9.66 / 9.914 Computational Cognitive Science

Josh Tenenbaum

What is this class?

- An attempt to see how recent work in computation (AI, machine learning, statistics) can inform some of the core questions of cognitive science.

- … and vice versa.

The questions

- What forms does our knowledge of the world take?
- What are the inductive principles that allow us to acquire new knowledge from the interaction of prior knowledge with observed data?
- What kinds of data must be available to human learners, and what kinds of innate knowledge (if any) must they have?

Goals for the term

- Setting up the big questions.
- Providing some computational tools for answering them rigorously and precisely.
- Exploring some case studies in cognition:
  - Concept learning and categorization.
  - Learning causal relations.
  - The structure and formation of intuitive theories of physical, biological and social systems.
  - The acquisition of natural language (syntax and semantics).
  - Theory of mind: how we understand the behavior and mental states of other people.

Requirements and grading

- Four problem sets. (40%)
  - Minimal programming in Matlab.
  - Experience working with real cognitive data.
  - For students already familiar with the relevant computational techniques, e.g., through 6.825 or 6.867, a more extensive modeling project can be substituted for one or more problem sets.

Requirements and grading

- Term paper or project. (40%)
  - Due on the last day of the term (before finals).
  - Two options:
    - Theoretical paper (15-20 pages).
    - Computational modeling project + short write-up (e.g. 4 page conference format).
  - Undergraduates may implement and extend an existing model.
  - Graduate students must make an original research contribution.

Requirements and grading

- Discussion notes responding to the readings (and questions) for each week. (20%)
  - Short (~1 paragraph) responses to the readings and questions -- not just summaries.
  - Due by 10 am on the day of class.
  - Submit electronically.
  - Must submit 20 notes for full credit. These can include short responses to other people's posts, as long as the responses are thoughtful and in some way address the readings and questions.

Relation to other classes

- 9.52
- 9.012
- Other computational classes in course 9

- 6.034 or 6.825
- 6.867

Background?

- Grad or undergrad?
- Matlab? C or Java?
- Taken graduate AI? Undergraduate AI? Machine learning?
- Who knows about:
  - Bayes' rule
  - Eigenvector
  - Support vector machine
  - Bayes net
  - Markov-equivalent networks
  - First-order logic
  - Probabilistic relational models

The problem of induction

Induction versus Deduction

- Deductive reasoning:

Socrates is a man.
All men are mortal.

Socrates is mortal.

- Inductive reasoning:

Socrates is a man.
Socrates is mortal.

All men are mortal.

Induction versus Deduction

- Deductive reasoning:
  - Conclusion follows with certainty.
  - Validity depends only on syntax (form).

Plato is a snark.
All snarks are boojums.

Plato is a boojum.

  - Argument evaluation is objective, independent of other knowledge available.

Aristotle

- *Prior analytics*: How can we reliably infer new truths from known truths?
- A theory: Symbolic logic.
- The syllogism:

All B are C
A is a B

A is a C

Socrates is a man
All men are mortal

Socrates is a mortal

Plato is a snark
All snarks are boojums

Plato is a boojum

. . .

Induction versus Deduction

- Inductive reasoning:
  - Conclusion follows with more or less probability.
  - Probability depends on semantics (meaning).

Socrates is a man.
Socrates is 47 years old.

All men are 47 years old?

  - Argument evaluation is subjective, depends on other knowledge available.

Aristotle

- *Prior analytics*: How can we reliably infer new truths from known truths?
- A theory: Symbolic logic.
- A schema for induction:

x_1 is P
x_2 is P
⋮
x_n is P
$x_1, x_2, \ldots, x_n$ are all the X's

Every X is P

Aristotle

- *Prior analytics*: How can we reliably infer new truths from known truths?
- A theory: Symbolic logic.
- A schema for induction:

x_1 is P
x_2 is P
⋮
x_n is P
$x_1, x_2, \ldots, x_n$ are all the X's

Every X is P

Horses are long-lived
Men are long-lived
Camels are long-lived
Horses, men, and camels are all the bile-less animals

All bile-less animals are long-lived

The great problem of philosophy

- John Stuart Mill (*A System of Logic*, 1843):

  "Why is a single instance, in some cases, sufficient for a complete induction, while in others myriads of concurring instances, without a single exception known or presumed, go such a very little way towards establishing a general proposition? Whoever can answer this question knows more of the philosophy of logic than the wisest of the ancients, and has solved the problem of Induction."

Computational approaches to induction

Two main ingredients:

- Knowledge representation: how to capture the *structure* of the world
- Statistics: how to *capture* the structure of the world.

