From Causal Models to Analogical Inference

Keith Holyoak
Dept. of Psychology
University of California, Los Angeles

"911, Analogy Police... State your bad analogy"

Recent Review

- Holyoak, K. J. (2005). Analogy. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 117-142). Cambridge, UK: Cambridge University Press.

Oldies

- Hesse, M. (1966). *Models and analogies in science*. Notre Dame, IN: University of Notre Dame Press.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.
- Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation (Vol. 19)*. New York: Academic Press.
- Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.

Structure Mapping Theory

- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155-170.
- Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*, 1-63.

Learning and Inference with Schemas and Analogies (LISA)

- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427-466.
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220-264.
- Hummel, J. E., & Holyoak, K. J. (2005). Relational reasoning in a neurally-plausible cognitive architecture: An overview of the LISA project. *Current Directions in Cognitive Science*, *14*, 153-157.

Forthcoming

- Bartha, P. (in press). *By parallel reasoning: The construction and evaluation of analogical arguments*. Oxford, UK: Oxford University Press.

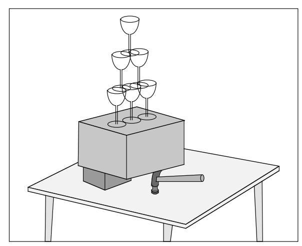
What is Analogy?

- novel *target* analog
- related to familiar *source* analog
- by a common pattern of *relations* among elements
- despite *different elements*
- to *draw inferences* about target

Relations in Perception and Cognition

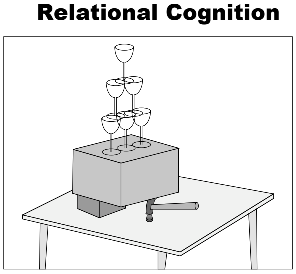
Perception and thinking are both constrained by *relations* between things rather than just the features of those things.

Relational Perception



Where are the wine glasses?

Relational Perception Meets Relational Cognition



You need a hammer. What do you *not* do?

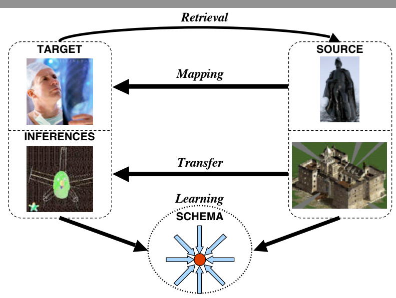
Analogy in Science
(Holyoak & Thagard, *Mental Leaps*, 1995)

"If genius has any common denominator, I would propose breadth of interest and the ability to construct fruitful analogies between fields." —Steven Jay Gould

Bohr/Rutherford model of atom

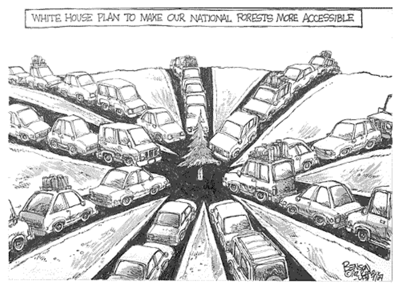
- •Sound / water waves (Vitruvius, 60 BC)
- •Earth / small magnet (Gilbert, 1600)
- •Earth / ship (Galileo, 1630)
- •Light / sound (Huygens, 1678)
- •Planet / projectile (Newton, 1687)
- •Heat / water (Carnot, 1824)
- •Natural / artificial selection (Darwin, 1859)
- •Chromosome / beaded string (Morgan, 1915)
- •Mind / computer (Turing, 1950)

Steps in Analogical Transfer



Gick & Holyoak (1980, 1983)

The convergence schema at work…



What does analogy have to do with probabilistic reasoning?

- Classic problem of induction from sparse data
- Long history in psychology and machine learning
- Builds on causal reasoning models
- Highlights issues of knowledge representation
- Highlights issues of cognitive capacity
- Provides a mechanism for forming new hypotheses

Causal models and analogical inference

How analogous is the moon to the earth? NO GENERAL ANSWER!

Is there life on the moon? Gimme a break…

Is there potential for mining on the moon? Sure, why not…

Analogy emphasizes relations such as:

- *physical-cause*(A, B)
- *logically-implies*(A, B)
- *enables*(A, B)
- *justifies*(A, B)
- *determines* (A, B).

Causal Relations and Analogical Inference: Experimental Tests

Lassaline (1996): inductive strength varies with relations

no connecting relation

Animal A: (1) weak immune system
(2) skin has no pigment
(3) dry flakey skin

Animal B: (1) weak immune system
(4) acute sense of smell

Does A have (4) acute sense of smell?

non-causal relation

Animal A: (1) weak immune system
(2) skin has no pigment
(3) dry flakey skin

Animal B: (1) weak immune system
(4) acute sense of smell
(1) develops before (4)

Does A have (4) acute sense of smell?

“cause” relation

Animal A: (1) weak immune system
(2) skin has no pigment
(3) dry flakey skin

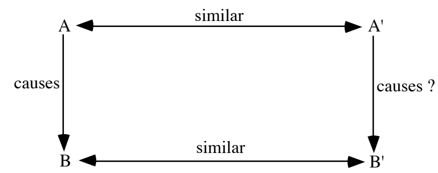
Animal B: (1) weak immune system
(4) acute sense of smell
(1) causes (4)

Does A have (4) acute sense of smell?

