Social Robotics

PDEEC PhD course on Social Robotics

João Silva Sequeira[1]

[1] Instituto Superior Técnico
Institute for Systems and Robotics

Spring 2021

TÉCNICO LISBOA

Doing research in Social Sciences I

“Systematic use of some set of theoretical and empirical tools to try to increase our understanding of some set of phenomena or events. In the social and behavioral sciences, the phenomena of interest involve states and actions of human systems – of individuals, groups, organizations, and larger social entities – and the by-products of those actions.”, [McGrath, 1995]

Researching in SSH involves

a *Content* that is of interest,

b *Ideas* that give meaning to that content, and

c *Techniques* or procedures by means of which those ideas and contents can be studied.

Biasing factors I

a **Giving instructions to participants**, e.g., trying to motivate them to try hard by telling them that there will be a valuable prize for the best product

b **Imposing constraints on features of the environment**, e.g., providing some participants with a particular software program that may help task performance, and providing other participants with a different or no program to carry out that function

c **Selecting materials for use**, e.g., trying to produce differences in task difficulty by giving some participants very difficult word problems to complete, and giving other participants easier problems of the same type

Biasing factors II

d **Giving feedback about prior performances**, e.g., trying to induce feelings of success or failure by telling some participants they did well, and telling others they did poorly, on a previous task

e **Using experimental confederates**, e.g., trying to establish different degrees of liking for fellow group members by having an experimental assistant who is pretending to be a normal participant work very hard in some groups and act indifferent in others

Abstract experiment goals – Criteria to maximize

A Generalizability of the evidence over the sets of actors (entities whose behavior is under study)

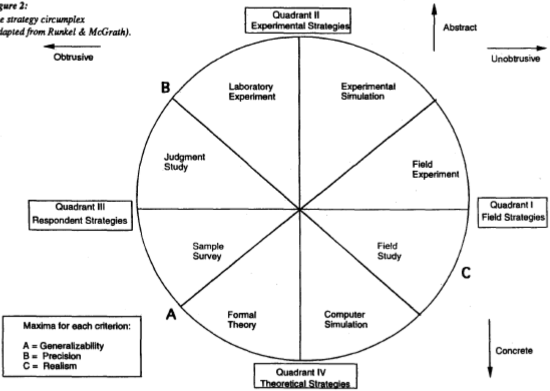
B Precision of measurement of the behaviors that are being studied (and precision of control over extraneous factors that are not being studied).

C Realism of the situation or context within which the evidence is gathered, in relation to the contexts to which you want your evidence to apply.

Not possible to maximize these 3 factors simultaneously, [McGrath, 1995]

Research methodologies

Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



Randomization and “true experiments” I

- You can only measure, match, control and manipulate a limited number of variables in any study
- In general, there will be many more factors that are potentially important to the phenomena being studied (and their existence is known)
- Randomization of the “extra” factors, or random assignment of cases to conditions, is the accepted practice

Techniques for manipulating variables I

Selection Selection is often the most convenient means to make sure that all cases of a given condition are alike on a certain variable -that all are 6-year-olds, or females, or juries-that-dealt-with-a-murder-case-and that all the cases of another comparison condition differ on that variable -being all 10-year-olds, or males, or juries-that-dealt-with-a-civil-suit

Direct intervention Manipulation by direct intervention in the structures and processes of the ongoing system that is being studied is the surest way of achieving a definite and specifiable manipulation, at least for those situations in which it can be done

Inductions Manipulations by less direct interventions are called experimental inductions

Results and validity I

a Results depend on methods. All methods have limitations. Hence, any set of results is limited.

b It is not possible to maximize all desirable features of method in anyone study; tradeoffs and dilemmas are involved.

c Each study (each set of results) must be interpreted in relation to other evidence bearing on the same questions.