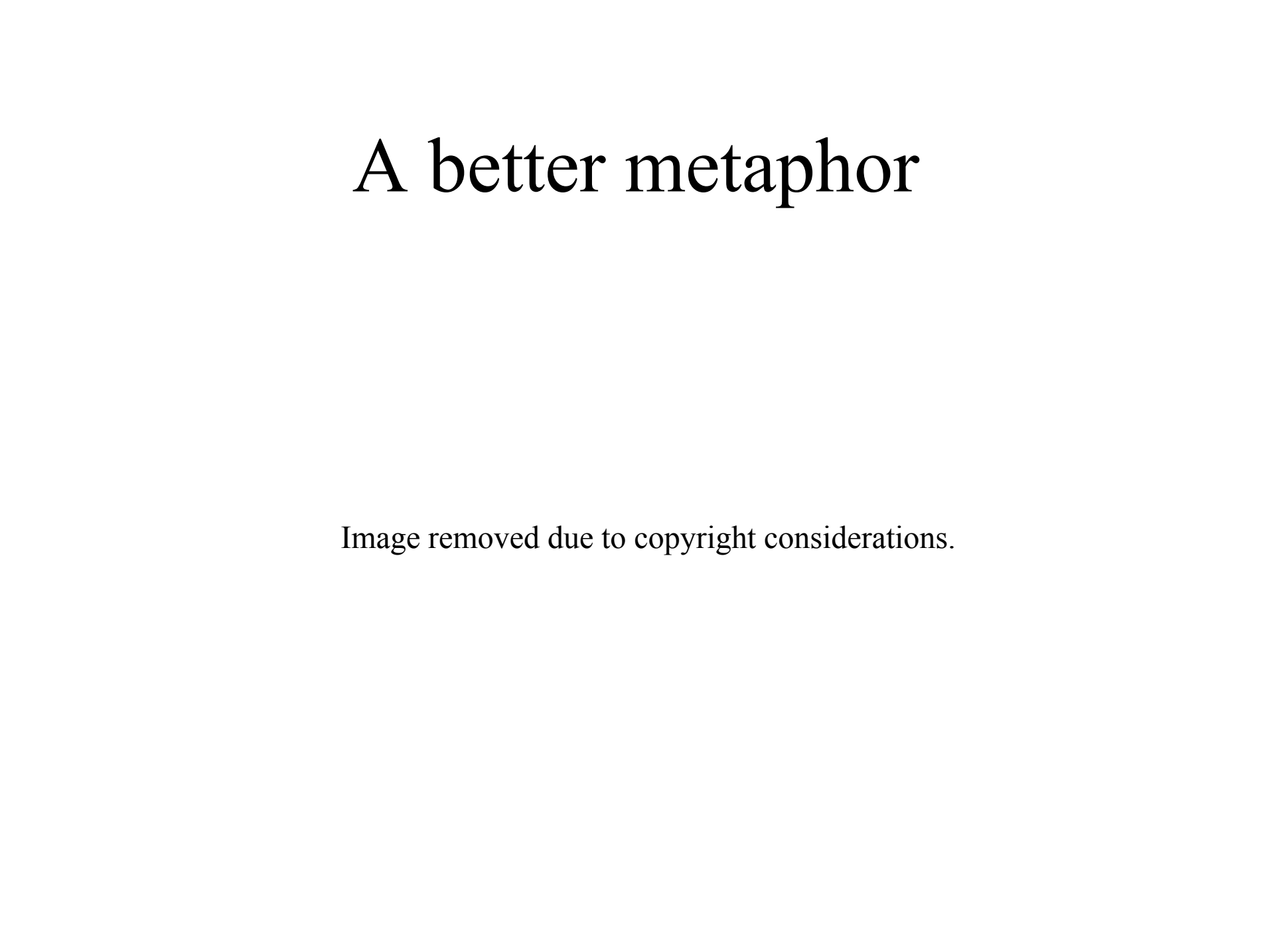
Structure versus statistics

Rules
Logic
Symbols

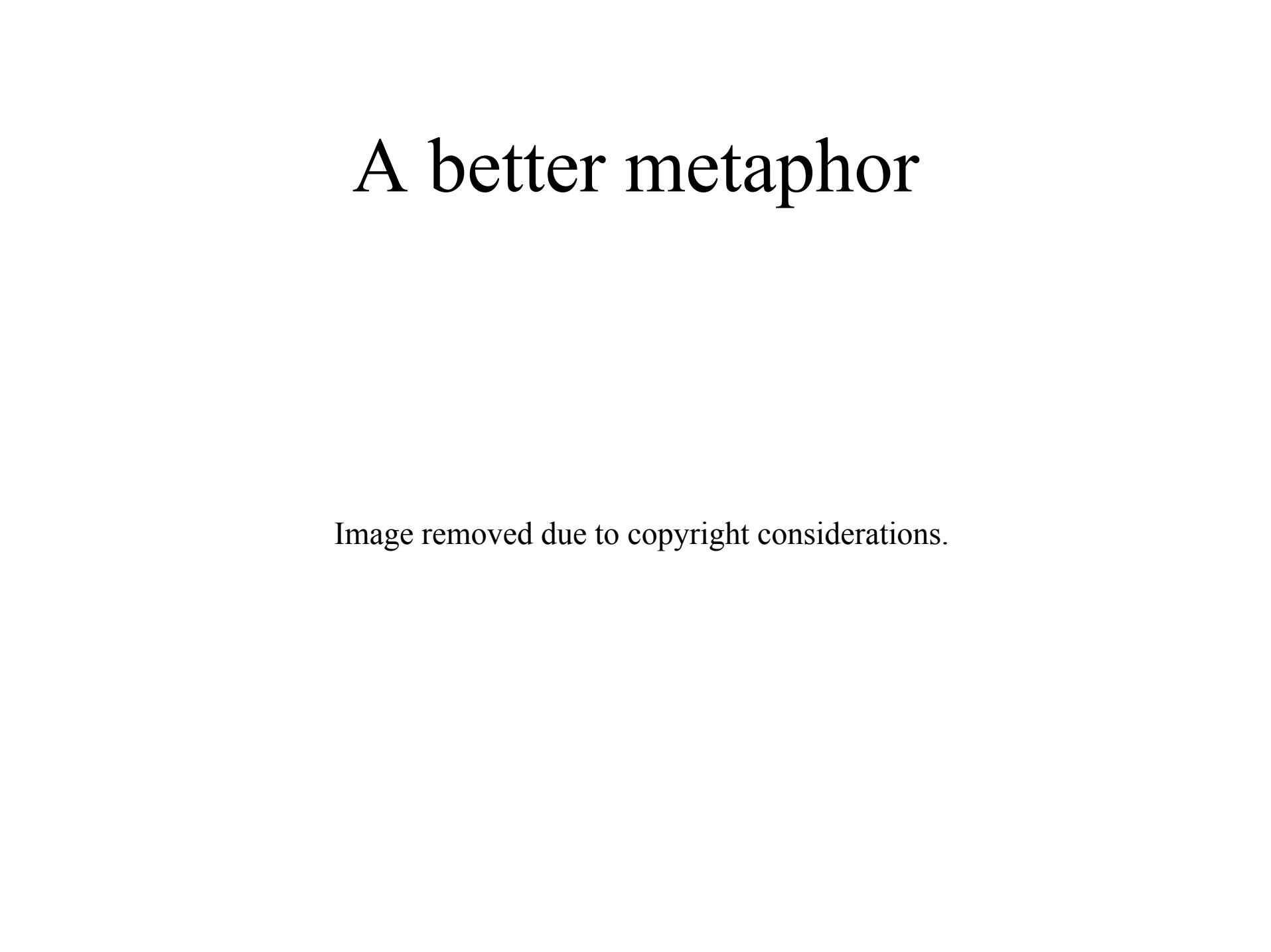
Image removed due to copyright considerations.