Inductive support: “cause” > “develops before” > no relation

Basic Scheme of Analogy

Similar causes are expected to have similar effects (cf. Hume):



P(B’|A’, A, B) ~ P(A’-->A, B’-->B) P(B|A)

mapping cause in source

Analogy as Isomorphism

$A = \langle S, T, m \rangle$

$S = \langle O_i, R_k, P_1, P_2, \ldots P_n \rangle$

where $P_i = R_k(o_i, o_j)$

$T = \langle O_i', R_k', P_1', P_2', \ldots P_n' \rangle$

$m: o_i \dashrightarrow o_i'; R_k \dashrightarrow R_k'; P_1 \dashrightarrow P_1'$

m defines an isomorphism iff

$R_k(o_i, o_j)$ implies $m(R_k)\ (m(o_i), m(o_j))$

Copy with Substitution & Generation (CWSG)

$A = \langle S, T, m \rangle$

$S = \langle O_i, R_k, P_1, P_2, \ldots P_n, P_m, P_{o\ldots} \rangle$

where $P_i = R_k(o_i, o_j)$

$T = \langle O_i', R_k', P_1', P_2', \ldots P_n' \rangle$

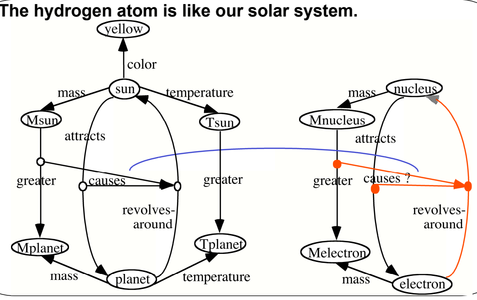
$m: o_i \dashrightarrow o_i'; R_k \dashrightarrow R_k'; P_1 \dashrightarrow P_1'$

Infer in T: $m(P_m), m(P_o)\ldots$

BIG QUESTION: How to constrain *plausible* inferences?

Learning by analogy: illustration

The hydrogen atom is like our solar system.



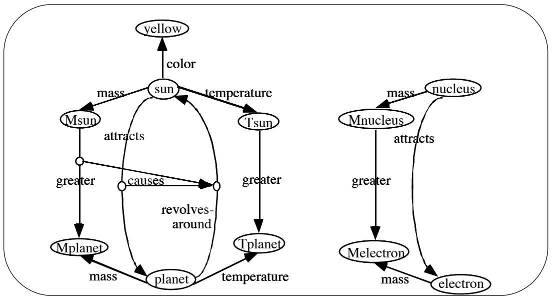
The Sun has a greater mass than the Earth and attracts it, causing the Earth to revolve around the Sun. The nucleus also has a greater mass then the electron and attracts it. Therefore it is plausible that the electron also revolves around the nucleus.

Learning by analogy: the general method

- **ACCESS: find a known entity S analogous to the novel entity T**
  How did Rutherford select the solar system as a source analog?
- **MAPPING: find correspondences between S and T**
  Map the nucleus to sun and the electron to planet.
- **INFERENCE: generate hypotheses by “copy with substitution and generation” (CWSG)**
  Perhaps electron revolves around the nucleus because the nucleus attracts the electron and the mass of the nucleus is greater than the mass of the electron.
- **EVALUATION: test the hypotheses**
  Design experiments to see if the electron revolves around nucleus.
- **LEARNING: generalize the new knowledge**
  By generalization from the solar system and the hydrogen atom, learn the abstract schema that a central force can cause revolution.

Potential mappings

Which are the possible mappings between the elements of S and the elements of T?



Potential mappings

There are several possible mappings between the elements of S and the elements of T, which need to be ordered by f_{map}:

Mapping1:
sun ↔ nucleus, planet ↔ electron, Msun ↔ Mnucleus, Mplanet ↔ Melectron,
which is supported by the following correspondences
mass(sun, Msun) ↔ mass(nucleus, Mnucleus)
mass(planet , Mplanet) ↔ mass(electron, Melectron)
greater(Msun, Mplanet) ↔ greater(Mnucleus, Melectron),
attracts(sun, planet) ↔ attracts(nucleus, electron)

Mapping2:
sun ↔ nucleus, planet ↔ electron, Tsun ↔ Mnucleus, Tplanet ↔ Melectron,
which is supported by the following correspondences
greater(Tsun, Tplanet) ↔ greater(Mnucleus, Melectron),
attracts(sun, planet) ↔ attracts(nucleus, electron)

Mapping3:
sun ↔ electron, planet ↔ nucleus, Msun ↔ Melectron, Mplanet ↔ Mnucleus

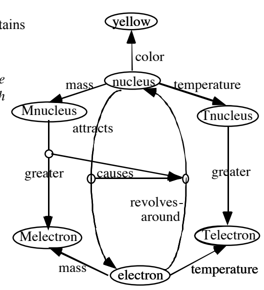
Possible analogical inferences

The best mapping is Mapping1 (because it leads to the highest number of common features of the solar system and the hydrogen atom), yielding correspondences:
m = (sun ← nucleus, planet ← electron, Msun ← Mnucleus, Mplanet ← Melectron)

