Summary

- Structured representations are important
  - Abstract
  - Recursive
  - Generative
- New primitive concepts can be learned
  - Learning the most parsimonious theory
- How to combine structured representations and statistical inference?
  - Statistical parsing in language
  - Statistical grammar induction
  - Probabilistic inferences about kin relations.
  - Statistical learning of relational concepts and theories.

Outline for today

- The debate about structure in people's mental representations of concepts
  - Hierarchies or hidden units?
  - Logical relations or hidden units?
  - Definitions or prototypes?
- Probabilistic inference

Semantic networks
(Quillian, 1968)

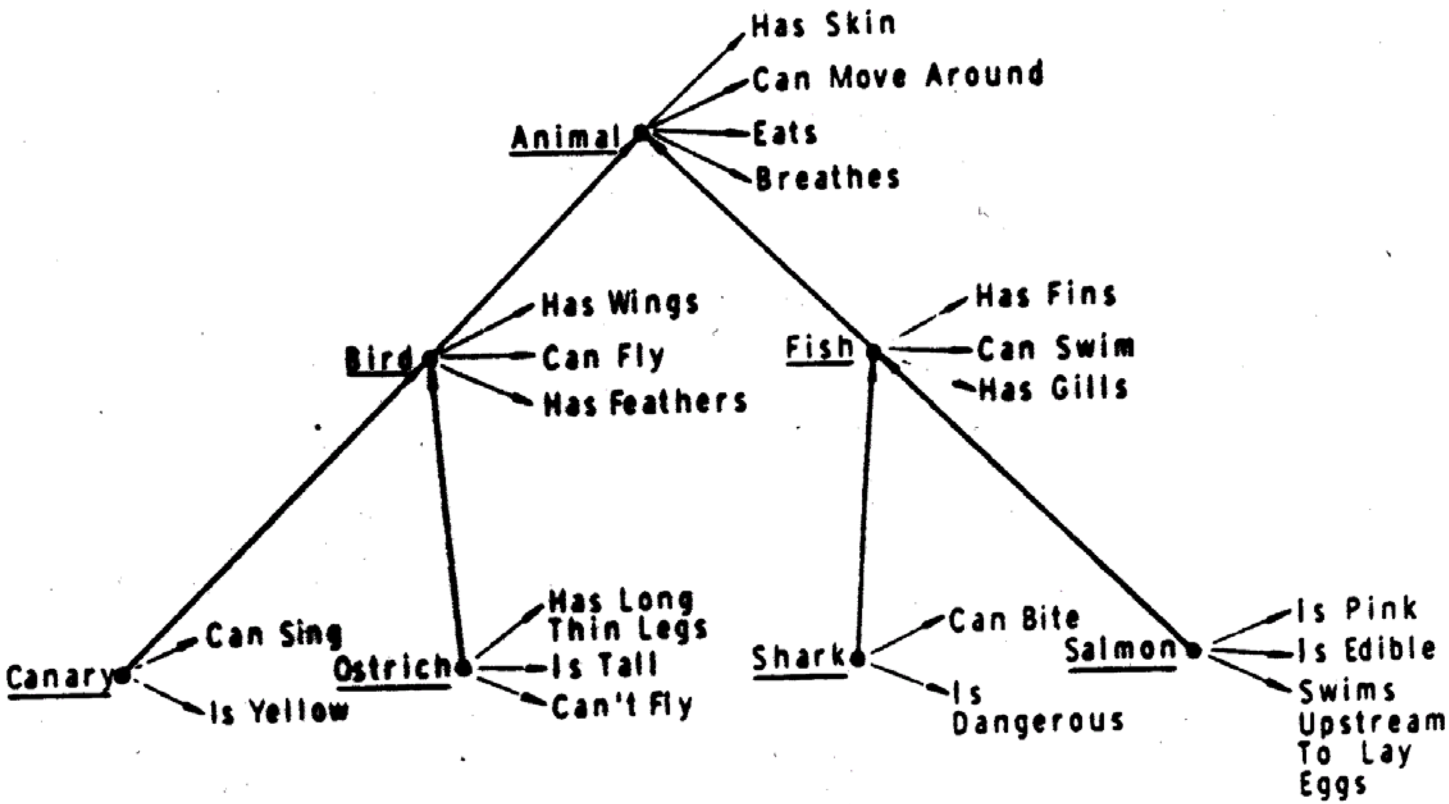


Figure of semantic trees from Quillian (1968). Quillian, M. R. "Semantic Memory." In Semantic *Information Processing*. Edited by M. Minsky. Cambridge, MA: MIT Press,1968, pp. 216-270. Courtesy of the MIT Press. Used with permission.