Statistics
Similarity
Typicality

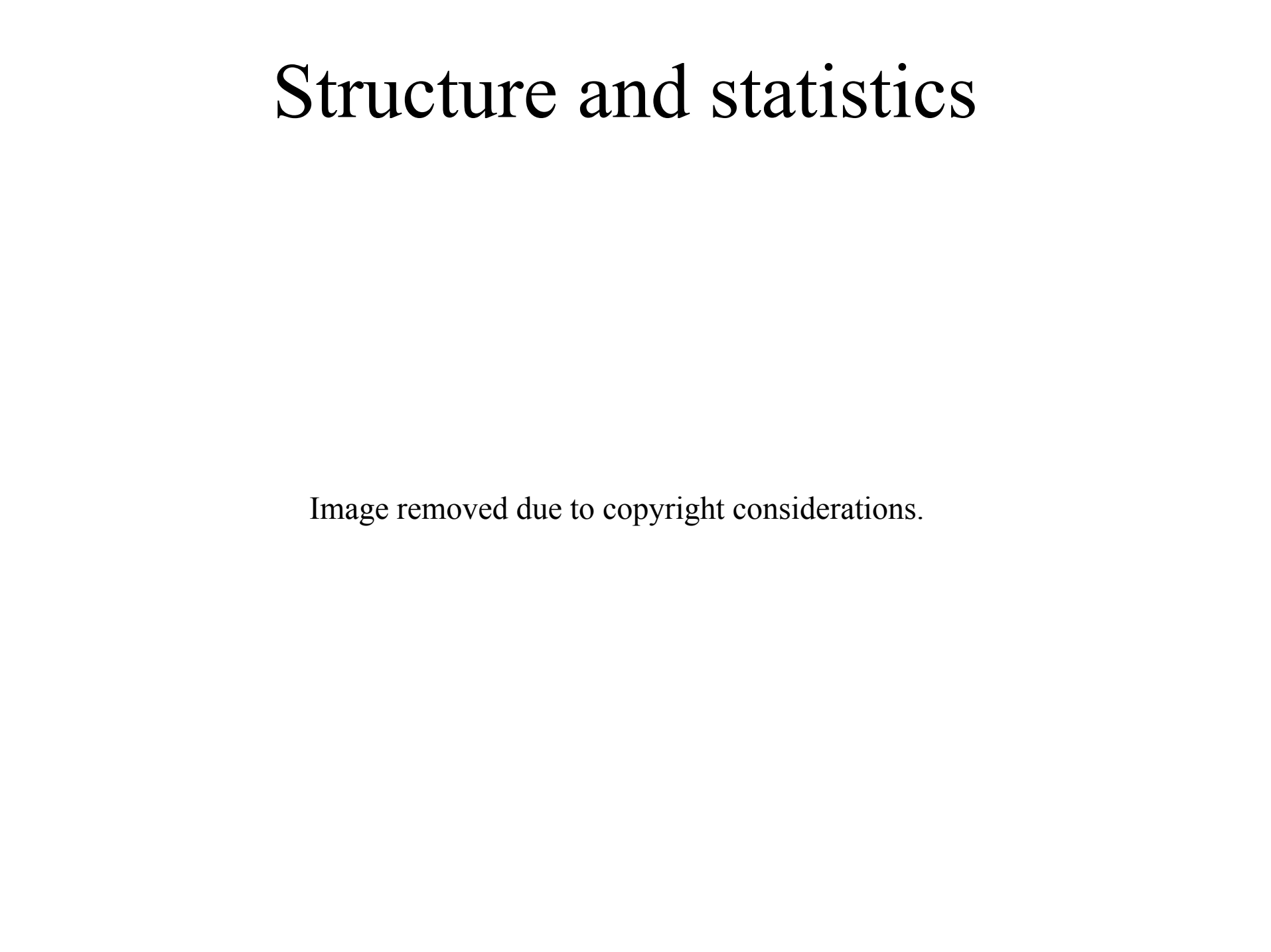
A better metaphor



A better metaphor



Structure and statistics



Cognition as inductive inference

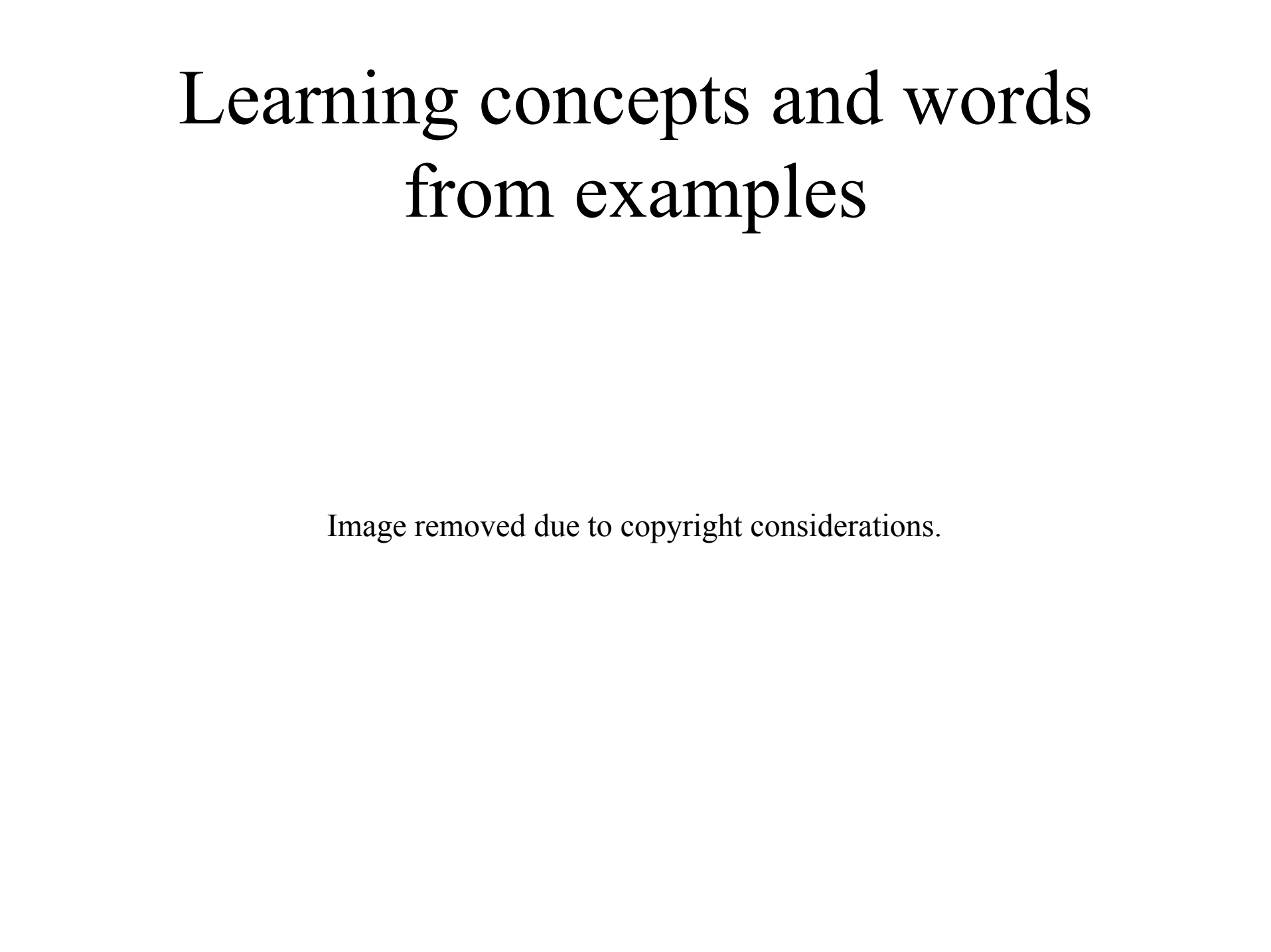
Induction in everyday reasoning

- Generalizing from examples.

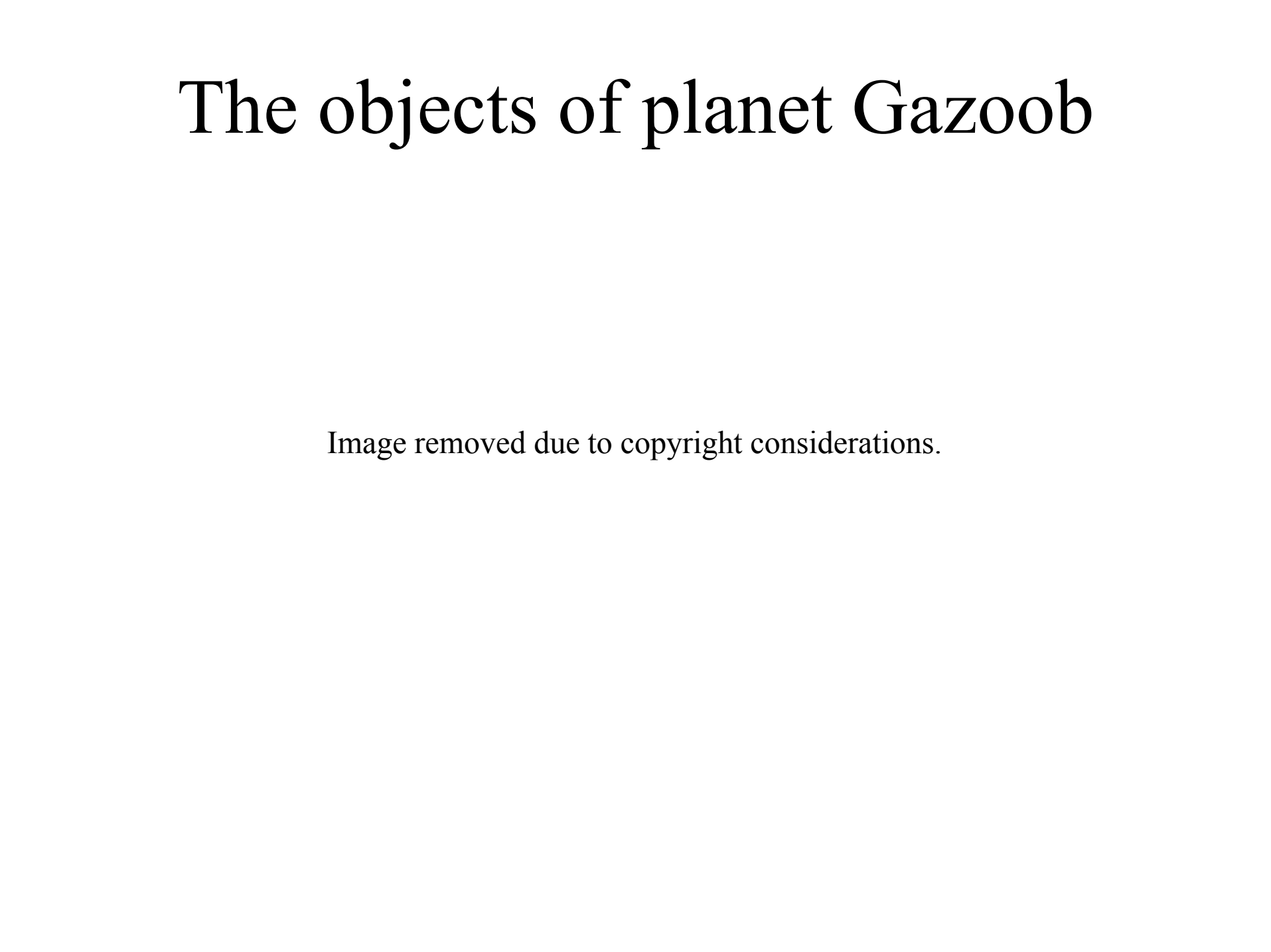
Squirrels can get avian flu.
Gorillas can get avian flu.

