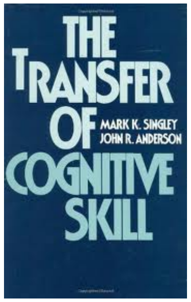
Learning and Transfer

Associated reading:
Singley, M. K. & Anderson, J. R. (1989). *Transfer of Cognitive Skill*. Cambridge, MA: Harvard University Press.



Vincent Aleven

8th Annual LearnLab Summer School

Pittsburgh, August 6-10, 2012

"The study of transfer is the study of how knowledge acquired in one situation applies (or fails to apply) in other situations."

Singley & Anderson, 1989, p. 1

" … the problem of transfer is perhaps the fundamental educational question. It is rare that people learn things in school which apply directly to life and work. For education to be effective, curricula must be designed with an eye toward transfer."

Singley & Anderson, 1989, p. 1-2

Examples of Transfer?

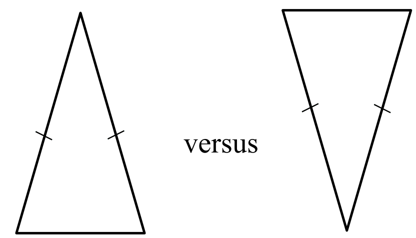
- Bubble sort for putting books on shelves
- Programming languages
- Fencing to boxing

- Other examples?

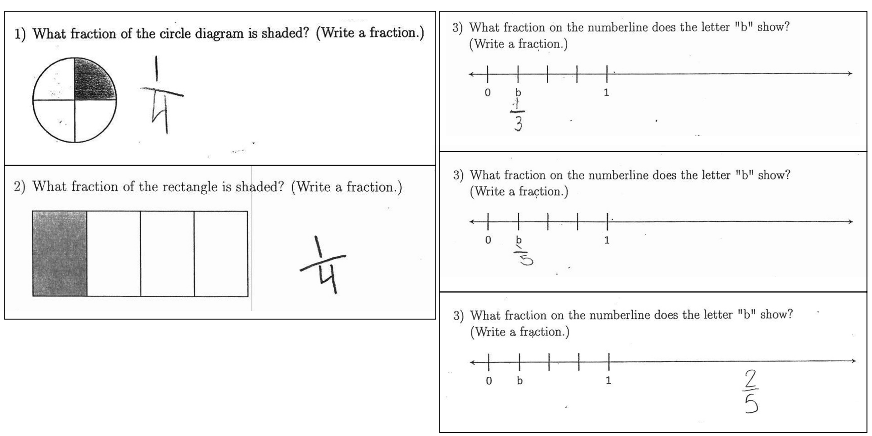
Examples of Lack of Transfer? Negative Transfer?

Example of Lack of Transfer?

Geometry Problem Solving



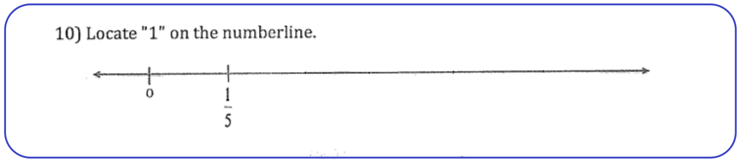
Lack of transfer?

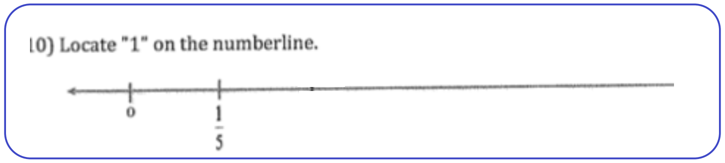


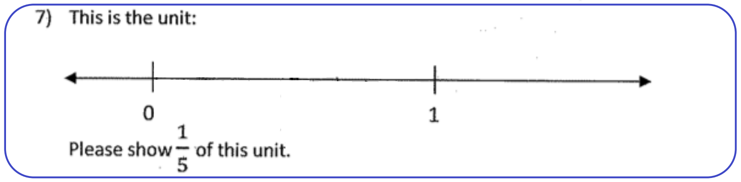
Unsuccessful Transfer?

- Student is able to deal with expressions of the form “2x + 3x” but not of the form “x + 3x”

Unsuccessful Transfer?