Why semantic networks?

- Economical encoding of information.

  (a big deal in 1968.)

- Supports generalization.
  - If you learn that a draxel is a bird, you can expect that a draxel has wings, can fly, and has feathers.

Generalization in a semantic network

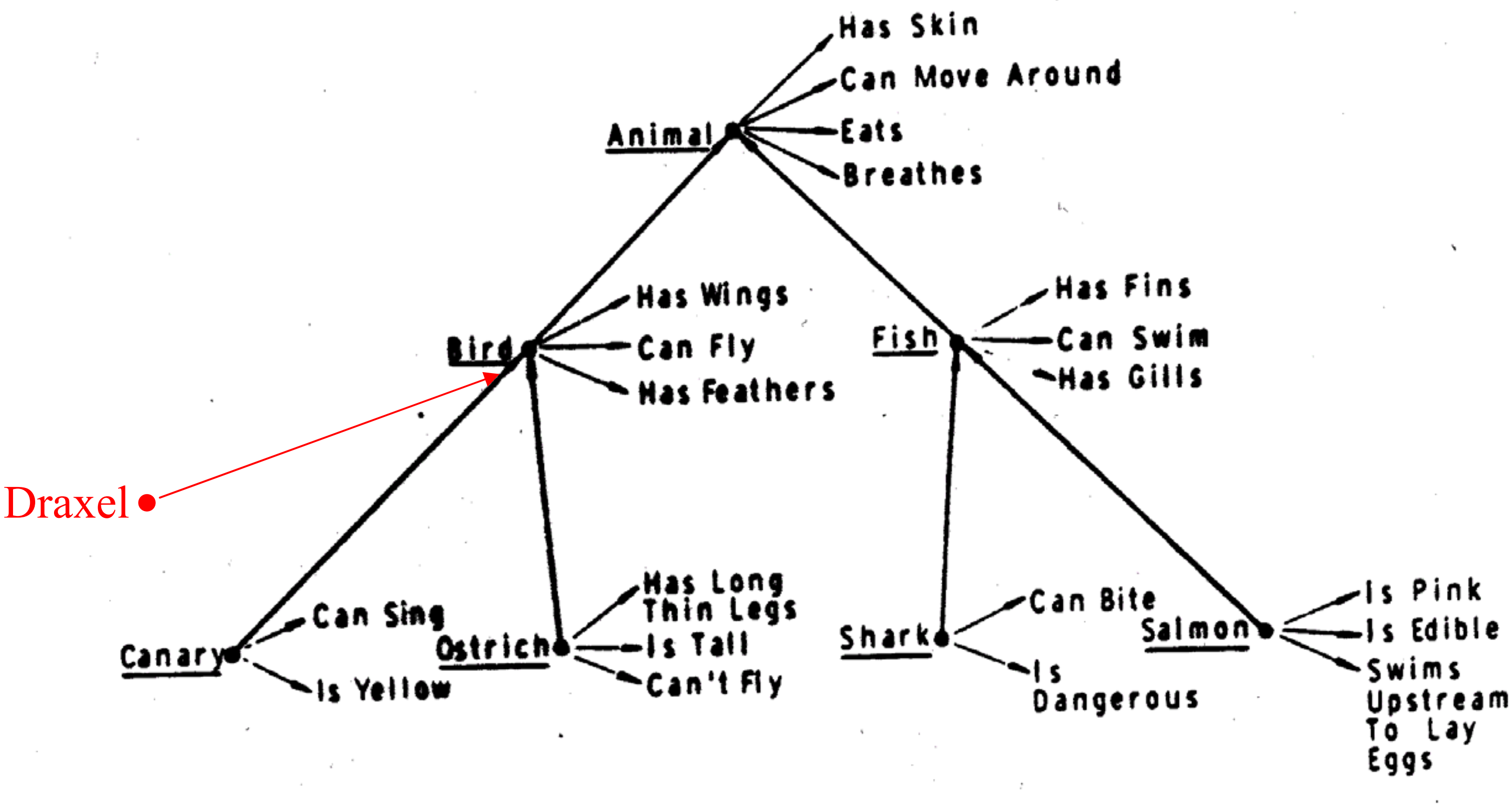


Figure of semantic trees from Quillian (1968). Quillian, M. R. "Semantic Memory." In Semantic *Information Processing*. Edited by M. Minsky. Cambridge, MA: MIT Press,1968, pp. 216-270.

Inferring mental structure through reaction times (Collins & Quillian, 1969)

Image removed due to copyright considerations. Please see:

Collins, A. M., and M. R. Quillian. "Retrieval Time from Semantic Memory." *Journal of Verbal Learning and Verbal Behavior* 8 (1969): 240-248.