Horses can get avian flu.

Learning concepts and words from examples



The objects of planet Gazoob



Induction in everyday reasoning

- Generalizing from examples.
- Diagnosis of causes given effects.
  - A woman at age 40 tests positive on a routine mammogram screening. How likely is it that she has breast cancer?

Inference from novel events

- Is a woman's chance of breast cancer higher or lower if:
  - she tests positive twice in a row?
  - she tests positive first, takes it again and tests negative the second time?
  - she is 45 years old instead of 40?
  - she lives near a nuclear power plant?
  - she has never had a routine screening, but instead chose to get a mammogram because she felt a lump?

Induction in everyday reasoning

- Generalizing from examples.
- Diagnosis of causes given effects.
- Inferring causal relations from patterns of correlation.
  - A drug intended to treat a chronic medical condition is found to improve the condition in 141/250 test patients. Does the drug work?
  - Does prayer improve recovery from illness?

Induction in everyday reasoning

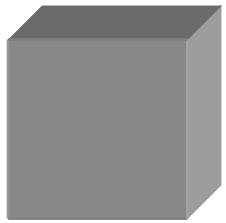
- Generalizing from examples.
- Diagnosis of causes given effects.
- Inferring causal relations from patterns of correlation.
- Discovering hidden causes from patterns of coincidence.
  - Is a cluster of disease cases just due to chance, or to a previously unknown cause?
  - Guillon-Barre, AIDS, Lyme disease.

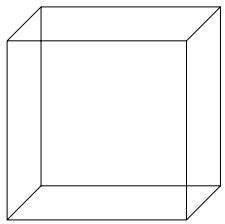
Visual perception

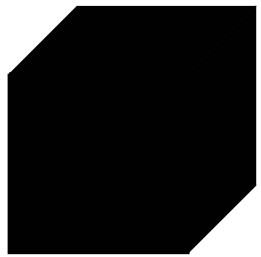
Image removed due to copyright considerations.

Ambiguity in visual perception

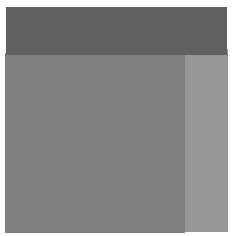
Three-dimensional:

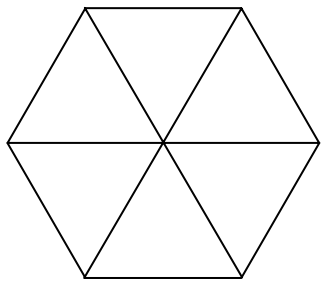


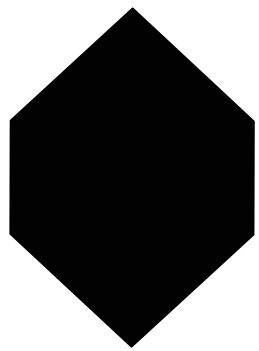




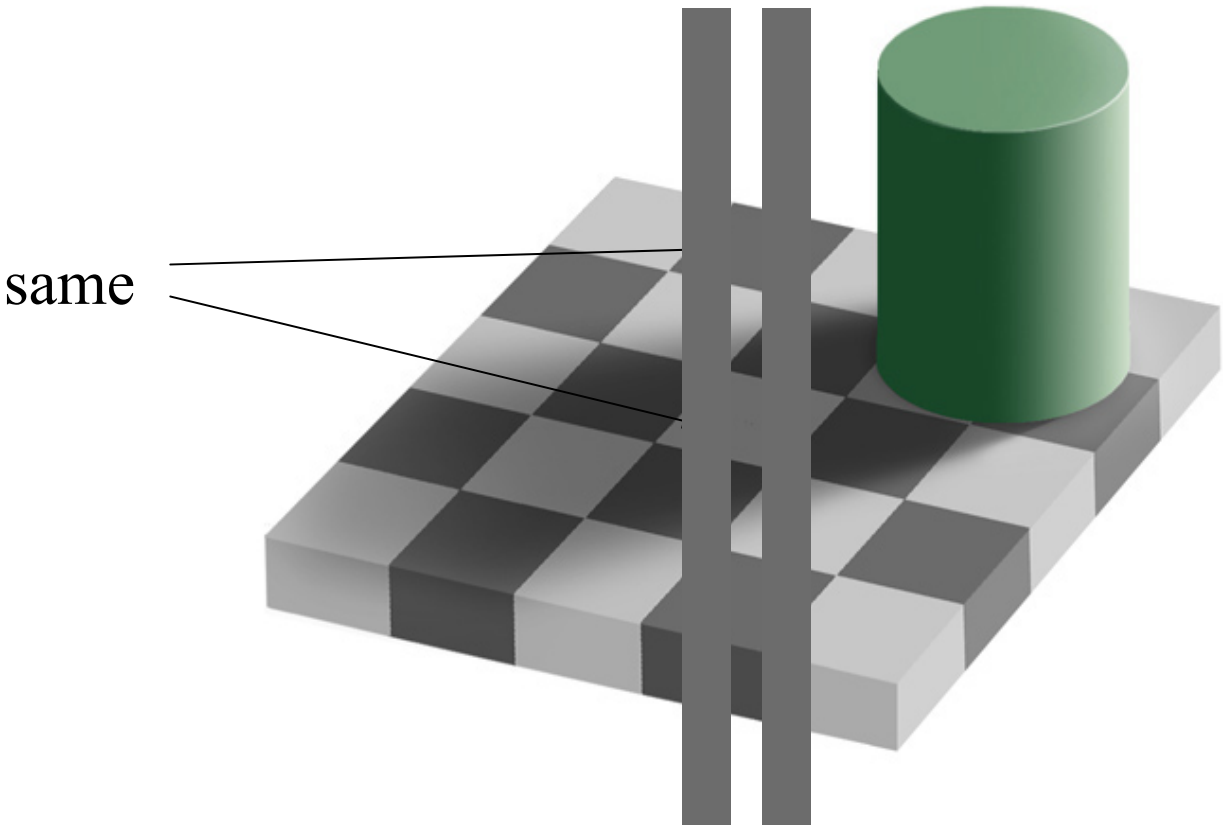
Two-dimensional:







The checkerboard illusion



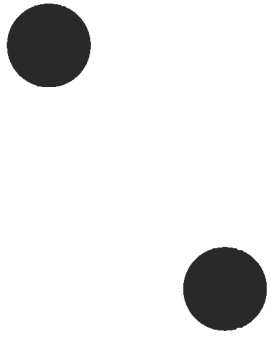
Courtesy of Edward Adelson. Used with permission.

