Cognitive Science:
Cognitive modelling + experimental design

Jon Oberlander

From the original slide sets by Frank Keller

School of informatics

Cognitive Modelling: Structure

Introduction

What is cognitive modeling?

Intuitive Definition

A model is an artificial system that behaves the same way as a natural system (in certain interesting respects).

Questions that need to be addressed:

- What kinds of natural systems are being modeled?
- What kinds of artificial systems are used for modeling?
- What does it mean to behave in the same way as a natural system?

Natural systems

Physical Processes

- meteorology: model development and interaction of weather conditions (e.g., forecasting, microclimates, climatic change)

Biological Processes

- molecular biology: model the structure, function and dynamics of biological macromolecules (e.g., protein folding);
- evolutionary biology: model the ecological processes and genetic mechanisms that cause evolutionary change.

Natural systems

Psychological Processes

- social psychology: model the behavior of a group of interacting agents;
- cognitive psychology:
  - memory: storage and retrieval; learning and practice
  - vision: feature detection, object recognition
  - reasoning: problem solving, deduction, categorization
  - language: comprehension, production, acquisition

Types of models

Models of increasing explicitness:

- schematic model (boxes and arrows);
- mathematical model (set of equations);
- computational model (algorithm).

Example: Cogent is a cognitive architecture that combines schematic and computational modeling.

Main elements of a cogent model

Schematic representation:

- buffers: store information; e.g., model short term memory, long term memory;
- processes: move information from buffer to buffer and change its representation; e.g., model input/output, rehearsal;
- model needs to specify how buffers and processes communicate (restrictions imposed by the experimental data).

Main elements of a cogent model

Algorithmic representation:

- buffers represent information as predicates (e.g., `word(yellow)`); Prolog-style unification is used;
- properties like buffer capacity, rate of decay from buffer can be specified;
- processes manipulate information using production rules (IF-THEN clauses); Prolog-style auxiliary clauses can also be defined;
- properties of processes can be specified (e.g., if rules fire in parallel or serially).

Example: model of memory retrieval in Cogent.

Example: Modal model of memory

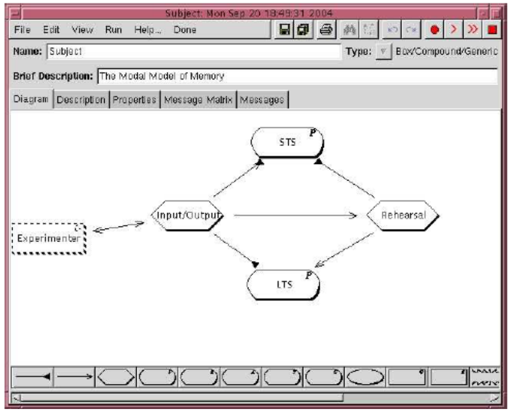
Experiment: subjects have to memorize lists of random words.

1. white	5. indigo	9. azure
2. cyan	6. brown	10. green
3. orange	7. black	11. magenta
4. yellow	8. beige	12. gray

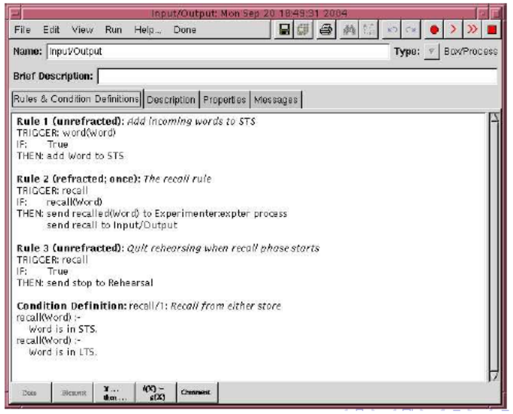
Then they have to recall as many words from memory as they can.

The results show that retention depends on the serial position of the word in the list. A U-shaped learning curve is obtained.