General finding: the more of the hierarchy a relation spans, the longer it takes to verify.

Reaction time data

Image removed due to copyright considerations. Please see:

Collins, A. M., and M. R. Quillian. "Retrieval Time from Semantic Memory." *Journal of Verbal Learning and Verbal Behavior* 8 (1969): 240-248.

“Cleaned up” reaction time data

Image removed due to copyright considerations. Please see:

Collins, A. M., and M. R. Quillian. "Retrieval Time from Semantic Memory." *Journal of Verbal Learning and Verbal Behavior* 8 (1969): 240-248.

Reaction time data

Image removed due to copyright considerations. Please see:

Collins, A. M., and M. R. Quillian. "Retrieval Time from Semantic Memory." *Journal of Verbal Learning and Verbal Behavior* 8 (1969): 240-248.

Problems

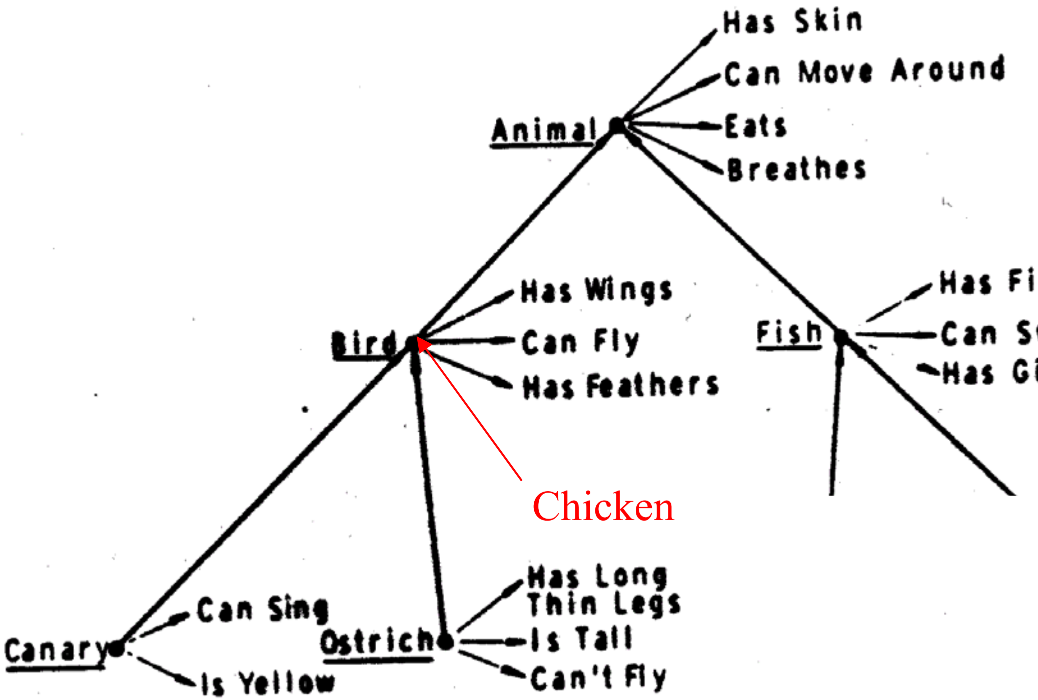
- Typicality effects.
  - "robin is a bird" faster than "chicken is a bird".
- Violations of hierarchy for atypical items.
  - "chicken is an animal" faster than "chicken is a bird."
- Rosch: Graded prototype representations more important than all-or-none "is a" relations.

Problems

- Typicality effects.
  - “robin is a bird” faster than “chicken is a bird”.
- Violations of hierarchy for atypical items.
  - “chicken is an animal” faster than “chicken is a bird.”
- But do these problems require us to give up on “is a” hierarchies?

Possible solutions

- We have multiple trees, a default and some alternative hypotheses.
  - In default tree: chicken falls under bird.
  - In alternative tree: chicken falls under animal.