Overview

- **Transfer in Historical Perspective**
  - General vs. specific transfer
  - Meaningful vs. rote learning
  - Lateral vs. vertical transfer
- Contemporary Studies of Transfer
  - Analogical transfer
  - Specificity of transfer
- The Ghost of General Transfer
- Modeling Transfer in ACT-R

General vs. specific transfer

- General: Doctrine of formal discipline
  - Mind is muscle you exercise with subjects like Latin and geometry
    - General faculties of mind: Observation, attention, discrimination, reasoning
  - => Transfer is broad & general, across domains
    - Example: Training in chess transfers to computer programming b/c both involve reasoning faculty
- Specific: Thorndike's theory of identical elements
  - Mind is made up of stimulus-response elements
  - Transfer only occurs between tasks with elements in common
  - => Transfer is narrow, within domains

Thorndike's 1922 Experiment on Transfer

- Slight changes in equations led to significantly worse performance
  - Familiar: Square $x + y$ -> 94% correct
    Novel: Square $b_1 + b_2$ -> 72% correct
  - Familiar: Multiply x^a & x^b -> 44% correct
    Novel: Multiply 4^a & 4^b -> 30% correct
- Thorndike's point:

  Slight changes in stimulus
  -> stimulus-response element does **not** apply
  -> no transfer

But, some transfer is apparent

- One of Thorndike's 6 novel problems was not harder
  - Familiar: Solve $e^2 + ef = 8/x$; x -> 48% crt
    Novel: Solve $e^2 + ef = 8/p$; p -> 47% crt
- Performance on novel problems was about 50% on average
  - well above 0% expected if there was no transfer
- How might you explain?

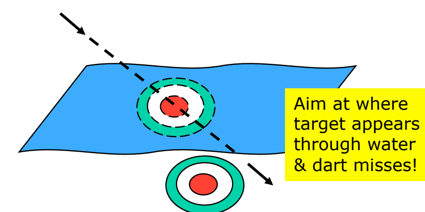
Weaknesses of Thorndike's Theory

- Did not allow for intelligent adaptation or flexible reconstruction of knowledge
  - ACT-R response: Declarative-procedural distinction. Circle the flexible one!
- No explicit representation language for cognitive skill
  => Vague about exact nature of "elements"
- Made no use of *abstract* mental representations
  - ACT-R response: Abstraction is a key feature of production rules

Meaningful vs. rote learning

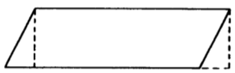
- Between the general-specific extremes:
  - Breadth of transfer dependent on type of instruction
  - *Transfer depends on whether a common representation can be found & communicated*

Judd's refraction study

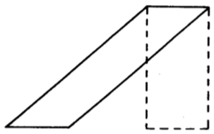


- Task: Throw darts at underwater target
  - Exper group instructed on refraction theory
  - Control group just practiced
  - Training task: Target was 12" under water
  - Transfer task: Target was 4" under water
- What happened during training?
  > *No difference in performance*
- What happened during transfer task?
  > *Experimental group did much better.* **Why?**
  > *Exper group had a better representation of the task & more flexibly adapted to new conditions*

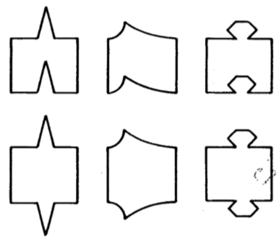
Wertheimer's Sensible and Insensible Transfer Tasks



(a) initial orientation and method



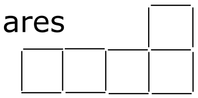
(b) new orientation and rote method



(c) sensible (top) and inse ısible transfer tasks

Katona's Puzzle Experiments

- Task: Move 3 sticks to make 4 squares



- Contrasted instruction on:
  - Rote strategy that applied to a particular problem
  - General strategy based on structural relations of an entire set of problems
    - Here are five squares composed of sixteen equal lines. We want to change these five squares into four similar squares. Since we have sixteen lines and want four squares, each square must have four independent side lines, which should not be side lines of any other square at the same time. Therefore, all lines with a *double function*, that is, limiting two squares at the same time, must be changed into lines with a *single function* (limiting one square only)"
- Rote Ss ~ better on trained problem, meaningful Ss much better on transfer

Lateral vs. vertical transfer

- Lateral transfer: Spreads over sets of same level of complexity
  - E.g., between different programming languages
- Vertical transfer: Spreads from lower-level to higher-level skills, from parts to whole
  - E.g., writing loops in isolation transfers to doing so in the context of a large problem
- Vertical transfer is common, lateral is rare
  - Vertical transfer was applied in early instructional design theories
    - Gagne & programmed instruction (Behaviorist)
    - Identify hierarchy of parts that need to be learned
    - Sequence instruction so that smaller parts are mastered first before larger wholes

