Issues in Transfer & Learning

Ken Koedinger
Human-Computer Interaction & Psychology
Carnegie Mellon University

CMU Director of the Pittsburgh Science of Learning Center

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

A task & a quick review of ACT-R theory before we get to transfer …

Read this story …

The General

A small country was ruled from a strong fortress by a dictator. The fortress was situated in the middle of the country, surrounded by farms and villages. Many roads led to the fortress through the countryside. A rebel general vowed to capture the fortress. The general knew that an attack by his entire army would capture the fortress. He gathered his army at the head of one of the roads, ready to launch a full-scale direct attack. However, the general then learned that the dictator had planted mines on each of the roads. The mines were set so that small bodies of men could pass over them safely, since the dictator needed to move his troops and workers to and from the fortress. However, any large force would detonate the mines. Not only would this blow up the road, but it would also destroy many neighboring villages. It therefore seemed impossible to capture the fortress.

However, the general devised a simple plan. He divided his army into small groups and dispatched each group to the head of a different road. When all was ready he gave the signal and each group marched down a different road. Each group continued down its road to the fortress so that the entire army arrived together at the fortress at the same time. In this way, the general captured the fortress and overthrew the dictator.

- Can you paraphrase it (to yourself)?
- Want to see it again?
- How plausible is the story?
  (Not at all) 1 2 3 4 5 (Very)
- How comprehensible is it?
  (Not at all) 1 2 3 4 5 (Very)

Quick review of ACT-R Theory underlying Cognitive Tutors

ACT-R: A Cognitive Theory of Learning and Performance

- Big theory … key tenets:
  - Learning by doing, not by listening or watching
  - Production rules represent performance knowledge:

These units are:		Instruction implications:
• modular	→	isolate skills, concepts, strategies
• context specific	→	address "when" as well as "how"

Anderson, J.R., & Lebiere, C. (1998). *Atomic Components of Thought*. Erlbaum.

ACT-R Theory of Cognition: Declarative-Procedural Distinction

- Declarative knowledge
  - Includes factual knowledge that people can report or describe, but can be non-verbal
  - Stores inputs of perception & includes visual memory
  - Is processed & transformed by procedural knowledge
  - Thus, it can be used *flexibly*, in multiple ways
- Procedural knowledge
  - Is only manifest in people's behavior, not open to inspection, cannot be directly verbalized
  - Is processed & transformed by fixed processes of the cognitive architecture
  - It is more specialized & *efficient*

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
- Knowledge Component Analysis

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: 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!
  - Note, there is substantial transfer as performance on novel tasks is far above 0%

Weaknesses of Thorndike's Theory

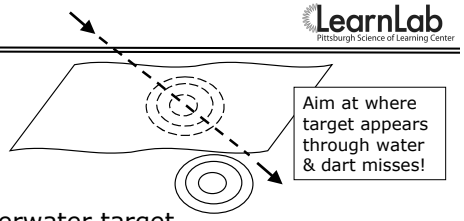
- Did not allow for intelligent adaptation or flexible reconstruction of knowledge
  - ACT-R response: Declarative-procedural distinction. Which is the flexible one?
- No explicit representation language for cognitive skill

  => 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*

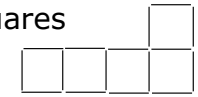
- What happened during transfer task?

> *Experimental group did much better.* **Why?**

> *Exper group had a better representation of the task & more flexibly adapted to new conditions*

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
- Knowledge Component Analysis

But before that a task …

Try this problem

Suppose you are a doctor faced with a patient who has a malignant tumor in his stomach. It is impossible to operate on the patient, but unless the tumor is destroyed the patient will die. There is a kind of ray that can be used to destroy the tumor. If the rays reach the tumor all at once at a sufficiently high intensity, the tumor will be destroyed. Unfortunately, at this intensity the healthy tissue that the rays pass through on the way to the tumor will also be destroyed. At lower intensities the rays are harmless to healthy tissue, but they will not affect the tumor either. What type of procedure might be used to destroy the tumor with the rays, and at the same time avoid destroying the healthy tissue?

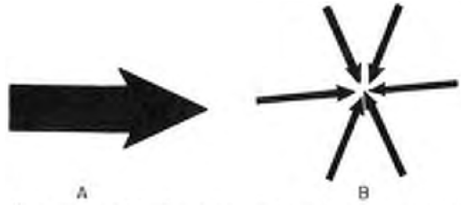
- How many think they have the answer?
- Here's a hint:
  Remember that story we started with …
- How many think they have it now?
- Here's another hint …

The Fire Chief

One night a fire broke out in a wood shed full of timber on Mr. Johnson's place. As soon as he saw flames he sounded the alarm, and within minutes dozens of neighbors were on the scene armed with buckets. The shed was already burning fiercely, and everyone was afraid that if it wasn't controlled quickly the house would go up next. Fortunately, the shed was right beside a lake, so there was plenty of water available. If a large volume of water could hit the fire at the same time, it would be extinguished. But with only small buckets to work with, it was hard to make any headway. The fire seemed to evaporate each bucket of water before it hit the wood. It looked like the house was doomed.

Just then the fire chief arrived. He immediately took charge and organized everyone. He had everyone fill their bucket and then wait in a circle surrounding the burning shed. As soon as the last man was prepared, the chief gave a shout and everyone threw their bucket of water at the fire. The force of all the water together dampened the fire right down, and it was quickly brought under control. Mr. Johnson was relieved that his house was saved, and the village council voted the fire chief a raise in pay.