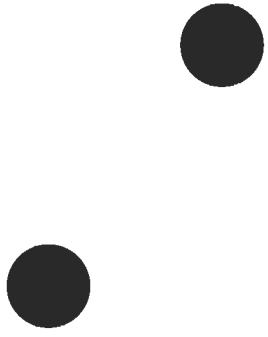
Inference to the casually deepest explanation:

- The visual system wants to infer the color of the checks in the world, not the gray value in the image.
- The "illusion" reflects the successful design of the visual system, not a quirky failure.

Apparent motion

- Visual system parses ambiguous experience into objects under several assumptions:
  - Objects typically do not disappear and appear spontaneously.
  - Objects typically follow “simple” space-time trajectories.





The perception of causality

- Michotte demos
- Heider and Simmel

Marr's three levels

- Level 1: Computational theory
  - What is the goal of the computation, and what is the logic by which it is carried out?
- Level 2: Representation and algorithm
  - How is information represented and processed to achieve the computational goal?
- Level 3: Hardware implementation
  - How is the computation realized in physical or biological hardware?

Language

- Parsing:
  - Two cars were reported stolen by the Groveton police yesterday.
  - The judge sentenced the killer to die in the electric chair for the second time.
  - No one was injured in the blast, which was attributed to a buildup of gas by one town official.
  - One witness told the commissioners that she had seen sexual intercourse taking place between two parked cars in front of her house.

Language

- Parsing
- Acquisition:
  - Learning the English past tense (rule vs. exceptions)
  - Learning the Spanish or Arabic past tense (multiple rules plus exceptions)
  - Learning verb argument structure (“give” vs. “donate”)
  - Learning to be bilingual.

Intuitive theories

- Physics
  - Parsing: Inferring support relations, or the causal history and properties of an object.
  - Acquisition: Learning about gravity and support.
    - Gravity -- what's that?
    - Contact is sufficient
    - Mass distribution and location is important
- Psychology
  - Parsing: Mind reading, or causal attribution.
  - Acquisition: Learning about agents
    - Recognizing intentionality, but without mental state reasoning
    - Reasoning about beliefs and desires
    - Reasoning about plans, rationality and "other minds".

Some philosophical puzzles

Hume's problem

- Can induction be justified?
  - E.g., Can we really *know* the sun will rise tomorrow?
- Only on the assumption of *uniformity of nature*: the future will be like the past.
- But what's the justification for that?

A modern answer to Hume?

Computational learning theory:

- PAC (Probably Approximately Correct):

  Given that a rule f has held for examples 1 … n, can we say with high probability (1-δ) that the rule will hold in most (1-ε) future cases?

- Often the answer is "yes", even without knowing how the examples were generated.
- But… still requires uniformity of nature.

Goodman's problem

- Why do some hypotheses receive confirmation from examples but not others?
  - "All piece of copper conduct electricity": yes
  - "All men in this room are third sons": no
- Distinguishing *lawlike* hypotheses from *accidental* hypotheses is not easy:
  - "All emeralds are green"
  - "All emeralds are grue", where *grue* means "if observed before *t*, green; else, blue."

Responses to Goodman

- First instinct is a syntactic response:
  - Hypotheses without arbitrary free parameters are more lawlike.
  - Simpler (shorter) hypotheses are more lawlike.

Syntactic levers for induction

- Which hypothesis is better supported by the evidence?
  - All blickets are chromium.
  - All blickets are chromium and arch-shaped.
  - All blickets are chromium or arch-shaped.
- Which curve is best supported by the data?

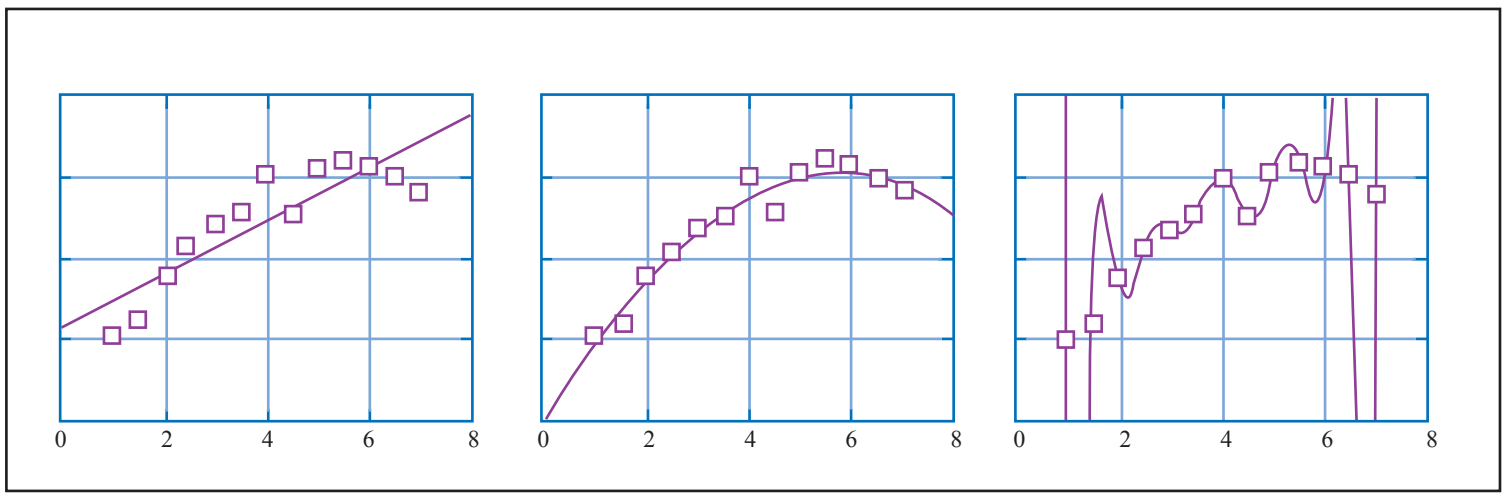


Figure by MIT OCW.

Responses to Goodman

- Hypotheses without arbitrary free parameters are more lawlike.
- Simpler (shorter) hypotheses are more lawlike.
- But “green” and “grue” are logically symmetric:
  - To a Martian who sees *grue* and *bleen*, green just means “if observed before *t*, grue; else, bleen.”

Responses to Goodman

- Hypotheses without arbitrary free parameters are more lawlike.
- Simpler (shorter) hypotheses are more lawlike.
- But “green” and “grue” are logically symmetric.
- *Lawlike* is a semantic (not syntactic) notion, and depends on prior subjective knowledge (not strictly objective world structure).