Default

Alternative

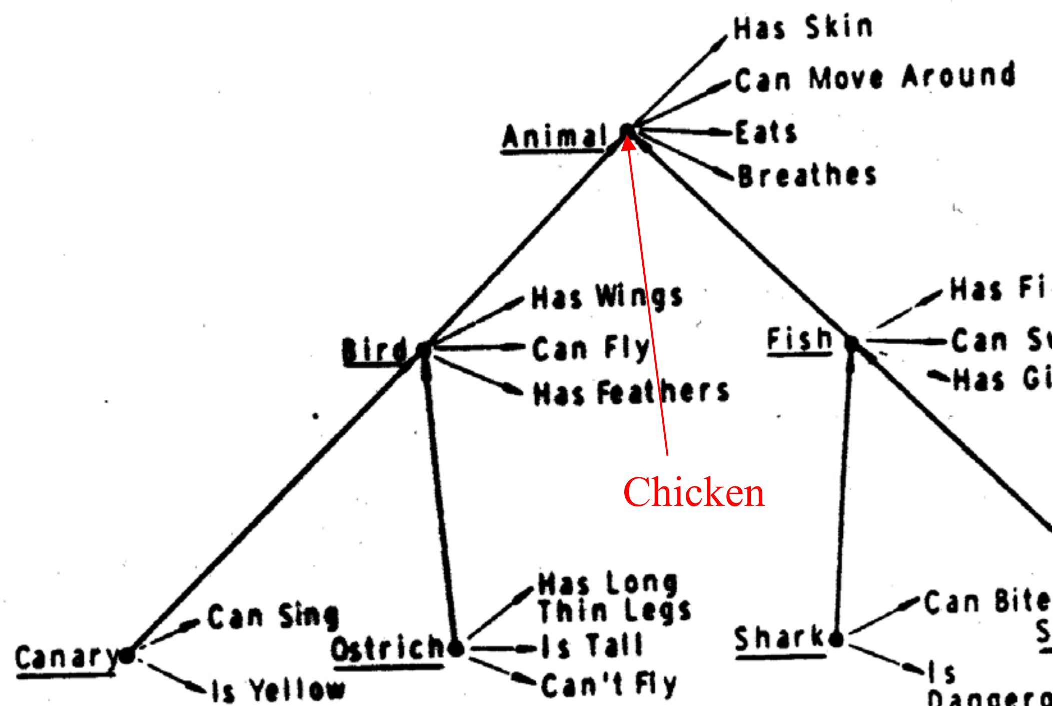


Figure of semantic trees from Quillian (1968). Quillian, M. R. "Semantic Memory." In Semantic *Information Processing*. Edited by M. Minsky. Cambridge, MA: MIT Press,1968, pp. 216-270. Courtesy of the MIT Press. Used with permission.

Possible solutions

- We have multiple trees, a default and some alternative hypotheses.
  - In default tree: chicken falls under bird.
  - In alternative tree: chicken falls under animal.
- The word “bird” maps onto two nodes, one referring just to typical birds and the other to *all* birds.