Overview

- Transfer in Historical Perspective
  - General vs. specific transfer
  - Meaningful vs. rote learning
  - Lateral vs. vertical transfer
- **Contemporary Studies of Transfer**
  - Analogical transfer
  - Specificity of transfer
- The Ghost of General Transfer
- Modeling Transfer in ACT-R

Analogical transfer

Perfetto, Bransford, & Franks study:

- Ss solved insight problems like:
  A man who lived in a small town in the U.S. married twenty different women of the same town. All are still living and he has never divorced any of them. Yet, he has broken no law. Can you explain?
- Exper subjects had previously been asked to rate the truthfulness of a number of sentences including: “A minister marries several people each week”
- Such exposure had **no effect** on insight problem performance!
- If Ss were told explicitly of sentences relevance, then performance was improved

Analogical transfer (2)

- Duncker’s radiation task: Being exposed to problem analogue improved performance from 10% to 30%
  - Indicates substantial lack of transfer
- Providing a general statement of common underlying principle or providing a diagram did not help
  - How can this be reconciled with theories of rote v. meaningful learning?

Simple 2-Stage Model of Analogical Transfer

1. Retrieve a similar prior problem
2. Map it on to your current situation

- Many studies, like Perfetto et al., show difficulties with retrieval (#1)
- But in more complex domains, mapping (#2) is also a challenge
  - Need deep, not surface feature encodings of problems to make a productive mapping

Deep vs. Shallow Features -- Chi, Feltovich, Glaser

- Novice physics students categorize problems by surface features
  - pulley or inclined plane in diagram, similar words in problem text
- Experts categorize based on abstract, solution-relevant features
  - Problems solved using the same principle, e.g., conservation of momentum

Specificity of transfer

- How much transfer occurs depends on the way in which people “encode” or “represent” the problem situation.

Wason Card Selection Task

- Test whether this rule is true:
  - If a card has a vowel on one side, then it has an even number on the other side
- Which cards must you turn over to test the rule?

E	B	4	7

Wason Selection Task - Concrete Version

- Test whether this rule is true:
  - If a person is drinking alcohol, then he must be over 21
- Which cards must you turn over to test the rule?

Someone Drinking Alcohol	Someone Drinking Soft Drink	Someone Over 21	Someone Under 21

Abstract Selection Task Results

- Subjects said:
  - E & 4 -> 46%
  - E only -> 33%
  - E & 7 -> 4%
  - Other -> 17%
- Subjects with formal training in logic do not perform significantly better

=> People do not apply abstract logic rules -> contradicts doctrine of formal discipline

Concrete Selection Task Results

- Subjects had no difficulty whatsoever correctly selecting “drinking alcohol” (if-part) & “under 21” (not of then-part)
- Other scenarios involving social rules yield same results, rule need not be familiar:
  - If a person enters the country, then he must be tested for cholera.

=> Neither doctrine of formal discipline nor Thorndike’s identical S-R elements account for these results

=> People’s knowledge is induced from the ground up & intermediate in abstraction

Overview

- Transfer in Historical Perspective
  - General vs. specific transfer
  - Meaningful vs. rote learning
  - Lateral vs. vertical transfer
- Contemporary Studies of Transfer
  - Analogical transfer
  - Specificity of transfer
- **The Ghost of General Transfer**
- Modeling Transfer in ACT-R

The Ghost of General Transfer

- General transfer could “liberate students & teachers from the shackles of narrow, disciplinary education”
- Is general transfer possible?

Evidence Against General Transfer

- Thorndike’s original experiments disconfirming formal discipline
  - Latin & geometry courses don’t increase reasoning test scores any better than bookkeeping or shop courses
- Problem solving study (Post & Brennan)
  - Heuristics: determine given, check result
  - Did not transfer to word problem solving

Evidence for General Transfer

- *No evidence for **G**eneral transfer!*
- Some evidence for *limited* general transfer:
- LOGO programming (Carver & Klahr)
  - LOGO programming & debugging instruction transfers to other debugging tasks
- Math problem solving (Schoenfeld)
  - Heuristics with a specific "if-part" led to transfer, heuristics with a vague if-part did not transfer

Why is even limited general transfer hard to produce?

