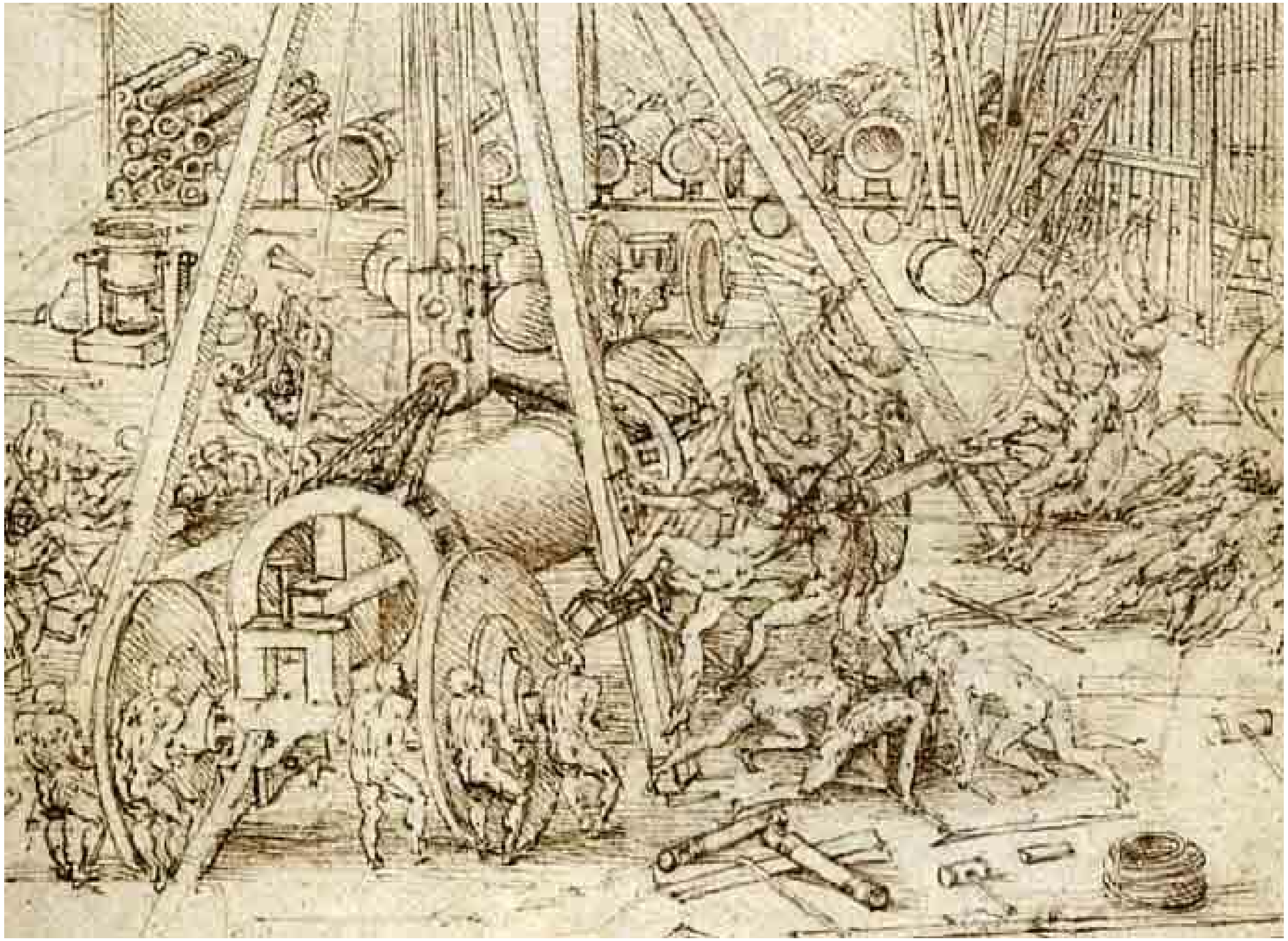
Neural Nets and Symbolic Reasoning

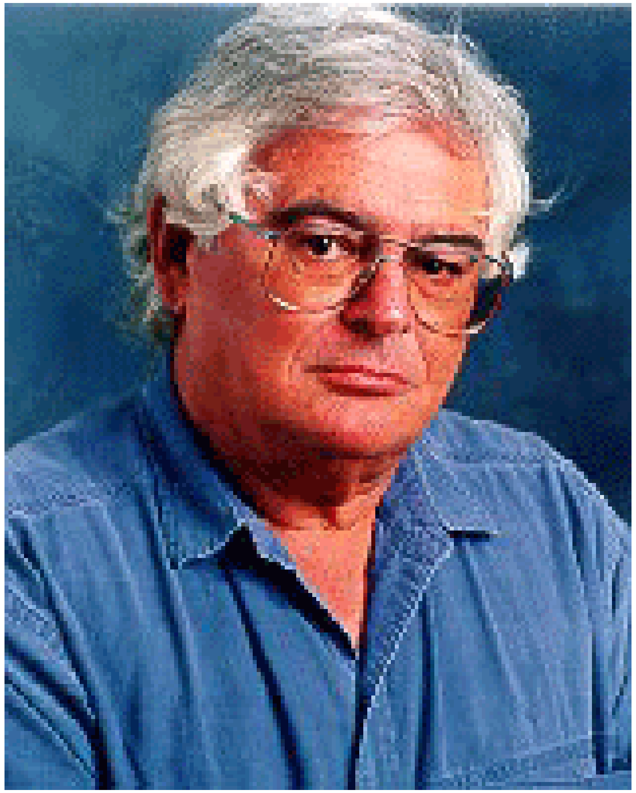
Opening the connectionist-symbolist debate