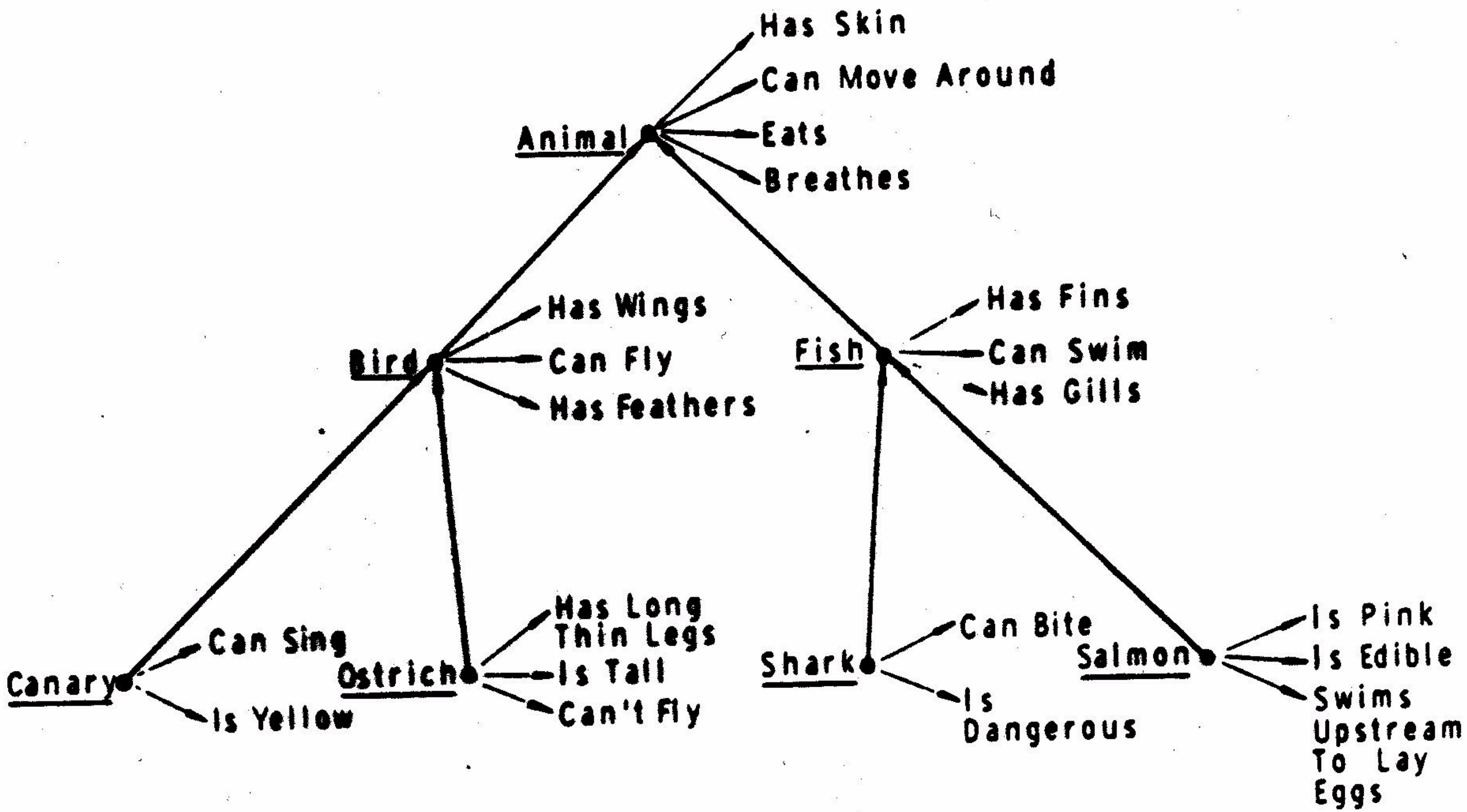


Figure of semantic trees from Quillian (1968). Quillian, M. R. "Semantic Memory." In Semantic *Information Processing*. Edited by M. Minsky. Cambridge, MA: MIT Press,1968, pp. 216-270. Courtesy of the MIT Press. Used with permission.