By applying CWSG to the solar system, one obtains the following structure:

Propositions that might be transferred to the hydrogen atom as a result of the analogy with the solar system:

revolves-around(nucleus, electron)
- causes((attracts(nucleus,electron), greater(Mnucleus, Melectron)), revolves-around(nucleus, electron))
- color(nucleus, yellow)
- temperature(nucleus, Tn)
- temperature(electron, Te)
- greater(Tn, Te)



Evaluation

The evaluation phase (i.e., doing experiments) shows that

For the hydrogen atom it is *true* that:
- revolves-around(nucleus, electron)
- causes((attracts(nucleus,electron), greater(Mnucleus, Melectron)), revolves-around(nucleus, electron))

For the hydrogen atom it is *false* that:
- temperature(nucleus, Tn)
- temperature(electron, En)
- greater(Tn, En)

Learning & Generalization

Store the new knowledge about the hydrogen atom:

- revolves-around(nucleus, electron)
- causes(attracts(nucleus,electron), greater(Mnucleus, Melectron)), revolves-around(nucleus, electron))

By generalization from the solar system and the hydrogen atom, induce the abstract schema that a central force can cause revolution:

- causes(attracts(x, y) & greater(Mx, My)), revolves-around(x, y))

Gentner's Structure Mapping Theory

How is CWSG constrained?

Predicates from the source are carried across to the target, using the substitutions dictated by the object correspondences, according to the following rules:

1. Discard attributes of objects $A(s_i)$ -/-> $A(t_i)$
For instance, the yellow color of the sun is not transferred to the hydrogen nucleus.

2. Try to preserve relations between objects $R(s_i, s_j)$ -?-> $R(t_i, t_j)$
Some relations are transferred to the target, but others are not.

3. The systematicity principle: the relations that are most likely to be transferred are those belonging to systems of higher-order relations
$R'(R_1(s_i,s_j), R_2(s_k,s_l)) \rightarrow R'(R_1(t_i,t_j), R_2(t_k,t_l))$

Some problems with Structure Mapping theory

- **rules for generating mappings and inferences are purely syntactic**

- no simple way to match non-identical predicates:
but *murder* (x, y) --> *kill* (x', y')

- Impossible to map predicates with different numbers of arguments (*n*-ary constraint):
but *murder* (Abe, Chad), *killer* (Dave) supports Abe --> Dave

- Not all attributes can be neglected:
red(flag) *cause attack* (bull, flag) -->
red(shirt) *cause attack* (bull, shirt)

- All higher-order relations are not equal: *cause* vs *prior-to*

- *cause* relations can have very different implications for inference:
generate vs *prevent*

Multiconstraint theory

(Holyoak & Thagard, 1989)

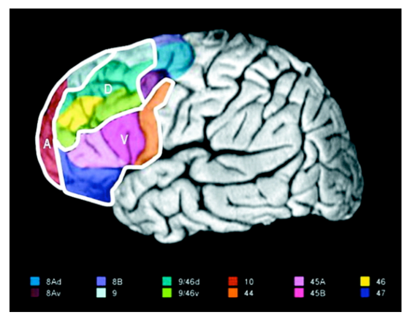
- **Isomorphism**: mappings should be structurally consistent and one-to-one
- **Semantic similarity**: mappings between similar elements are preferred
- **Pragmatic centrality**: mappings involving goal-relevant elements are preferred

Human Analogical Thinking

So common it seems easy… But it's not…

- Late evolutionary development
- Linked to size and complexity of frontal cortex
- Late to develop in children
- Perhaps uniquely human (Penn, Holyoak & Povinelli, forthcoming in *BBS*)

Diagram of the human prefrontal cortex (PFC; left lateral view)



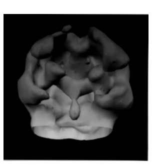
Rajah, M. N. et al. Brain 2005 128:1964-1983; doi:10.1093/brain/awh608

BRAIN
A journal of neurology

Relational Integration in Frontal Cortex
(Waltz et al., *Psych. Science*, 1999)

- **Frontotemporal Dementia patients**
- **Broad bilateral damage**
- **Two major variants**

Normal Brain



Frontal-Variant

- **personality changes**
- **dysexecutive changes**

Reasoning in the Brain FTD

- **FTD patients**
- **Two major variants**

Normal Brain

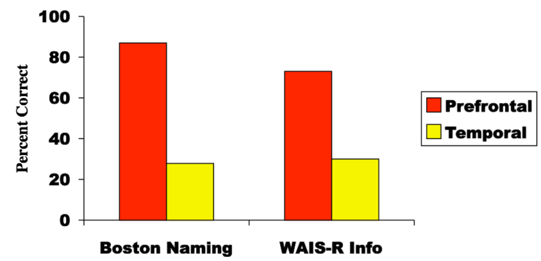
Temporal-Variant

- **semantic memory**
- **emotional changes**
- **preserved episodic & working memory**