Results and validity II

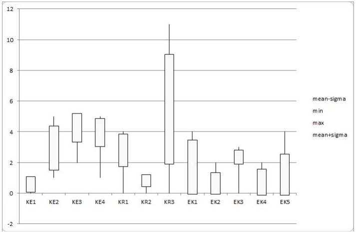
- Internal validity – to which degree to which results of a study permit strong inferences about causal relations ?
- Statistical validity – can results be due to chance ?
- Construct validity – how well defined are the theoretical ideas of your study?
- External validity – how confident you can be that the findings of your study will hold up upon replication, and how confidently you can predict both the range over which your findings will hold and the limits beyond which they will not hold ?

Evaluating experiments in social robotics I

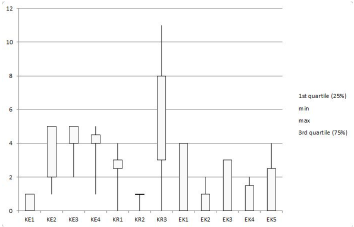
- Time and event-based data
- Noise and outliers always present
- Statistical analysis often used (statistical inference, nonparametric statistical techniques, distance between distribution estimates)
    - Is data consistent with a specific distribution ?
    - What's the amount of “bad data” are we willing to accept ?

Evaluating experiments in social robotics II

- Boxplot graphs provide information on mean, median and dispersion



1σ dispersion around the mean



1st and 3rd quartile dispersion around median

TÉCNICO LISBOA

Hypothesis testing I

- Define the *a priori* accepted hypothesis, H_0, (the *null hypothesis* - formulating H_0 in terms of statistical variables may not be simple)
- H_0 must be defined in terms of the data obtained from the experiments
- H_0 must be stated such that no one can think that you are biasing the process

  H_0 must be such that you “dare” to accept it, i.e, it represents your “worst fears”

Hypothesis testing II

- Is the data consistent with some statistical model, i.e., can it have been generated by a given statistical model ?
- Compute a distance between data statistics and the assumed model, and for some threshold, decide what to do with H_0

Hypothesis testing III

- Try to disprove H_0 for some confidence level (usually denoted by α), i.e., when disproving H_0 it is assumed that you accept a max of α% of the data obtained is purely random

- Compute a p-value and check if it is smaller/bigger than α

- A p-value is the evidence against H_0
  - $p \leq \alpha$ —> Reject H_0 —> Too much random data indicate that rejection is best option

  - $p > \alpha$ —> Accept H_0 —> Too few random data indicate that acceptance is best option

TÉCNICO LISBOA

Hypothesis testing IV

- Get the data under testing
- Compute adequate statistics for the hypothesis being tested
- Check the limit (critical) value for the chosen admissible error (significance) in a table for the statistics distribution
- Decision can be taken by comparing the computed statistics with the limit value found
- Alternatively, compute the probability mass at a tail of the distribution (p-value) and compare it with the significance value

Hypothesis testing V

- Distance between statistical models can be evaluated by multiple forms, e.g., Kullback-Liebler metric

  The KL Divergence (information discrimination) between distributions

  Often interpreted as the information gain by using distribution P instead to that by using Q

$$D_{KL}(P, Q) = -\sum_i P(i) \log \left(\frac{Q(i)}{P(i)} \right)$$

Hypothesis testing VI

- Bhattacharyya coefficients / Bhattacharyya distance / Hellinger distance

$$BC(Q,P) = \sum_i \sqrt{q_i\, p_i}$$

$$BD(Q,P) = -\log(BC(Q,P))$$

$$HD(Q,P) = \sqrt{1 - BC(Q,P)}$$

Comparisons among similar types of assessment component within different years using Bhattacharyya coefficients Theory (T), Lab (L), and Free (F) components

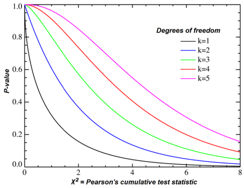
Years	T-T	L-L	F-F
17-18 / 18-19	0.9037	0.9617	0.7646
18-19 / 19-20	0.9711	0.9227	0.8286
19-20 / 17-18	0.9290	0.9736	0.7067
20-21 / 17-18	0.6032	0.8656	0.5306
20-21 / 18-19	0.5801	0.7664	0.6764
20-21 / 19-20	0.6919	0.9102	0.8504