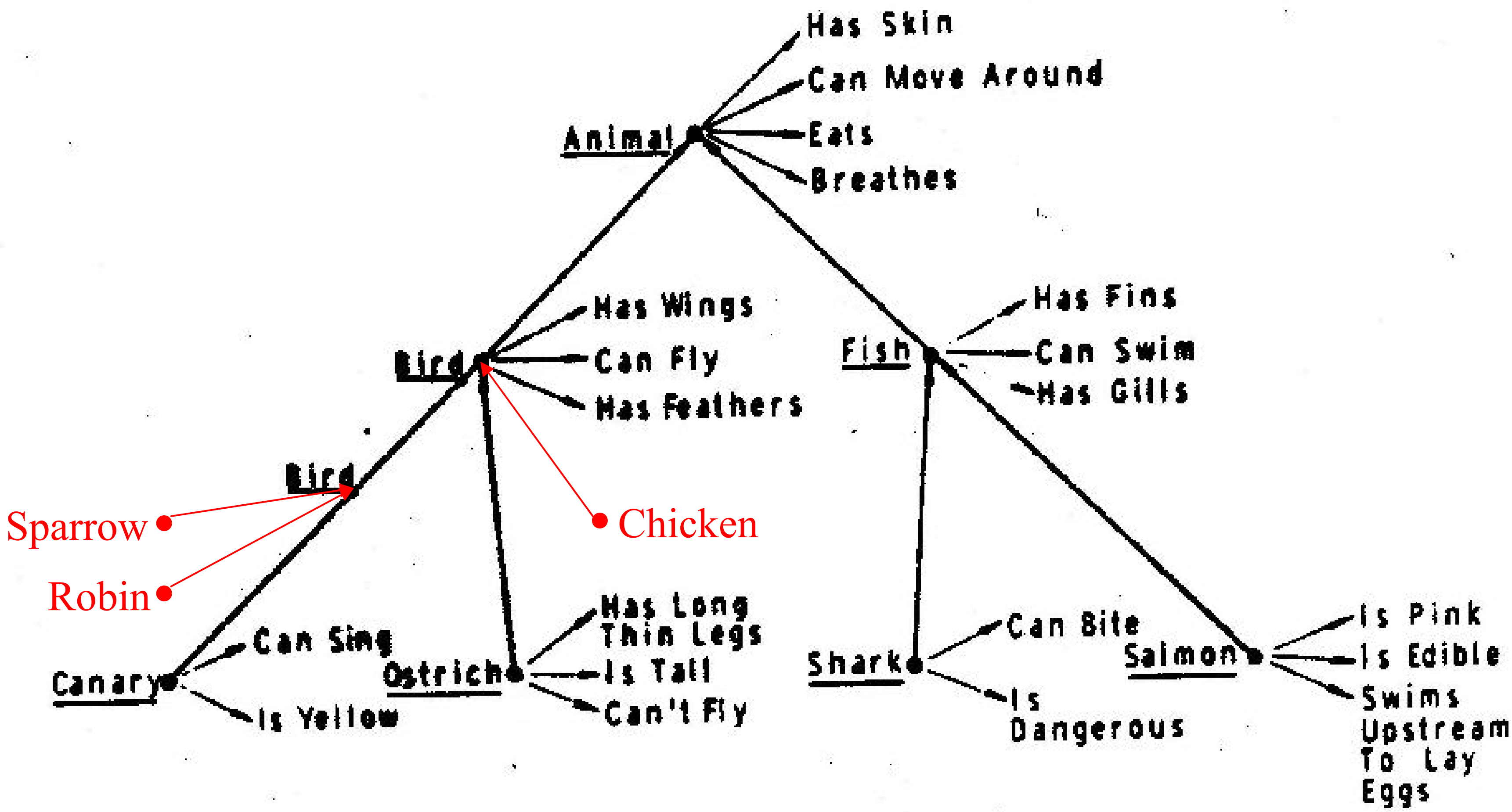


Figure of semantic trees from Quillian (1968). Quillian, M. R. "Semantic Memory." In Semantic *Information Processing*. Edited by M. Minsky. Cambridge, MA: MIT Press,1968, pp. 216-270.