The origin of good hypotheses

- Nativism
  - Plato, Kant
  - Chomsky, Fodor
- Empiricism
  - Strong: Watson, Skinner
  - Weak: Bruner, cognitive psychology, statistical machine learning
- Constructivism
  - Goodman, Piaget, Carey, Gopnik
  - AI threads….

Plato

- *Meno*: Where does our knowledge of abstract concepts (e.g., virtue, geometry) come from?
- The puzzle: “A man cannot enquire about that which he does not know, for he does not know the very subject about which he is to enquire.”

Plato

- *Meno*: Where does our knowledge of abstract concepts (e.g., virtue, geometry) come from?
- A theory: Learning as “recollection”.
- The Talmud’s version:

“Before we are born, while in our mother's womb, the Almighty sends an angel to sit beside us and teach us all the wisdom we will ever need to know about living. Then, just before we are born, the angel taps us under the nose (forming the philtrum, the indentation that everyone has under their nose), and we forget everything the angel taught us.”

Plato meets Matlab^{tm}

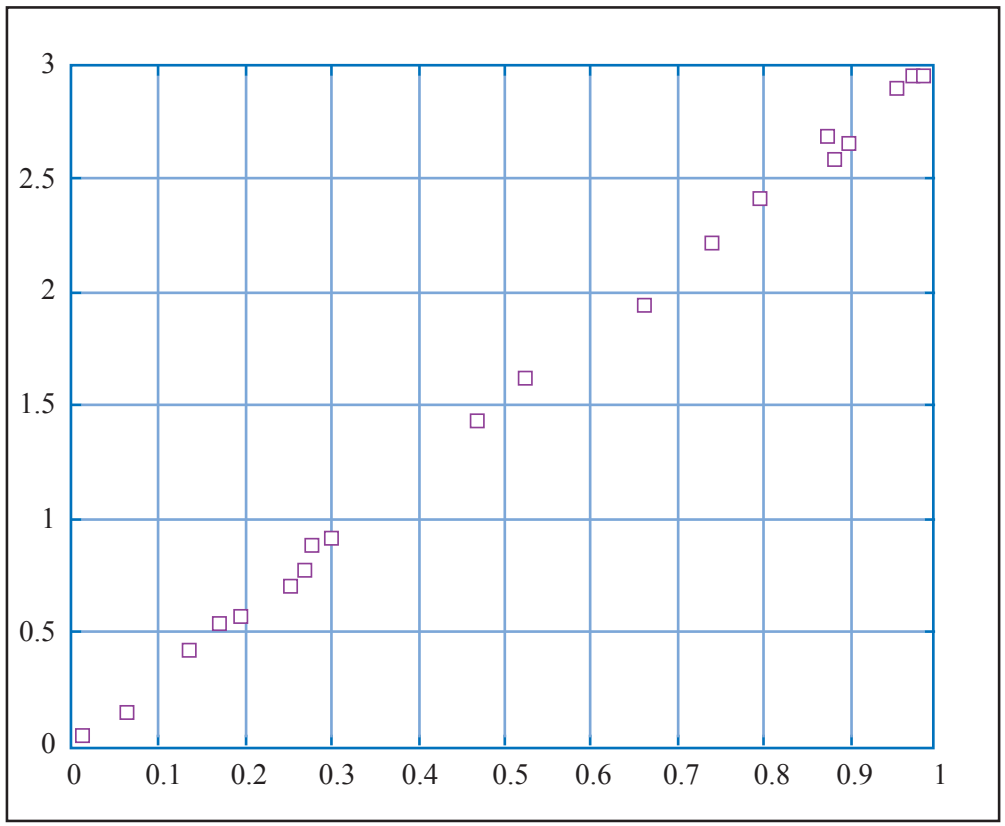


Figure by MIT OCW.

What is the relation between y and x?

Plato meets Matlab^{tm}

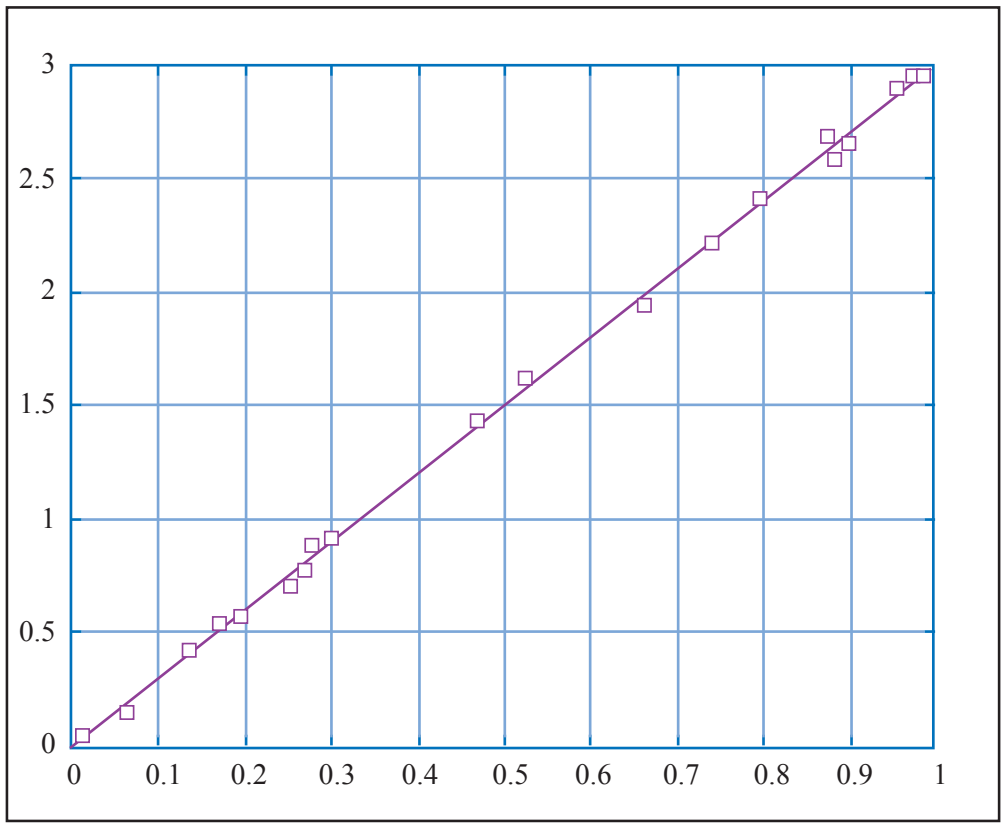


Figure by MIT OCW.

What is the relation between y and x?