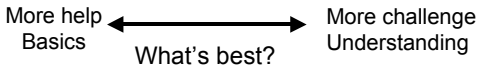
- How many think they have it now?
- Here's another hint …
  Success in these stories can be attributed to this important principle: If you need a large force to accomplish some purpose, but are prevented from applying such a force directly, many smaller forces applied simultaneously from different directions may work just as well.
- How many think they have it now?
- Here's another hint …

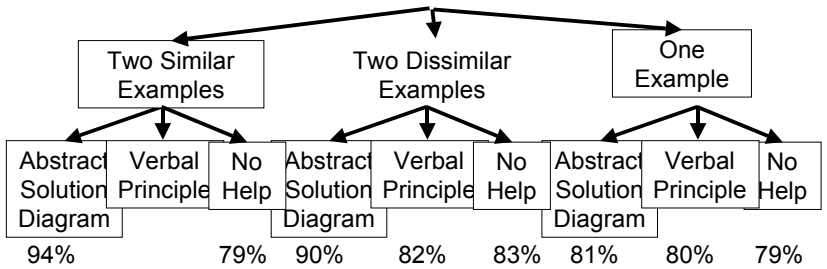


Analogical transfer

- These were materials from
  - Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1-38.
- Baseline results
  - Less than 10% solve Radiation problem at first exposure
  - After reading General story, only 29% without hint
  - After a hint of story relevance, 79% solve
- G & H had six experiments addressing various instructional strategies & some of their combinations …

Instructional options explored by Gick & Holyoak





Simple 2-Stage Model of Analogical Transfer

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

- Many studies, like G & H, 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**
- Knowledge Component Analysis

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
- **Knowledge Component Analysis**

Knowledge component (KC) analysis to model transfer

- Two kinds of KC's in ACT-R: declarative chunks & production rules
  - *Transfer occurs to the extent components overlap between instructional tasks & target tasks*
- Productions & chunks can be acquired at varying levels of generality
  - Analogy process induces productions of limited generality
  - Depends on how person encodes or views the task, what features they notice

Transfer-Enabling Knowledge Components (the "germs")

- In the Card Selection task, what are the relevant KCs that explain partial transfer?
  - What are "transfer-enabling" KCs that apply in novel versions of concrete task?
    - If entering country, then cholera test
  - Why don't these KCs apply in abstract task?
- In other words, what KC is general, but not too general?

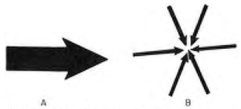
How is the Gick & Holyoak task like academic learning?

- Is the following a good *analogy*?!

Convergence problems

↓

Convergence solution



Algebra story problems

↓

Algebraic equation solution ***s***

800 - 40x d - 1/3d

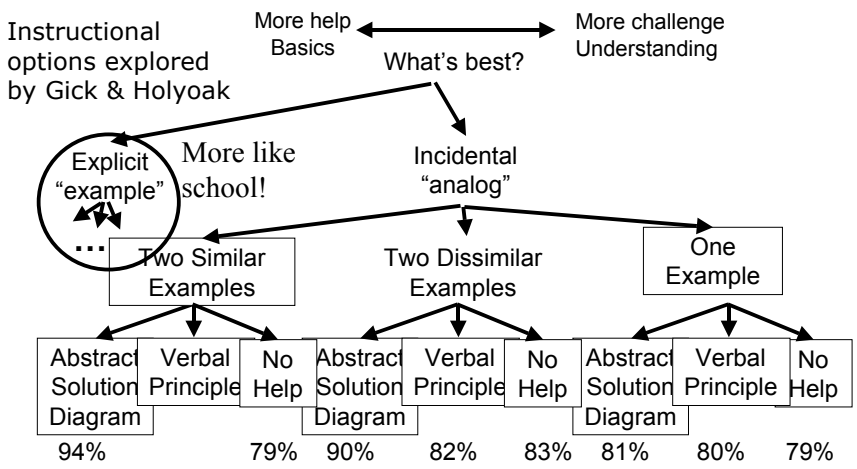
(n - 20)/3 …

How is the Gick & Holyoak task like academic learning?

- Is the following a good *analogy*?!

Improve transfer across *convergence* problems
↓
Optimal instruction for *convergence*

Improve transfer across *algebra* problems
↓
Optimal instruction for *algebra*



Summary

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

For more information on this topic see:
Singley, M. K., & Anderson, J. R. (1989). *Transfer of Cognitive Skill*. Hillsdale, NJ: Erlbaum.

END

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 vowel-clause
    then-part even-num-clause
```

- What productions fire?
  - Permission productions do not apply to this chunk
  - Productions resulting from experience with if-then sentences -- not typically result of logic training.

Types of Transfer

Source knowledge \ Target knowledge	Declarative	Procedural
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

Questions to check your understanding

- What kind of transfer is going on in Judd's Refraction study?
  - Declarative-declarative, declarative-procedural, procedural-declarative, procedural-procedural?
- What kind of learning occurred in training?
- What kind of knowledge was required at transfer?

ACT-R Analogy Mechanism

- Steps:
  - *Find* an example that had a similar goal
  - *Map* goal structure of example to problem
  - *Apply* mapping to response structure of example to get response structure for current goal
  - *Check* preconditions