Source: “Profiling ECE students through horizontal skills”, J. Sequeira, CISPEE 2021

Hypothesis testing VII

- χ^2-test - Goodness of fit for a multinomial distribution - how significant are the differences between expected and observed frequencies for K data classes
  - Compute the χ^2-square statistic

$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$



where O_i are the observations and the E_i the expected values

Q follows a χ^2 distribution of $k - 1$ degrees of freedom

  - Use a χ^2 distribution to obtain a p-value and accept/reject the null hypothesis

Hypothesis testing VIII

- Lilliefors test – Normality test

  The test creates its own baseline statistical estimates from the dataset

- Kruskal-Wallis test – Two samples originate from the same population

  Assume that k independent samples from a source population are drawn

  Assume that the source population has a normal distribution

Hypothesis testing IX

- Mann-Whitney test – Samples from independent populations are equally likely (come from identical distributions)

  Normality of the populations not necessary

  Independence between samples is assumed

- F-test – Compares two samples

$$F = \frac{2\,P\,R}{P + R} = \frac{\sigma^2_{explainable}}{\sigma^2_{not\ \ explainable}}$$

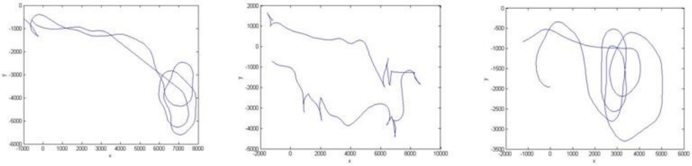
  Reject the null hypothesis when the F is high, meaning that your “worst fears” are not true because what you can’t explain is “smaller” than what you can

- In general it is a good idea to use multiple techniques to test H_0

TÉCNICO LISBOA

Non-statistical metrics - expressiveness of motion I

- The expressiveness of a trajectory/path is often related to the level of chaos/entropy



Source: [Sequeira, 2008]

- Typical measures for the degree of “order/chaos” in a trajectory include
  - Fractal dimension of the trajectory (measures self-similarity characteristics)
  - Fourier components for the trajectory
- Some authors suggest that (expressiveness) features can be combined linearly (see [Ceruti, Robin, 2007])

$$E = c_1 e_1 + c_2 e_2 + c_3 e_3 + c_4 e_4 + \cdots$$

Getting data I

- Direct video analysis – video segmentation (in micro-behaviors)
- Questionnaires – the sequence, number, and type of questions must be carefully designed

Getting data II

Source: FP7 URUS project

- Pioneer wheeled robot
- Robot head and torso developed by Scuola Superiore Sant'Anna and RoboTech, Pisa, Italy
- Does this setup allows quality HRI ?

TÉCNICO LISBOA

Data from Likert questionnaire I

- The questionnaire consisted of:

  a) 4 questions in Likert scale:

  1) I liked to interact with the robot
  2) The robot scares me
  3) I like the robot physical appearance
  4) It's been easy to interact with the robot

  b) 1 multiple choice test:

  5) What did you like most during the interacting with the robot? Possible answers: a) The physical appearance, b) the facial expressions, c) the voice, d) the touch-screen.

  C) 1 closed questions:

  6) Interacting with the robot has been more like interacting with a machine or with a human being?

Cuestionario URUS project

Barcelona, 15-07-2009

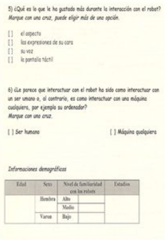
Marque con una cruz.

1) Me ha gustado interactuar con el robot

2) El robot me da miedo.

3) Me gusta el aspecto del robot

4) Ha sido fácil interactuar con el robot



5) ¿Qué es lo que le ha gustado más durante la interacción con el robot?
Marque con una cruz, puede eligir más de una opción.