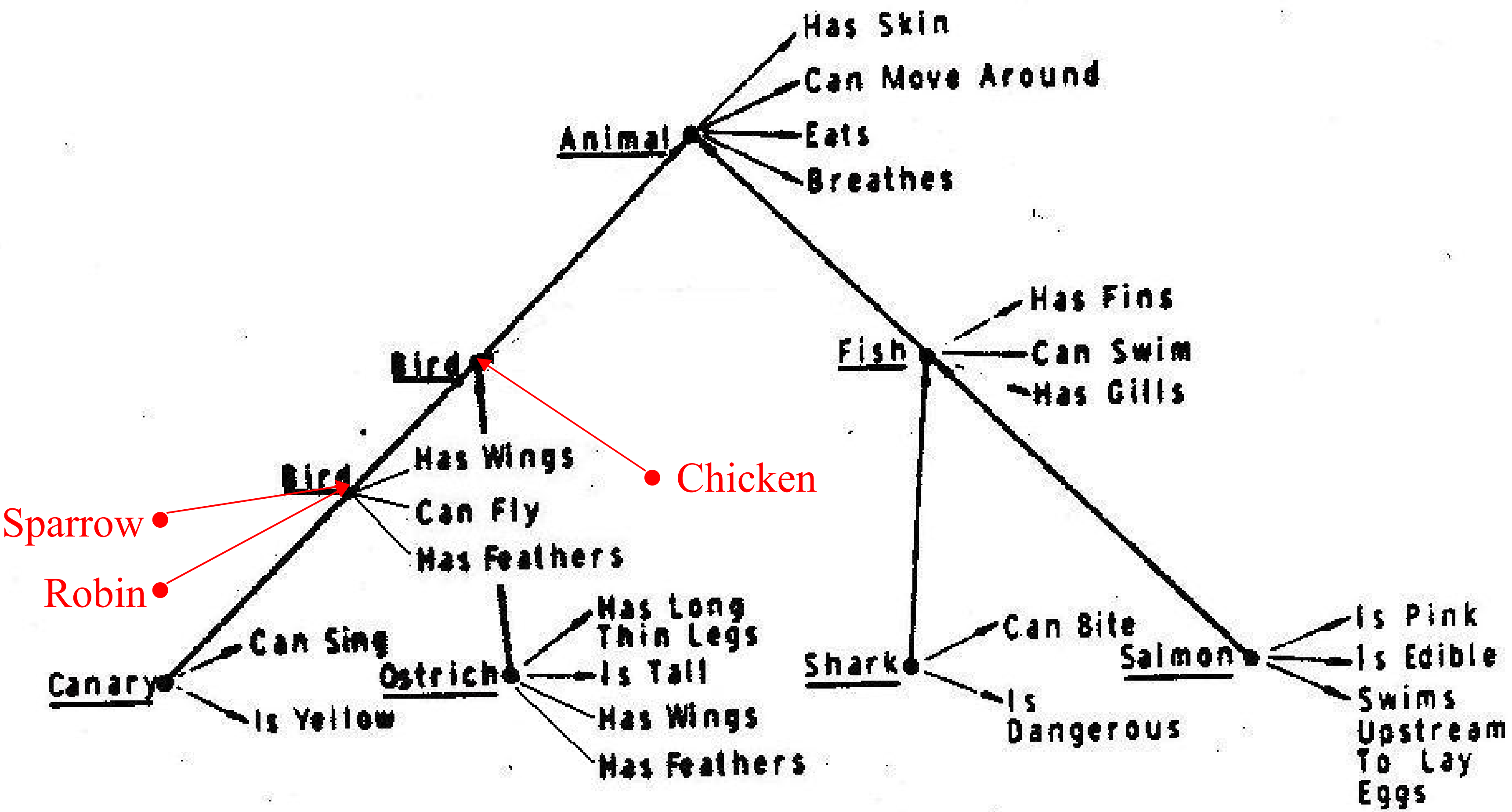


Figure of semantic trees from Quillian (1968). Quillian, M. R. "Semantic Memory." In Semantic *Information Processing*. Edited by M. Minsky. Cambridge, MA: MIT Press,1968, pp. 216-270.

Possible solutions

- We have multiple trees, a default and some alternative hypotheses.
  - In default tree: chicken falls under bird.
  - In alternative tree: chicken falls under animal.
- The word “bird” maps onto two nodes, one referring just to typical birds and the other to *all* birds.
- Deny that prototype effects are diagnostic of core representations.

Armstrong, Gleitman & Gleitman

Image removed due to copyright considerations. Please see:

Armstrong, S. L., L. R. Gleitman, and H. Gleitman. “What Some Concepts might not be.” *Cognition* 13, no. 3 (May 1983): 263-308.

Armstrong, Gleitman & Gleitman

Image removed due to copyright considerations. Please see:

Armstrong, S. L., L. R. Gleitman, and H. Gleitman. “What Some Concepts might not be.” *Cognition* 13, no. 3 (May 1983): 263-308.

- Prototype ratings and reaction time effects for clearly definitional concepts shows that these data are not diagnostic of conceptual structure.

Why not give up on a definitional core for concepts?

- Reasoning, e.g., “Consider a new person, Boris.”
  - Is the mother of Boris’s father his grandmother?
  - Is the mother of Boris’s sister his mother?
  - Is Boris’s uncle his grandfather?
  - Is the son of Boris’s sister his son?

- Compositionality in concepts and language
  - e.g., Greatgrandmother = mother of a grandparent.
  - “Colorless green idea”
  - “Big”

Why not give up on a definitional core for concepts?

- Even without definitions, need a distinction between typicality and degree of membership.
  - At some level we know for certain that chickens are birds. (Consider a bet....)
  - Some categories really are graded in their membership:

green or blue?

Images removed due to copyright considerations.

cup or bowl?

Other problems for all-or-none semantic relations

- Graded generalization
  - Which is a stronger inference?

Canaries have sesamoid bones.	Chickens have sesamoid bones.
All birds have sesamoid bones.	All birds have sesamoid bones.

  - More of a problem, as generalization is the main function that “is a” hierarchies are supposed to fulfill.
- Others?

An alternative architecture

- Semantic networks are symbolic:
  - encode discrete, localized bits of knowledge.
- Neural networks are subsymbolic:
  - inspired by long-term memory in the brain (synaptic plasticity).
  - graded representations that can approximate symbolic models, e.g., "is a" hierarchies, while still capturing prototypicality.

Image removed due to copyright considerations.

Training set

Image removed due to copyright considerations.

Learned distributed representation

Image removed due to copyright considerations.

Generalization test

- Train on one fact for new object:
  - Draxel ISA bird
- Network then believes that a Draxel has other properties in common to most other birds . . .
  - can fly, has wings, has feathers.
- . . . but not properties distinctive to individual birds (e.g., is red or is yellow).

Hierarchical structure in conceptual development

Image removed due to copyright considerations. Please see:

Keil. "The Development of the Young Child's Ability to Anticipate the Outcomes of Simple Causal Events." *Child Development* 50 (1979): 455-462.

Image removed due to copyright considerations. Please see:

Keil. "The Development of the Young Child's Ability to Anticipate the Outcomes of Simple Causal Events." *Child Development* 50 (1979): 455-462.

Development of hierarchy in network

Image removed due to copyright considerations.

Problems

- Collapses typicality and graded membership.
  - *Chicken* activates the *Bird* unit less than *Canary* does.
  - But recall:
    - At some level we know for certain that chickens are birds.
    - Some categories really are graded in their membership:

Image removed due to copyright considerations.

cup or bowl?