- Knowledge is largely domain specific
  - Simon estimate from chess studies: Expert's acquire > 10,000 chunks of knowledge
- General methods are often either:
  - Too vague to effectively apply
    - Heuristics like "avoid detail" depend on substantial domain-specific knowledge (which novices lack!) to distinguish irrelevant detail from key features
    - "Search paths simultaneously, use signs of progress" again depends on dom spec k to detect signs of progress
  - Effective ones may already be known by novices
    - Working backwards, means-ends analysis

Fundamental Design Challenge

- Specificity of transfer:
  - "The fundamental issue concerns the acquisition of a particular use of knowledge and *the range of circumstances over which that use will extend.*"
- *If-part* of production rules model this range of knowledge applicability
- Design challenge: How to identify this range in the domain we want to tutor?
  - Cognitive Task Analysis!

Overview

- Transfer in Historical Perspective
  - General vs. specific transfer
  - Meaningful vs. rote learning
  - Lateral vs. vertical transfer
- Contemporary Studies of Transfer
  - Analogical transfer
  - Specificity of transfer
- The Ghost of General Transfer
- **Modeling Transfer in ACT-R**
  - Declarative-procedural taxonomy

Modeling Transfer in ACT-R

- Chunks & production rules are the elements of transfer
- Transfer will occur to the extent these overlap between instructional & target task
- Productions & chunks can be acquired at varying levels of generality
  - Analogy induces productions of limited generality
  - Depends on how person encodes or views the task, what features they notice
  - The brain is conservative, makes minimal generalizations. Why? The more generalization, the greater risk of error (i.e., buggy productions)

Production rules have limited generality -- depending on purpose & context of acquisition

Overly general

IF "Num1 + Num2" appears in an expression
THEN
replace it with the sum

Leads to order of operations error:
"x * 3 + 4" is rewritten as "x * 7"

Overly specific

IF "ax + bx" appears in an expression and c = a + b
THEN
replace it with "cx"

Works for "2x + 3x" but not for "x + 3x"

Not explicitly taught

IF you want to find Unknown and the final result is Known-Result and the last step was to apply Last-Op to Last-Num,
THEN
Work backwards by inverting Last-Op and applying it to Known-Result and Last-Num

In "3x + 48 = 63":
63
- 48
-----
15 / 3 = 5 (no use of equations!)

Chunk Encoding & Transfer

- Chunks describe how instruction is "encoded" or understood
  - Drive how much learning transfer occurs
- Example: Different chunk representation of fractions

Shallow encoding:

QUANTITY2/3>
isa number
whole 0
top-number 2
bottom-number 3

Deep encoding:

QUANTITY2/3>
isa number
whole 0
top-number 2
bottom-number 3
parts-per-whole 3
needed-parts 2

Example ACT-R representations of card task

- In ACT-R, diff chunk types represent diff "encodings"
- Permission schema
  - Intermediate abstraction for encoding situations where social rules apply.
- ACT-R chunk for the encoding of drinking rule:
  Some-Chunk>
  isa permission
  what-you-can-do drink
  when-you-can-do-it older-21
- Associated production rules determine whether situation violates the permission
- Letter-number rule not encoded as a permission
  - Which part of this rule (vowel or even) goes in the "what-you-can-do" slot?
- Encoding of this rule in language processing chunk:
  Another-chunk>
  isa if-then-sentence
  if-part consonant-clause
  then-part even-num-clause
- What productions fire?
  - Permission productions do not apply to this chunk
  - Prods resulting from experience with if-then sentences -- not typically result of logic training.

Types of Transfer

		Target knowledge: Declarative	Target knowledge: Procedural
Source knowledge	Declarative		
	Procedural		

Check your understanding

- What kind of transfer going on in the studies discussed in this lecture?
  - Declarative-declarative, declarative-procedural, procedural-declarative, procedural-procedural?
- What kind of learning occurred in training?
- What kind of knowledge was required at transfer?

Types of Transfer

		Target knowledge: Declarative	Target knowledge: Procedural
Source knowledge	Declarative	Verbal reasoning. Ex: Duncker's radiation prob	Interpretive analogy. Ex: Judd refraction task
	Procedural	Meta-cognitive, learning skills. Ex: Reading comprehension	Production rule overlap. Ex:Text editor to text editor

Implications for instructional design?

- E.g., designing a Cognitive Tutor?

Summary

- Enhancing performance does not necessarily enhance learning
  - Enhancing learning requires transfer
  - Transfer depends on how students *encode* instructional tasks & target tasks
- General transfer *does not* occur
  - People must learn "details" of a domain
  - If-part of production determines generality
- ACT-R provides a way to:
  - think about learning & transfer issues
  - assess how much vertical/lateral transfer is likely from instructional tasks to real world tasks

END