[ ] el aspecto
[ ] las expresiones de su cara
[ ] su voz
[ ] la pantalla táctil

6) ¿Le parece que interactuar con el robot ha sido como interactuar con un ser umano o, al contrario, es como interactuar con una máquina cualquiera, por ejemplo su ordenador?
Marque con una cruz.

[ ] Ser humano [ ] Máquina qualquiera

Informaciones demográficas

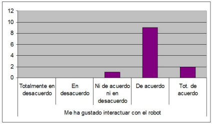
Edad	Sexo	Nivel de familiaridad con los robots		Estudios
	Hembra	Alto		
		Medio		
	Varon	Bajo		

- The above is an example of a 5-points questionnaire

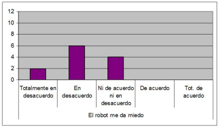
- 7-point and 10-point questionnaires are also commonly used

Data from Likert questionnaire II

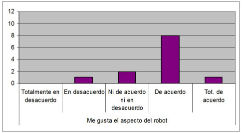
Question no. 1:
"I liked to interact with the robot"



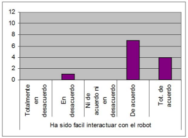
Question no. 2
"The robot scares me"



Question no. 3
"I like the robot physical appearance



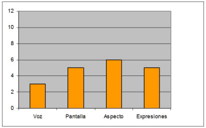
Question no. 4
"It's been easy to interact with the robot"



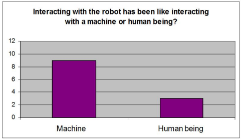
TÉCNICO LISBOA

Data from Likert questionnaire III

Question no. 5
“What did you like most during the interacting with the robot? ”



Question no. 6
“Interacting with the robot has been more like interacting with a machine or with a human being?”



Setting up an experiment in social robotics I

1. Do not stay in the activity area
2. Do not interfere, even if the experiment is not going your way
3. Plan well ahead the data collection
4. Avoid large data (bag) files – often difficult to deal with by computer programs
5. Expect the unexpected (computer programs are never error free) and observe “rules” 1 and 2
6. Have a Wizard-of-Oz mode running in parallel ready to take control of the experiment (the backup plan)

Towards a new Psychology for HRI ?

- Some researchers advocate that a new psychology is needed to study HRI (see, for instance [Turkle et al, 2006]) – robots may be extended egos of one's self

'Heinz Kohut describes how some people may shore up their fragile sense of self by turning another person into a self object.

In the role of self object, the other is experienced as part of the self, thus in perfect tune with the fragile individual's inner state. Disappointments inevitably follow. Relational artifacts (not only as they exist now but as their designers promise they will soon be) clearly present themselves as candidates for such a role. If they can give the appearance of aliveness and yet not disappoint, they may even have a comparative advantage over people, and open new possibilities for narcissistic experience with machines. One might even say that when people turn other people into self-objects, they are making an effort to turn a person into a kind of spare part. From this point of view, relational artifacts make a certain amount of sense as successors to the always-resistant human material.'

- **So, ... interesting times for Roboticists if robots start visiting Psychologists ...** 

TÉCNICO LISBOA

Local bibliography

Joseph E. McGrath
"Methodology Matters: Doing Research in the Behavioral and Social Sciences"
Readings in Human–Computer Interaction (Second Edition): Toward the Year 2000
Interactive Technologies, 1995, Pages 152-169

S. Turkle, W. Taggart, C. Kidd, O. Dasté
Relational artifacts with children and elders: the complexities of cybercompanionship
Connection Science, 18(4):347-361, December 2006

M. Ceruti, S. Robin
Infodynamics: Analogical analysis of states of matter and information
In Int. Journal on Information Sciences 177:969-987, 2007

J. Sequeira
Measuring expressiveness in wheeled mobile robots
In European Robotics Symposium 2008, EUROS 2008