Outline

- Overview of criticism
- Productivity of thought and systematicity of representations
- Compositionality of representations
- Inferential coherence
- General conclusions

1 Overview of criticism



The starting point

- Distinguishing *representationalist* and *eliminativist* approaches to theoretizing about cognition
- Representationalists claim that the internal states of the cognitive system are "representational" (*intentional, semantic*) states that *encode states of the world*
- Eliminativists (*Watson, Churchland, Stich*) dispense with such semantic notions
- Connectionism is on the **representationalist** side
- Connectionist systems are inadequate as representational systems.

Fodor's language of thought hypothesis

The character of representations is crucial

- Cognitive activities require a language-like representational medium
- Symbolic representations have a combinatorial syntax and semantics
- This means that representations are composed of constituents, which may themselves be composed of smaller constituents, and so forth
- Unpacking the structure eventually yields atomic elements (the elements of representation)
- Rule-governed processes operate on representations. They are syntactic, i.e. they are applied with respect to form
- Compositional semantic interpretation: the syntactic engine mimics a semantic engine.

Claimed advantages of Mentalese

- Provides an account of beliefs, intentions and doubts etc. these are also expressed in terms of Mentalese sentences
- Provides an account of productivity and systematicity
- Human language is productive: No limit to number of sentences we can produce
- And systematic: if you can say *John loves Mary*, can also say *Mary loves John*
- Human thought: productive and systematic because it relies on Mentalese, which is productive and systematic.

Main critic points

Connectionism does not realize the essentials of a language of thought

Connectionist systems lack a combinatorial syntax and semantics. Constituent structure cannot be defined. Therefore connectionist systems cannot deal with three important properties of cognitive systems:

1. The productivity of thought: the capacity to understand and produce indefinitely many propositions
2. The sytematicity of thoughts: the intrinsic connection between the ability to comprehend or think one thought and the ability to comprehend or think certain other thoughts
3. Compositional semantic interpretation
4. The coherence of inference: the ability to make systematic inferences.

2 Productivity of thought and systematicity of representations

Productivity of thought

- The idea of productivity is pretty clear (the capacity to produce indefinitely many sentences/propositions) and adequately formalized (Turing machines)
- Several papers show that already finite discrete-time recurrent networks have the computational power of Turing machines

  See, e.g.

  - Hava Siegelmann & Eduardo Sontag: On the computational power of neural nets
  - Jiri Sima & Pekka Orponen: A computational taxonomy and survey of neural network models

Siegelman & Sontag

- If one moves from binary-state to analog-state neurons, then arbitrary Turing machines may be simulated by single, finite recurrent networks

- The original construction recurred 1058 saturated-linear neurons to simulate a universal Turing machine, but this has later been improved to at least 114 neurons or even 25 neurons

- In his criticism of connectionism, Fodor & Pylyshyn don't exclude such results, what they exclude, however, is that the connectionist solution finds a non-artificial solution to the puzzle of productivity that follows from the basic nature of architecture alone, and not merely be compatible with the architecture.

Systematicity according to Fodor & Pylyshyn

All cognitive systems (humans and other animals) are systematic, i.e., are such that their ability to do some things of a given cognitive type (including at least "thinking a thought" and making an inference) is intrinsically connected with their ability to do other, structurally related things of that type.

Examples

- If I understand what it means that *Peter likes Maria* I also understand what it means that *Maria likes Peter*
- If I understand the concept of *a brown cow* and *a black horse* I also understand the concept of *a black cow* and *a brown horse*

Cognitive architecture must *explain* systematicity

- Fodor & Pylyshyn place a strict constraint on what counts as genuine explanation. For an hypothesis H to explain some phenomenon S, H alone must entail S
- This constraint is essential to their dismissal of connectionism. They imagine a defender of connectionism building systematicity into a particular connectionist model and claiming, on this basis, that connectionism can indeed explain systematicity
- Fodor & Pylyshyn reply that this would be insufficient. Since all natural cognitive systems are systematic, systematicity must follow from the basic nature of architecture alone, and not merely be compatible with the architecture (p.50).

Can classical architecture *explain* systematicity?