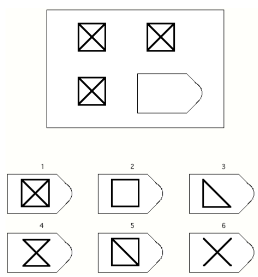
Prefrontal Cortex and Relational Integration

- **Approach**: systematically vary number of relations to be integrated
- **Prediction**: patients with prefrontal damage will exhibit a deficit with relational complexity >1
- **Subjects**: Patients with early stage fronto-temporal dementia (FTD)
- Patients divided into frontal variant and temporal variant groups

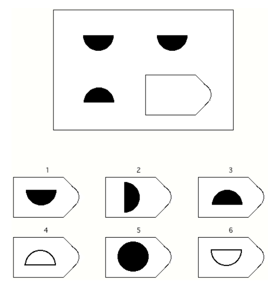
Performance on semantic knowledge tests



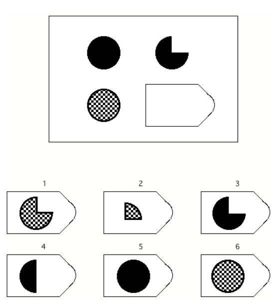
Level 0 Matrix Problem



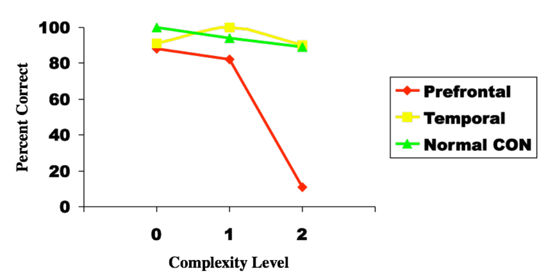
Level 1 Matrix Problem



Level 2 Matrix Problem

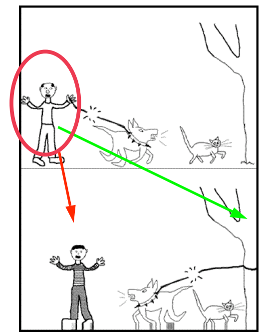


Matrix Problems



1

2



- Featural
- Relational

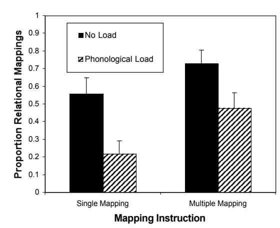
Analogical Mapping Under Dual-Task Conditions

(Waltz, Lau, Grewal & Holyoak, 2000)

- Map 3 objects or just 1
- Memory load of 7 random digits, visually presented before each picture pair (or no load)

Mapping Results

(Waltz et al., 2000, Ex 1)



Ways to impair human relational reasoning

- Impose memory load
- Add salient superficial distractors
- Cause anxiety
- Be a child
- Get old
- Inflict frontal damage

Human Relational Representations

Two key properties

1) *Explicitly Relational*

- Represent relational roles *explicitly* --> independently of their arguments
- Roles and arguments bound dynamically into compositional structures
- i.e., *Symbolic*
- Traditional symbolic approaches capture this aspect of human mental representation
- Traditional distributed connectionist representations **do not**

Feature vectors and propositional logics of Week 1: caveat!

Human Relational Representations
Two key properties

2) Semantically Rich

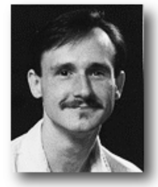
- Relational roles and their arguments have meaning:

  murder (*x*, *y*) --> *kill* (*x'*, *y'*) vs *greet* (x', y')

  killer (Abe), *murder* (Chad, Dave) supports Abe --> Chad
- Traditional distributed connectionist representations capture this aspect of human mental representation
- Traditional distributed symbolic representations **do not**

Predicate logics & grammars of Week 2: caveat!

Models of Analogy



- **LISA (Hummel & Holyoak, 1997, 2003)**
- **Algorithmic Level Model of Analogy**
  - **sensitive to computational level constraints**
  - **neurally & psychologically plausible**
  - **learning relational generalizations (schemas)**
  - **intrinsic working memory limits**

Working Memory, Inhibition, and Mapping

- LISA links the number of "active" relational roles to the capacity of WM
- LISA's performance depends on inhibitory control
- Both WM for relations and inhibitory control depend on prefrontal cortex

Knowledge Representation in LISA
("LISAese")

Symbolic Connectionism

- Neural-style computing architecture that gives rise to symbolic representations and processes
- Captures relations that are both explicit and semantically rich

LISAese

Hierarchy of distributed and localist codes

Bottom of the hierarchy: Distributed *semantic* units

LISAese

Distributed *semantic* units

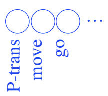
go-to (John, LAX)

LISAese

Distributed *semantic* units

go-to (John, LAX)

traveler role of the *go-to* relation

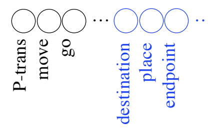


LISAese

Distributed *semantic* units

go-to (John, LAX)

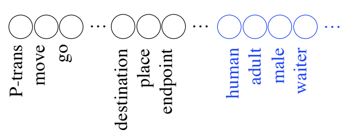
destination role of the *go-to* relation



