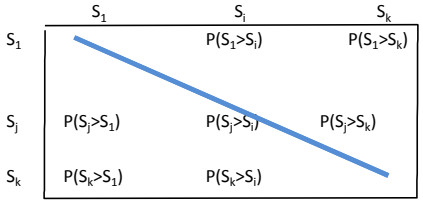
k stimuli may be assigned to n categories, k≠n

- Identification data require a model that relates properties of the stimuli to properties of the responses or categories
- this may take the form of modelling similarities, or psychological distances between stimuli and categories
- a favourable relation between # parameters and degrees of freedom is paramount.

3TU School Human Technology Interaction

The two data models — 2. Dominance Matrix

mostly no diagonal



3TU School Human Technology Interaction

The two data models — 2. Dominance Matrix

- The data represent whether the row stimulus exceeds the column stimulus in some well-specified attribute
- The extent to which one stimulus exceeds, dominates the other is usually modelled with Thurstone scaling
- Easier on the subject than rating as no internal criterion needs to be maintained; much used in quality judgments.
- statistical approach non-existent

3TU School Human Technology Interaction

Similarity or Distance data

- are usually represented as points in a 1- or multidimensional space.

possible solutions are obtained by

- factor analysis (Principal component analysis)
- non-metric multidimensional scaling (NMDS)

Factor analysis only operates on correlations.

NMDS operates on (ordinal) order relations and usually produces one dimension less than factor analysis.

Similarity or Distance data

problems

- NMDS very sensitive to initial configuration.
- leads easily to degenerate solutions
- Factor analysis has uninformative error measure

shared problems

- solution is only unique up to k-dimensional rotation, when k factors have been retained. Meaningful interpretation of the dimensions, or factors can be nearly impossible.
- Flexible rotation-, projection- and visualization- procedures are necessary, but offer no guarantee of success.

Similarity or Distance data

problem of degenerate NMDS solutions

- The first obtained solution of an NMDS analysis is almost certainly useless and has a too high stress value.
- Geometric solutions with a 'hole' in the middle (horseshoe shape, circle) are certainly degenerate. New stimuli should be added and analyses rerun.
- All available information should be used to generate a likely initial configuration, and reruns should be made until convergence is attained. The stress value can attain a zero value.

Measures in Context

ISO 9241

- effectiveness
- efficiency
- satisfaction
- learnability
- memorability

all measures directed at office tasks

Measures in Context

consumers do not so much consider the interaction, but the **result** of the interaction

- enjoyment
- fun
- emotion
- relief
- Quality of Experience — QoE