Plato meets Matlab^{tm}

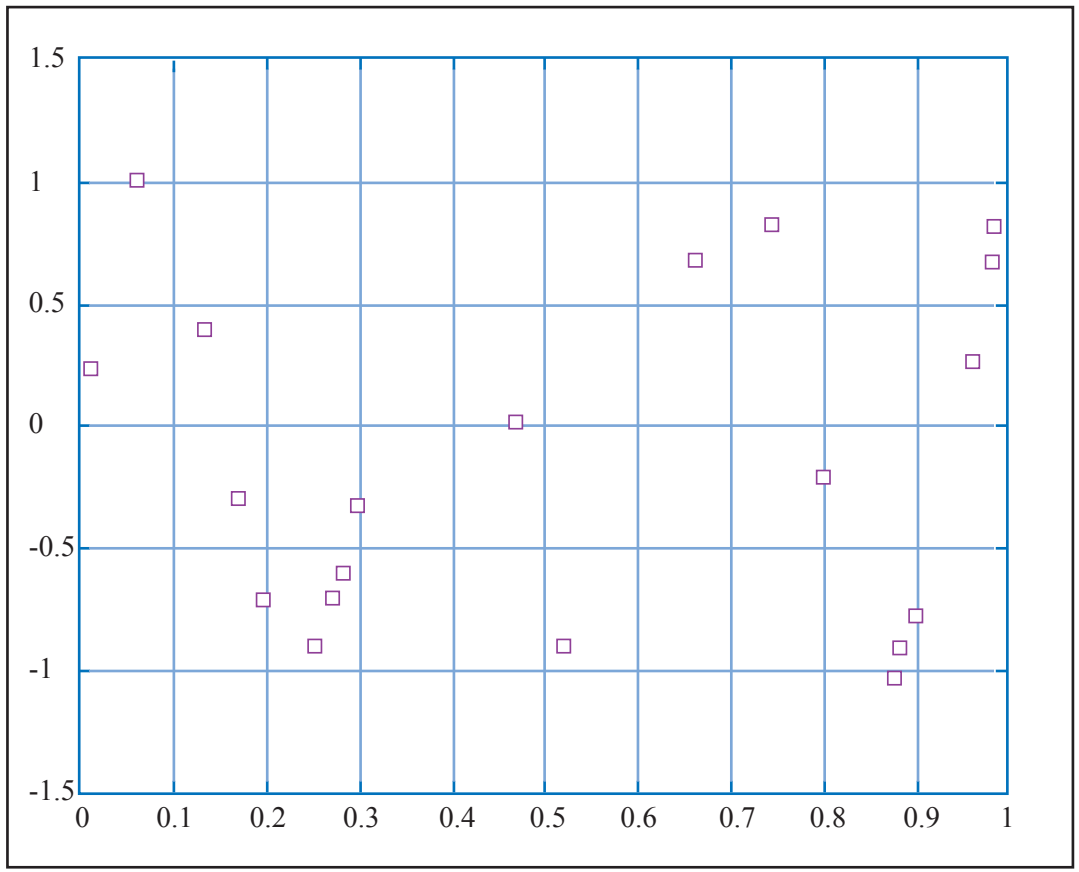


Figure by MIT OCW.

What is the relation between y and x?

Plato meets Matlab^{tm}

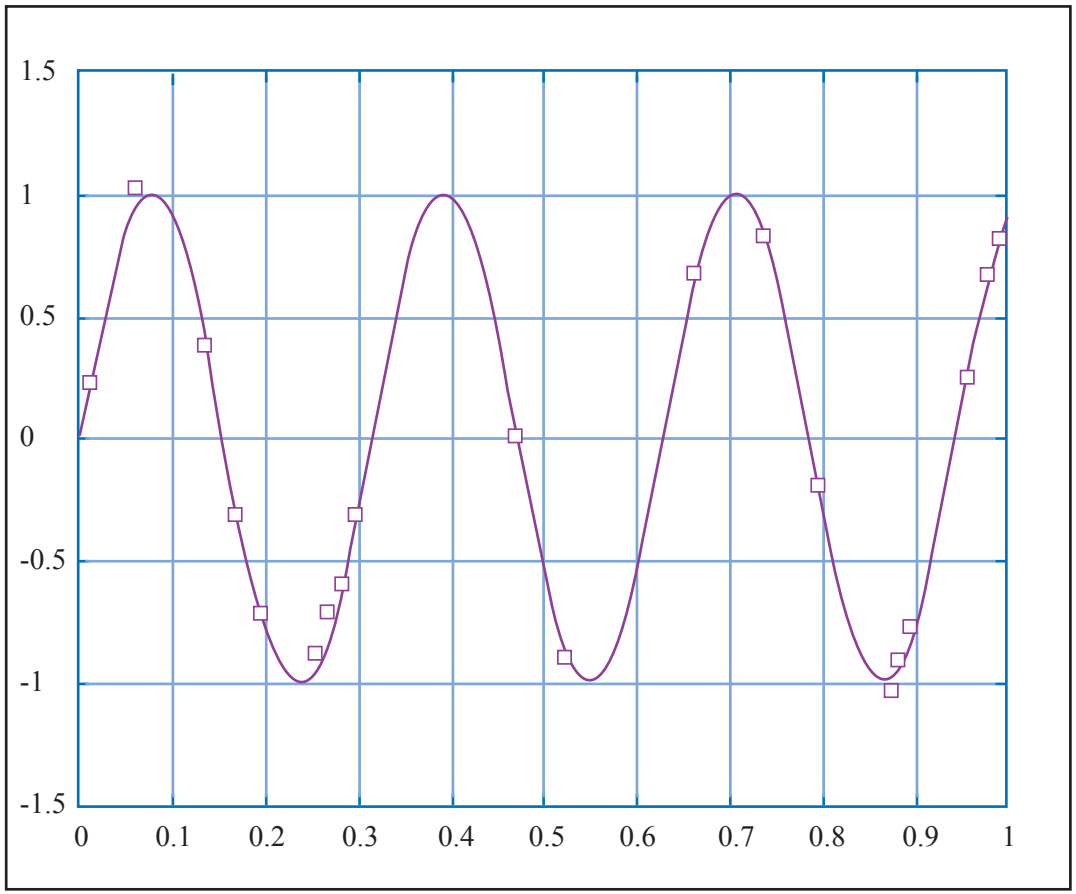


Figure by MIT OCW.

What is the relation between y and x?