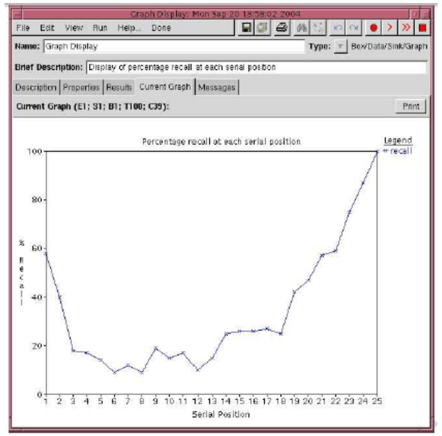
Example: Modal model of memory



Example: Modal model of memory



Example: Modal model of memory



Cognitive architectures vs Rational analysis

Complementary approaches to cognitive modeling (Chater and Oaksford 1999):

Traditional: **mechanistic approach:**

- analyze cognitive phenomena (memory, reasoning, language) regarding their causal structure
- stipulate architectures, algorithms, cognitive constraints
- the goal structure (i.e., functional relationships) remain largely unspecified

Rational analysis

Alternative: **goal-oriented approach:**

- analyze cognitive phenomena regarding their functions (similar approach in biology, social sciences, economics);
- rational analysis: assume that the cognitive system is optimally adapted to the task it has to perform (but: resource limitations)
- historically, this approach is related to probability theory; Bayesian mathematics often used to formulate models.

Rational analysis

Methodology (Anderson 1990):

1. **Goals:** specify the goals of the cognitive system
2. **Environment:** develop a formal model of the environment to which the systems is adapted
3. **Computational Limitations:** make minimal assumptions regarding the cognitive limitations of the system
4. **Optimization:** derive an optimal behavioral function based on (1)–(3)
5. **Data:** evaluate the optimal behavioral function based on empirical data
6. **Iteration:** repeat (1)–(5); iterative refinement

Example: Wason selection task

A	K	2	7

Experimental task:

- every card has a letter on one side, and a number on the other side
- test the following rule: if there is an A on the one side, then there is a 2 on the other
- turn over the cards that allow you to decide whether the rule holds or not.

Example: Wason selection task

A	K	2	7

Logical:

- $p \to q \Rightarrow \neg q \to \neg p$
- $A \to 2 \Rightarrow \neg 2 \to \neg A \Rightarrow 7 \to \neg A$

Empirical:

- A only (33%)
- A and 2 (46%)
- A and 2 and 7 (7%)
- A and 7 (4%)

Rational analysis is not the same as Optimal

Explanation for seemingly irrational behavior:

- logical principles are not very helpful for day-to-day reasoning because some events (p, q) are rare
- e.g., if you drop a plate (p), you hear a noise (q)
- for rare events, confirming evidence is more informative that falsifying evidence
- obtaining evidence if often costly; there is an advantage for rare, but informative evidence over frequent, but less informative evidence

Rational analysis of the Wason task

1. **Goals:** select the data that have the highest expected information gain (I) in determining whether the rule is true or false
2. **Environment:** confirming evidence q, p is rare $\Rightarrow I(q) > I(\neg q)$
3. **Computational Limitations:** obtaining evidence is costly $\Rightarrow$ minimize the amount of evidence required
4. **Optimization:** ODS (optimal data selection) model: subjects select the most informative evidence given (1) and (2) (formally: Bayesian model) $\Rightarrow I(A) > I(2) > I(7) > I(K)$
5. **Data:** prediction: for one cards, A is selected most of the time; for two cards, A and 2; for three cards, A, 2, 7
6. **Iteration:** new prediction: performance in the Wason task should change if (2) (rarity) is violated

Cognitive Modelling: Summary

- a cognitive model is an artificial system that behaves like a natural cognitive system;
- models can be schematic, mathematical, or computational;
- cognitive architectures (e.g., Cogent):
  - symbolic representations;
  - buffers: store information; processes: manipulate information;
  - emphasis on architecture, i.e., on how buffers and processes communicate.
- rational analysis:
  - emphasis on the function of a cognitive system;
  - analyze the goals, environment, limitations of the system;
  - assume that the system is optimally adapted;
  - implemented using Bayesian reasoning or probability theory.