Problems

- Requires special care in training.
  - Lots of training data, must be randomly interleaved throughout training.
  - Potential for “catastrophic interference” when learning a new fact, without freezing weights.
  - As knowledge base grows, need to add hidden units (to preserve bottleneck ratio for good generalization).
  - When learning a novel proposition, “blickets may queem”, some controller needs to specify that *blicket* initializes a new input node, *queem* a new output node, and *may* a new relation ndoe.

Problems

- Doesn't know certain obvious things unless explicitly trained:
  - "A bird is a bird": we don't have to check that fact the same way we check "a bird is an animal".
  - "A blicket is a blicket".
  - If these are living things,

Image removed due to copyright considerations.

they are either plants or animals. If animal, they are some kind of animal -- not "just" an animal. (i.e., Must initialize an unlabeled node under *animal*, which is then a candidate word meaning.)

Fodor and Pylyshyn: What’s missing from connectionism?

- Systematicity
  - The thoughts a cognitive system is capable of are not a random collection (like the phrases in a tourist’s foreign language-phrasebook) but a systematic set (like the sentences that can be produced by a fluent speaker of a language).
  - If it can think *Sandy loves Kim*, then it can entertain the thought *Kim loves Sandy*.

Learning family relationships (Hinton, 1986)

- Tree structure generates relations:

Image removed due to copyright considerations. Please see:

Hinton, G. E. "Learning Distributed Representations of Concepts." *Proc. Ann. Conf. of the Cognitive Science Society* 1 (1986).

Learning family relationships
(Hinton, 1986)

- Network architecture:

Image removed due to copyright considerations. Please see:

Hinton, G. E. “Learning Distributed Representations of Concepts.” *Proc. Ann. Conf. of the Cognitive Science Society* 1 (1986).

Learning family relationships
(Hinton, 1986)

- 112 possible facts of the form:

  <person1, relation, person2>

  <Christopher, father-of, Victoria>,

  <Colin, son-of, Victoria>,

  <Jennifer, aunt-of, Colin> . . .

- Trained on 108 examples, network usually generalizes well to the other 4.
  – Doesn’t work with less training.

Learning family relationships
(Hinton, 1986)

- Does this really count as systematicity?
  - With so much training required, and so little generalization ability?
  - Every time you learn about a new person, still need an external controller to add that person to both the input layer and output layer. That's the *real* source of systematicity.

Linear Relational Embedding

Image removed due to copyright considerations.

- Minor improvement, from 4 to 8 or 12 generalization trials.

Learning family relationships
(Hinton, 1986)

- Problem: Consider a new person, Boris.
  - Is the mother of Boris's father his grandmother?
  - Is the mother of Boris's sister his mother?
  - Is Boris's uncle his grandfather?
  - Is the son of Boris's sister his son?

A Big open question

- How to integrate abstract knowledge with probabilistic (or typicality-based) reasoning?
  - Is the son of Boris's sister his son? *(Note: Boris and his family were stranded on a desert island when he was a young boy.)*
  - Is Boris's son his wife's son?
  - Boris has five aunts. How many cousins does he have?

A challenge for either approach

- “Because Sarah loves him, John hates Bill.”
  - Who does “him” refer to?
  - How to represent this thought?

- Two hypotheses:

  cause(loves(Sarah,Bill), hates(John,Bill))

  cause(loves(Sarah,John), hates(John,Bill))

- Why prefer the first?
  - Inference rules:

    implies(and(cause(x,y), cause(y,z)), cause(x,z))

    implies(cause(and(x,y),z), cause(x,z))

  - Beliefs with high probability:

    cause(and(loves(x,y), loves(y,z), not(loves(y,x))), jealous(x,z))

    cause(jealous(x,y), hates(x,y))

  - First hypothesis would be true if:

    loves(John,Sarah)

    not(loves(Sarah,John))

  - No such simple explanation for second hypothesis.

So...

… why do we keep having this debate: rules/symbols vs. prototypes/connections?

Other cases:

- Language acquisition and processing, e.g. past tense
- Schemas and scripts for events and actions
- Visual object recognition and scene perception

So...

… why do we keep having this debate: rules/symbols vs. prototypes/connections?

… and why do none of the standard approaches seem to be satisfying?

So…

The real problem: a spurious contest between logic and probability.

- Neither logic nor probability on its own is sufficient to account for human cognition:
  - Generativity
  - Systematicity
  - Recursion and abstraction
  - Flexibility
  - Effective under great uncertainty (e.g., sparse data)
- What we really need is to understand how logic and probability can work together.

So…

The real problem: a spurious contest between logic and probability.

- A confusion between knowledge representations and inference processes:
  - Gradedness or fuzziness doesn't necessarily mean that the knowledge representations lack structure or rules -- merely that the inference processes incorporate uncertainty.
- Probabilistic inference over structured representations is what we need.