LISAese

Distributed *semantic* units

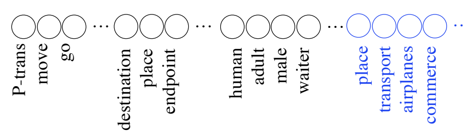
go-to (John, LAX)



LISAese

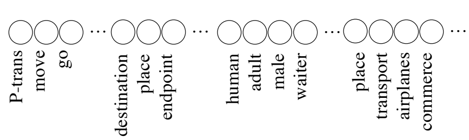
Distributed *semantic* units

go-to (John, LAX)



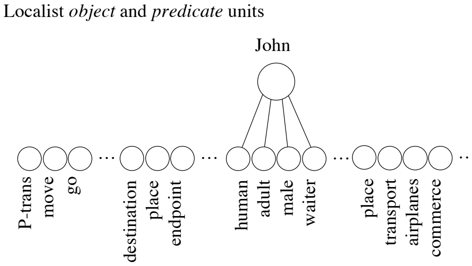
LISAese

Localist *object* and *predicate* units



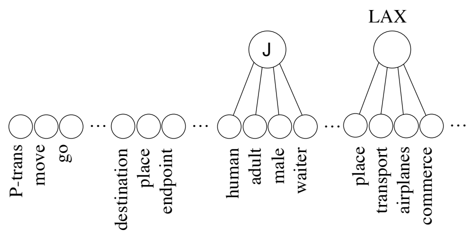
LISAese

Localist *object* and *predicate* units



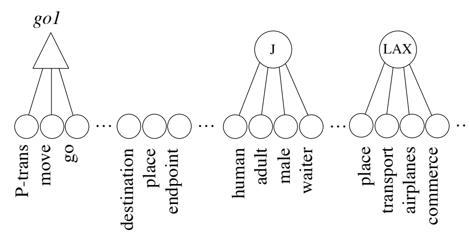
LISAese

Localist *object* and *predicate* units



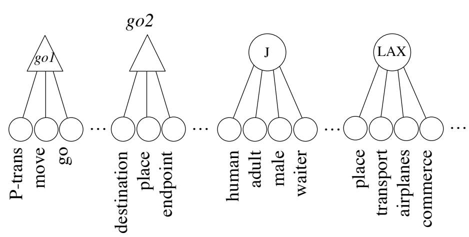
LISAese

Localist *object* and *predicate* units



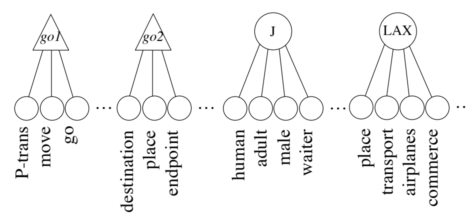
LISAese

Localist *object* and *predicate* units



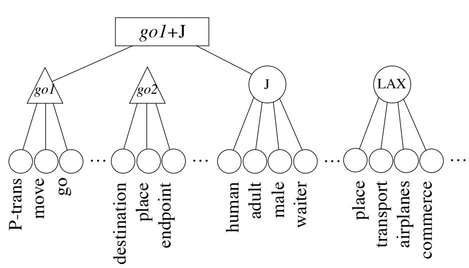
LISAese

Localist role-filler binding units (aka *sub-propositions* or *SPs*)



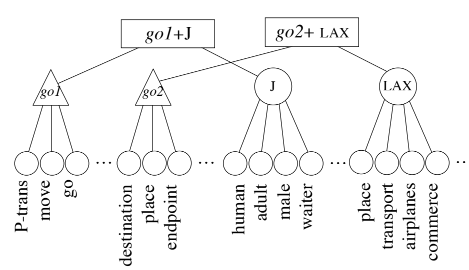
LISAese

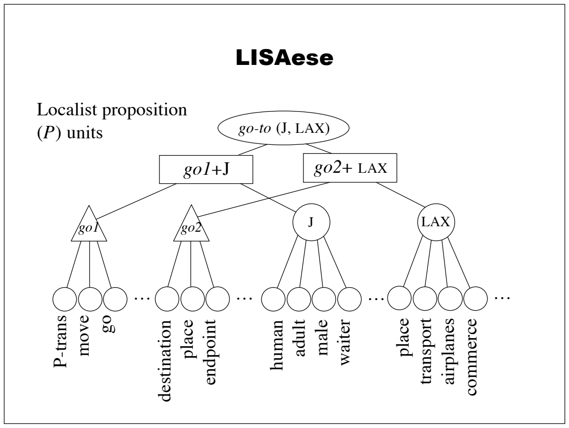
Localist role-filler binding units (aka *sub-propositions* or *SPs*)