Experimental Design: Structure

Introduction: An example problem

Scenario: You have designed and built an intelligent tutoring system to help students learn to program in Lisp.

- You are not sure though about giving feedback to students using it – what should you say and when?
- How do you go about making such a decision?
- Do some experiments with the tutoring system, with some students.

Based loosely on the experimental study of on Corbett and Anderson (1990).

Hypotheses

"Immediate Feedback is best!"
Hard to test

"There is a difference in performance between students given no feedback and students given immediate feedback."
More specific = **the experimental hypothesis**

"There is no difference in performance between students given no feedback and students given immediate feedback."
No effect = **the null hypothesis**

Variables

Things the experimenter can manipulate:

- Whether or not feedback is given.
- When it is given – immediately? After 3 errors of the same type? After certain types of errors? At the end of the session?
- What is given as feedback – correct or incorrect; detailed explanation; further examples?
- How much control does student have over feedback?

Variables

Things the experimenter can measure:

- How long does the student take to complete an exercise?
- What is the student's level of performance?
- How does the student feel about the different types of feedback – which do they prefer? Which do they feel they learn most from? Which do they learn most quickly with?
- How good are students at estimating their performance on a task?

Experimental conditions

Different conditions the experiment could compare:

1. immediate error feedback and correction
2. immediate error flagging but no correction
3. feedback on demand

Control condition:

- no feedback

Control condition (minimal experimental manipulation) is used as a standard of comparison.

Experimental variables

Independent Variable: manipulated by experimenter

Dependent Variable: not manipulated, but look to see if manipulating the independent variable has an effect on it (but not necessarily a causal relationship)

In our example:

- Independent Variable: type of feedback
- Dependent variable: time to complete the exercises; post-test performance
- What was taught remained constant; slight variations in the environments for teaching Lisp (assumed to be unimportant).

Alternative design

Independent Variables:

- immediate vs. delayed feedback
- short (right/wrong) vs. long (explanation) feedback

Control condition:

- no feedback

Experimental conditions (in a full factorial design):

1. immediate error feedback with explanation
2. immediate error feedback with right/wrong
3. delayed feedback with explanation
4. delayed feedback with right/wrong

Experimental subjects

Do we use the same subjects for each condition? Or different groups of subjects for each condition? Or matched subjects?

A. Same subjects (within group comparisons):

- each subject uses the tutor under all 4 conditions
- vary order of conditions to avoid order effects
- use isomorphic problems of equivalent difficulty, and vary these also across conditions

Pros

+ avoids individual differences
+ needs fewer subjects

Cons

– more complex design
– need isomorphic problems
– may still get order effects
– testing vs. learning issues
– fatigue/boredom

Subjects

B. Different subjects (between group comparison):

- different subjects undergo different conditions
- assume all from the same population

Pros

+ less order effects
+ simpler design

Cons

– individual differences
– needs more subjects

Subjects

C. Matched subjects (between groups, where pairs of subjects across groups are matched):

Could match on:

- intelligence
- previous number of years computing experience
- previous performance in another language (Prolog?)

Pros

+ as between groups
+ reduces individual differences

Cons

– hard to get good and appropriate matches

This study used between groups design, 55 students from the same undergraduate class. Assumed roughly same experience.

Correlational design

If this study showed that immediate feedback was best, we might want to follow it up by looking at the relationship between:

- performance on later Lisp assignments
- the amount of time spent using the tutor over the year

Does spending more time on the tutor correlate with better performance on later tests?

Warning: Correlation is not causation.

Example: if it doesn't rain, reservoirs dry out; if it doesn't rain, people stop using umbrellas.
Correlation between use of umbrellas and dry reservoirs; but no causation.

Levels of measurement

Nominal: Data are in categories; e.g., feedback vs. no feedback.

Ordinal: Data are rank ordered; e.g., spend no time/little time/lot of time on tutor; get good/medium/poor performance on post-test.