The legacy of Plato

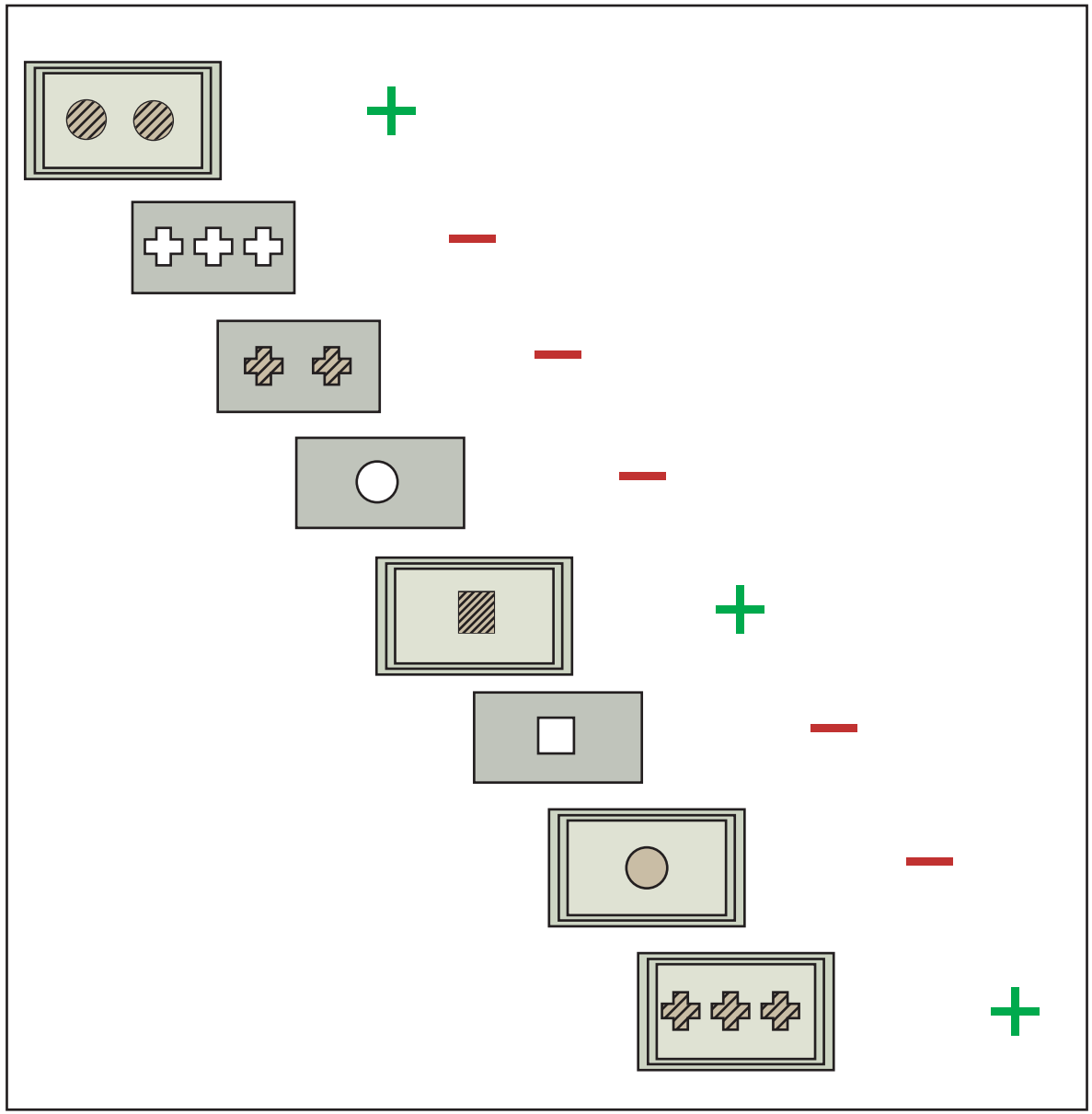
- "A man cannot enquire about that which he does not know, for he does not know the very subject about which he is to enquire."
- We can't learn abstractions from data if in some sense we didn't already know what to look for.
  - Chomsky's "poverty of the stimulus" argument for the innateness of language.
  - Fodor's argument for the innateness of all concepts.

The origin of good hypotheses

- Nativism
  - Plato, Kant
  - Chomsky, Fodor
- Empiricists
  - Strong: Watson, Skinner
  - Weak: Bruner, cognitive psychology, statistical machine learning
- Constructivists
  - Goodman, Piaget, Carey, Gopnik
  - AI threads….

Image removed due to copyright considerations. Please see:
Bruner, Jerome S., Jacqueline J. Goodnow, and George Austin. *A Study in Thinking*. Somerset, NJ: Transaction Publishers, 1986. ISBN: 0887386563.

Image removed due to copyright considerations. Please see:
Hull. "Qualitative Aspects of the Evolution of Concepts."
Psychological Monograph 28, no. 123 (1920).



"striped and three borders"

Figure by MIT OCW.

Fodor's critique

- This isn't really *concept learning*, it's just *belief fixation*.
  - To learn the rule "striped and three borders", the learner must already have the concepts "striped", "three borders", and "and", and the capacity to put these components together.
  - In other words, the learner already has the concept, and is just forming a new belief about how to respond on this particular task.
- More generally, all inductive learning seems to require the constraints of a hypothesis space -- so the learner must begin life with all the concepts they will ever learn. How depressing.

Fodor's critique

Raises major questions for cognitive development, machine learning, and AI:

- Is it ever possible to learn *truly new* concepts, which were not part of your hypothesis space to begin with?
- What conceptual resources must be innate?
  - Objects?
  - First-order logic?
  - Recursion?
  - Causality?

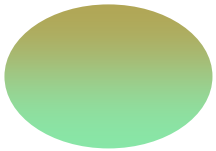
The origin of good hypotheses

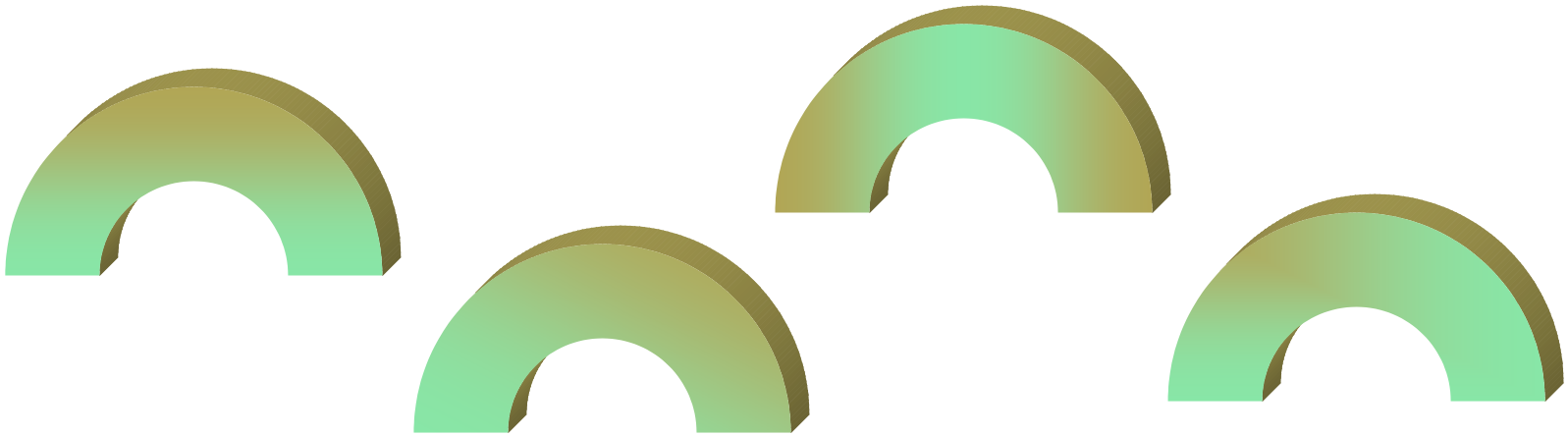
- Nativism
  - Plato, Kant
  - Chomsky, Fodor
- Empiricists
  - Strong: Watson, Skinner
  - Weak: Bruner, cognitive psychology, statistical machine learning
- Constructivists
  - Goodman, Piaget, Carey, Gopnik
  - AI threads….

Goodman's answer to Goodman

- More lawlike hypotheses are based on "entrenched" predicates: *green* is more entrenched than *grue*.
- How does a predicate become entrenched? Is it simple statistics: how often the predicate has supported successful inductions in the past?
- Suppose *grue* means "If observed on Earth, green; if on Mars, blue."
- Entrenchment could come through experience, but could also derive from a causal theory. Theory supported by experience seems best.

How do theories work?

- See this look?  It's called "chromium".
- Here are some blickets:



- Which hypothesis is more lawlike?
  - "All blickets are chromium"
  - "All blickets are chromirose", where *chromirose* means "if observed before *t*, chromium; else rose-colored."

How do theories work?

- Theories depend on abstract categories.
  - E.g., *chromium* is a kind of color or material.
- Abstract categories depend on theories.
  - E.g., species, magnetic pole
- Theories support hypotheses for completely novel situations.
- Big open questions:
  - What is a theory, formally?
  - How are theories learned?