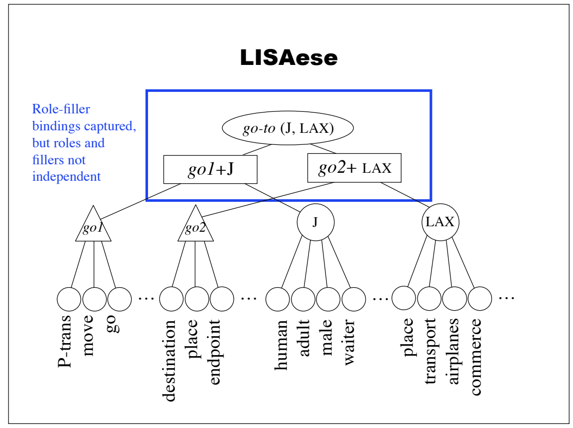


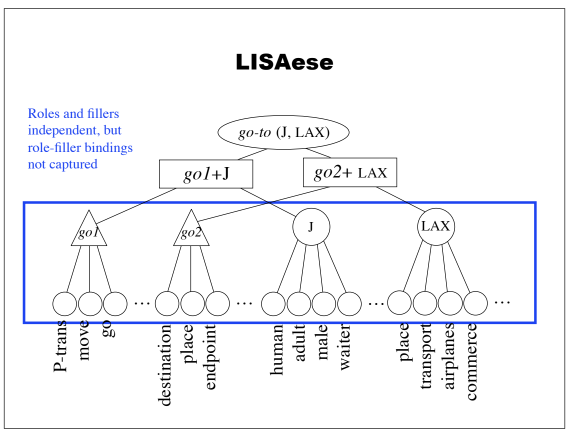
LISAese

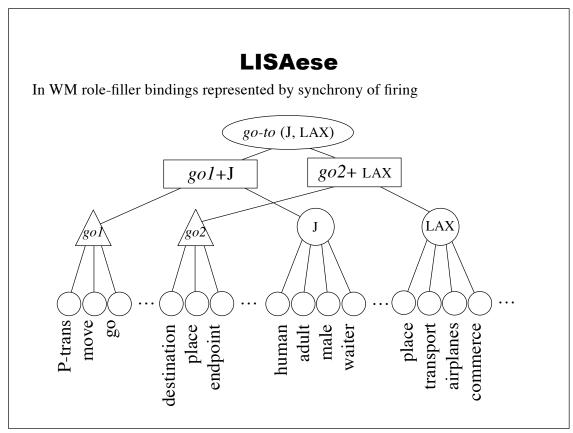
Localist role-filler binding units (aka *sub-propositions* or *SPs*)

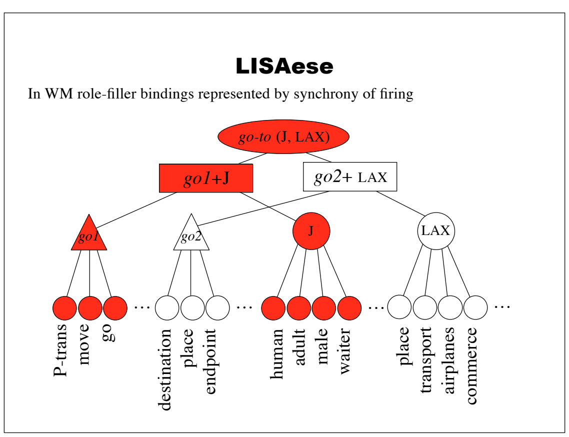


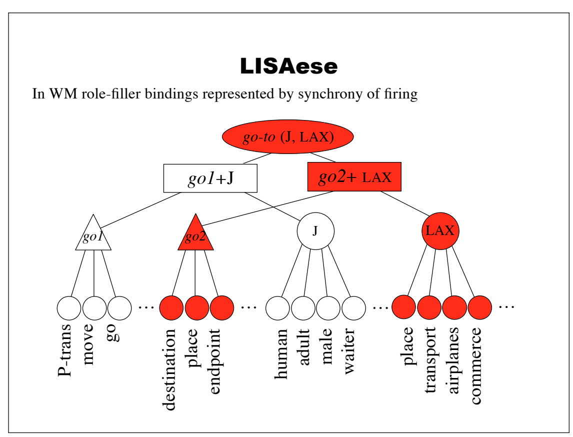

LISAese
Localist proposition (*P*) units
go-to (J, LAX)
go1+J
go2+ LAX
go1
go2
J
LAX
P-trans
move
go
destination
place
endpoint
human
adult
male
waiter
place
transport
airplanes
commerce


LISAese
Role-filler bindings captured, but roles and fillers not independent
go-to (J, LAX)
go1+J
go2+ LAX
go1
go2
J
LAX
P-trans
move
go
destination
place
endpoint
human
adult
male
waiter
place
transport
airplanes
commerce


LISAese
Roles and fillers independent, but role-filler bindings not captured
go-to (J, LAX)
go1+J
go2+ LAX
go1
go2
J
LAX
P-trans
move
go
destination
place
endpoint
human
adult
male
waiter
place
transport
airplanes
commerce

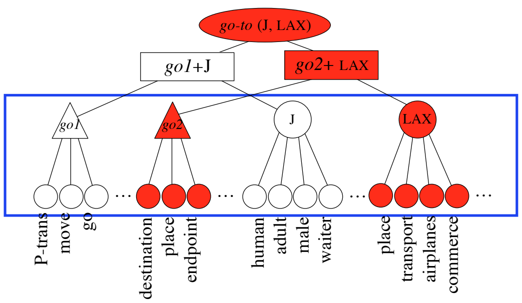

LISAese
In WM role-filler bindings represented by synchrony of firing
go-to (J, LAX)
go1+J
go2+ LAX
go1
go2
J
LAX
P-trans
move
go
destination
place
endpoint
human
adult
male
waiter
place
transport
airplanes
commerce