Interval: Data are on a continuous numerical scale with equal intervals between points; e.g., test scores (though sometimes really ordinal).

Ratio: As interval, with an absolute zero; e.g., time taken to complete exercises.

The level of measurement affects which statistical tests can be used on the data.

Results

Table 1 (Corbett and Anderson 1990):

	Immediate Feedback	Error Flagging	Demand Feedback	No Feedback
Post-test scores	55%	75%	75%	70%
Exercise times	4.6	3.9	4.5	4.5

Results

Table 3 (Corbett and Anderson 1990):

	Immediate Feedback	Error Flagging	Demand Feedback	No Feedback
1. difficulty	4.1	3.9	3.4	2.8*
2. learn material	5.2	4.6	5.4	5.8*
3. like tutor	5.2	4.5	4.8	4.9
4. help finish	5.1	4.6	4.7	4.5
5. help understand	5.3	4.9	4.7	4.7
6. like assistance	5.3	5.0	4.7	4.7
7. more assistance	4.3	4.9	4.5	4.6

Mean ratings on a scale from 1 to 7.

Discussion and conclusions

From Table 1:

- The effect of tutor type, as measured by post-test scores and mean exercise completion times, is not statistically significant.

So there is no evidence that feedback manipulation affected learning.

Discussion and conclusions

From Table 3: there were significant differences among the four groups in the questions:

1. How difficult were the exercises?
2. How well did you learn the material?

Interestingly, there is an inverse relationship between perceived difficulty and amount of feedback.

Discussion and conclusions

From Table 3: there were no significant differences among the four groups in the questions:

3. How much did you like the tutor?
4. Did the tutor help you finish more quickly?
5. Did the tutor help you understand better?
6. Did you like the tutor's assistance?
7. would you like more or less assistance?

Further data collected indicated however that students who received less assistance seemed more confident – but their confidence did not correlate with performance on the post-test.

Write-up

Abstract: summary of the problem, the results and the conclusion.

Introduction: problem statement; related work; derive testable hypothesis from general problem statement.

Method:

- **Subjects:** number, background and other relevant details of the subjects.
- **Materials:** test/teaching materials used; examples.
- **Procedure:** what data was collected and how: description of each stage in the experiment; enough information to replicate the experiment.

Write-up

Results: summary of the data; statistical analysis if appropriate; tables or graphs displaying the data; no interpretation.

Discussion: interpretation of the results; restating of the hypothesis and the implications of the results; discussion of any methodological problems (weaknesses in the design, unanticipated difficulties, confounding variables, etc.).

Conclusion: statement of overall conclusion of the study.

(This standard structure for write-up of experiments is borrowed from experimental psychology.)

Some issues and problems

Natural environment vs. ability to control variables (e.g., test in classroom vs. bring into laboratory).

Interference with subjects; ethical issues:

- Should you use a method of teaching that you don't think is going to work on your subjects?
- Should everyone get the opportunity to use the best approach?
- Will getting poor scores on a test that is not relevant to the curriculum affect student's morale (and their other work)?
- Should you use teaching time to do experiments?

Some issues and problems

Problems of measurement:

- What is improvement? Are we using the correct dependent variable?
- How long does the improvement last? Will students show an improvement if we retest them in a week's time?
- Do the results generalize? Does the improvement also show up in other tests and in other learning situations?

Experimental Design: Summary

- Before running an experiment, a testable hypothesis has to be formulated.
- An experimental design has to be developed that tests this hypothesis.
- It includes independent variables (manipulated by the experimenter) and dependent variables (measured by the experimenter) and control conditions.
- The design can include between groups, within groups, or matched subject comparisons.
- Levels of measurement: nominal, ordinal, interval, ratio.
- Standard structure for experimental write-ups.

References

Anderson, John R. 1990. *The Adaptive Character of Thought*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Chater, Nicholas and Mike Oaksford. 1999. Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences* 3(2):57–65.

Corbett, A. T. and J. R. Anderson. 1990. The effect of feedback control on learning to program with the Lisp Tutor. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Mahwah, NJ, pages 796–803.