LISAese
In WM role-filler bindings represented by synchrony of firing
go-to (J, LAX)
go1+J
go2+ LAX
go1
go2
J
LAX
P-trans
move
go
destination
place
endpoint
human
adult
male
waiter
place
transport
airplanes
commerce

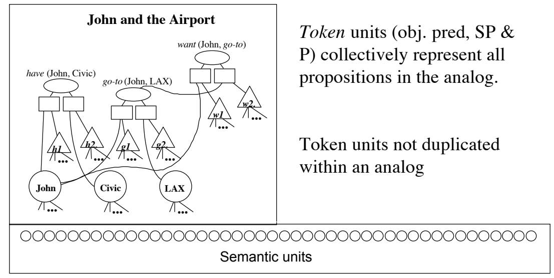

LISAese
In WM role-filler bindings represented by synchrony of firing
go-to (J, LAX)
go1+J
go2+ LAX
go1
go2
J
LAX
P-trans
move
go
destination
place
endpoint
human
adult
male
waiter
place
transport
airplanes
commerce

LISAese

In WM roles and fillers are independent and bindings are explicit



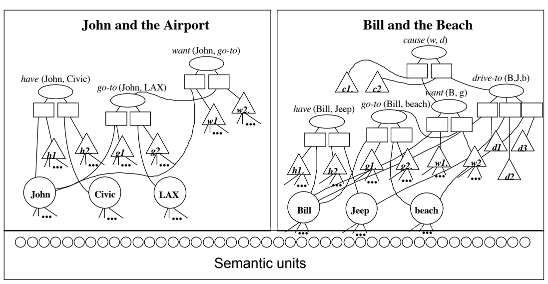
Complete Analog



Token units (obj. pred, SP & P) collectively represent all propositions in the analog.

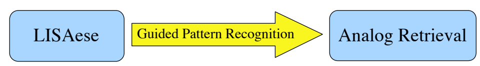
Token units not duplicated within an analog

Multiple Analogs



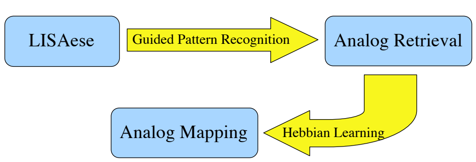
No shared token units across analogs; all semantic units are shared.

Operations on LISAese



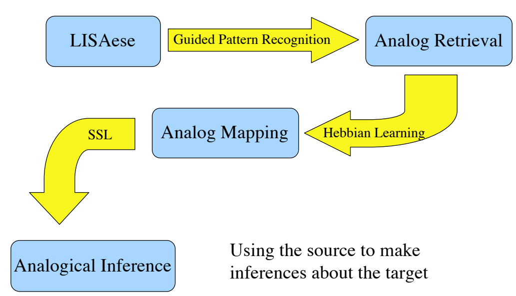
Retrieving a *source* analog or schema from LTM given a novel *target* as a cue

Operations on LISAese



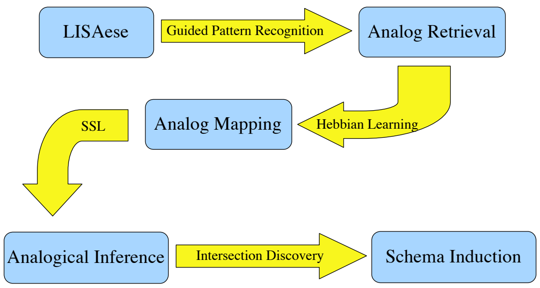
Finding the relational correspondences between source and target

Operations on LISAese



Using the source to make inferences about the target

Operations on LISAese



Using the source and target together to induce a more general schema or rule

Measure of mapping quality in LISA

Proportion of "clear" mappings weighted by "importance"

$$q(T,S) = \frac{\sum_{t,s \in T,S} i_t [m(t, s_{\max}) - m(t, s_{\max 2})]}{1 + \sum_{t \in T} i_t},$$

What Makes an Analogical Inference Plausible?

- Source analog well understood
- More similarities, fewer differences
- Similarities causally relevant to inference
- Multiple source analogs

Bad Analogies

(Style Invitational Report,
Washington Post, July 23, 1995)

The little boat drifted across the pond exactly the way a bowling ball wouldn't.

John and Mary had never met. They were like two hummingbirds who had also never met.

I felt a nameless dread.... It's a dread that nobody knows the name for, like those little square plastic gizmos that close your bread bags. I don't know the name for those either.

Questionable analogies

"Agreed: The national interest requires that all children be educated and that all taxpayers contribute. But it doesn't follow that we need public schools. We need military aircraft; all taxpayers help pay for them. Which doesn't mean that we need public aircraft companies. Schools aren't the same as airplane factories, but the analogy is illuminating."
David Gelernter, Professor of Computer Science, Yale (and Unabomber victim), LA Times, May 2005

Questionable analogies

There's a big problem with the Endangered Species Act: only 10 species have recovered enough to be removed from the list. The act is "a failed managed care program that checks species in but never checks them out."
Congressman Richard Pombo (R-Calif), Southern Sierran, April 2006

Questionable analogies

US President George Bush compared the war in Iraq with the US war for independence in his 4th of July speech. Like the revolutionaries who "dropped their pitchforks and picked up their muskets to fight for liberty", Mr. Bush said American soldiers were fighting "a new and unprecedented war" to protect US freedom.

LA Times, 7/6/2007

Lightning as electricity

Nov. 7, 1749. Electrical fluid agrees with lightning in these particulars: 1. Giving light. 2. Color of the light. 3. Crooked direction. 4. Swift motion. 5. Being conducted by metals. 6. Crack or noise in exploding. 7. Subsisting in water or ice. 8. Rending bodies it passes through. 9. Destroying animals. 10. Melting metals. 11. Firing inflammable substances. 12. Sulphureous smell. -- The electric fluid is attracted by points. -- We do not know whether this property is in lightning. -- But since they agree in all the particulars wherein we can already compare them, is it not probable they agree likewise in this? Let the experiment be made.
Journal of Ben Franklin

Discovery of Demerol

- Synthetic compound, structure similar to morphine
- Induced S-shaped tail curvature in mice
- Effect previously observed only with morphine (but causal mechanism unknown)
- INFERENCE: Demerol would have narcotic effects

Has there ever been life on Mars?

- Only one analog (earth)
- Negative analogs (moon)
- Origin of life on earth not well understood
- Water once flowed on Mars
- Microbes thrive in Antarctica
- Atmosphere once present on Mars (but for relatively short time)
- Some theories (role of tidal pools) fail for Mars (no large moon)

Analogy in Ethnography: Infer Function of Artifacts

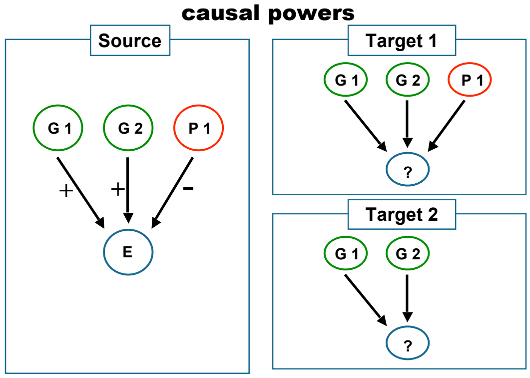
- **Target**: Neolithic Greek clay fragments, individual female legs manufactured as pairs but broken apart
- **Sources**: other paired tokens used to seal a contract and provide special evidence of the identity of the bearer
  - Greece, Rome, Japan, China
  - American mafiosi (tear a monetary bill in half)

Causal Models in Analogical Inference

Lee & Holyoak, 2007 Cog Science meeting

- Pit degree of relational match against causal powers
- Suppose in the source, the effect was produced despite negative factors (i.e., preventive cause).
- Then ***absence*** of a correspondence in target for a ***preventive*** cause might actually ***strengthen*** argument from analogy

Pitting shared relations against causal powers



Example materials

Animal A has blocked oil glands, elevated blood sugar, an extra chromosome, and dry flaky skin.

For **animal A**, blocked oil glands tend to PRODUCE dry flaky skin;
elevated blood sugar tends to PRODUCE dry flaky skin;
an extra chromosome tends to PREVENT dry flaky skin.

Animal B has blocked oil glands and elevated blood sugar.

Animal B also has dry flaky skin.

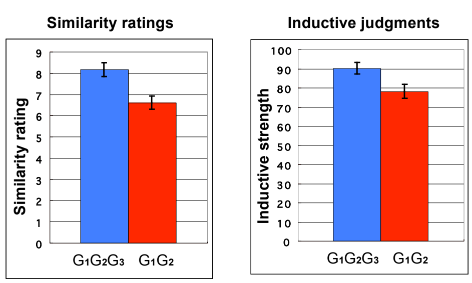
How likely will the conclusion be true? Frequency (0-100):

Judge how similar animal A and animal B are.

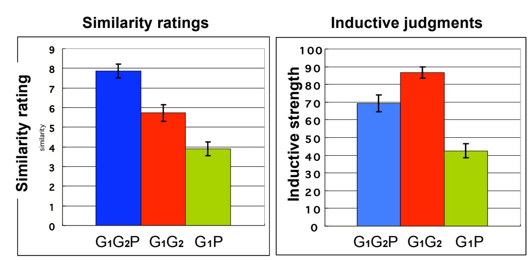
0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

totally different identical

Generative causes



Preventive causes



Issues for Bayesian Inference

1. How to search long-term memory for optimal S_k ?

Define access function

2. How to measure the degree of mapping between S_k and T ?

Define mapping function

3. How to infer the plausibility of analogical inferences?

Assign initial probabilities to inferences potentially generated by CWSG.

4. How to integrate probabilities based on analogy with direct data about T?

Standard Bayesian updating (?)

5. How to infer the plausibility of analogical generalizations ?

Assign initial probabilities to generalized inferences