- Consequently, if the classical conception of cognitive architecture explains systematicity, it likewise must entail systematicity from the basic nature of the architecture alone.
- Classical Architecture (H): All natural cognitive systems contain (a) mental representations with combinatorial constituent structure and compositional semantics, and (b) mental processes that are sensitive to the combinatorial structure of the representations.
- Empirical phenomenon (S): All natural cognitive systems are systematic.

This entailment is neither immediate nor obvious. Whether it goes through depends on what the phenomenon of ***Systematicity*** actually is.

Systematicity Schema

For every organism O, and any given series of cognitive performances *t* of type T, there is some set $M_{O,t}$ of *structurally related* performances such that O is capable of all and only the performances in $M_{O,t}$

(van Gelder & Niklasson)

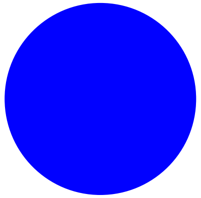
- In the absence of any particular specification of how the Systematicity Schema is to be filled out, there is no determinate answer to whether classical architectures entail and hence explain systematicity

- Suppose that $C_{O,t}$ is the set of systematically related performances as predicted by classical architectures, and for the real set: $M_{O,t} \subset C_{O,t}$. Then, the classical architecture over-generates and fails to explain systematicity. This seems to be the case in many particular domains.

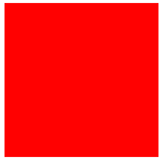
Adjectival modification

Let *t* be the cognitive performance of understanding the meanings of two (absolute) adjective – noun combination, e.g. *blue circle* and *red square*, then $C_{O,t}$ would be the set of all combinations {*blue,red*}{*circle,square*}.

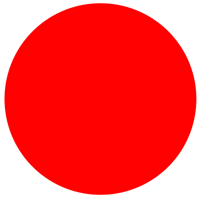
blue circle



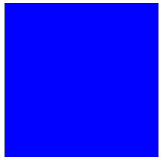
red square



red circle



blue square



Such examples suggest $M_{O,t} = C_{O,t}$, and classical architecture seems to explain systematicity in the selected domain.

Details??

(A) Lexicon

e.g. blue ⇒ BLUE, *square* ⇒ SQUARE

(B) Translation rule

e.g. $(x_{adj}\ y_{noun})_{N'} \Rightarrow X \cap Y$ (assuming $x \Rightarrow X$ and $y \Rightarrow Y$)

(C) Extensions

for all primes P the extensions $\|P\|$ are given *a priori*.

Assume now the system understands *blue circle*. This means there is a translation like $(\textit{blue circle})_{N'} \Rightarrow \text{BLUE} \cap \text{CIRCLE}$. From this we conclude *blue* ⇒ BLUE, *circle* ⇒ CIRCLE, *adjective-noun* combination ⇒ intersection operation ∩. Similarly for *red square*. This is sufficient then to calculate the translations of all four combinations {*blue,red*} {*circle,square*}.

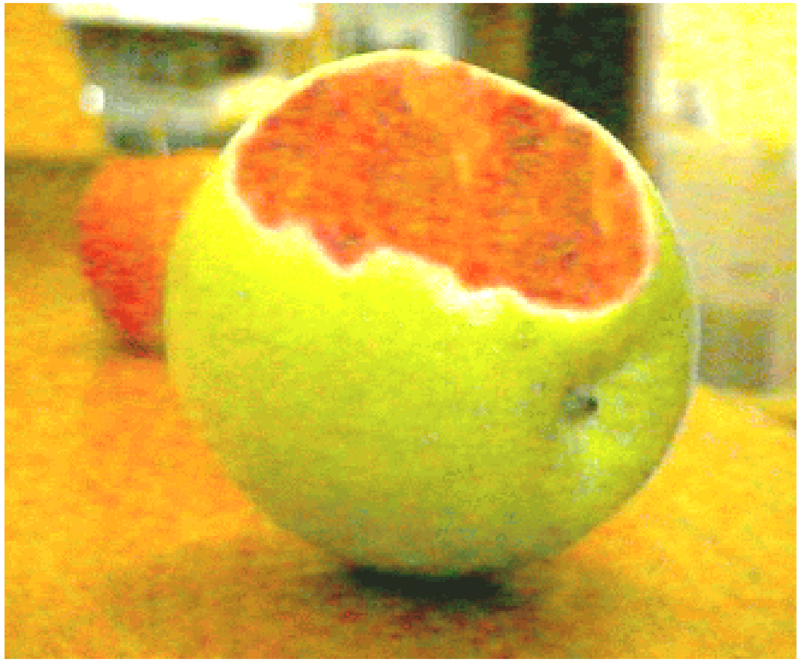
Lahav

Lahav (1993) claimed that the intersection operation that is crucially involved in the scheme does not work for many (absolute) adjective–noun combinations if the noun refers to objects with an extended part-whole structure.

> In order for a cow to be brown most of its body's surface should be brown, though not its udders, eyes, or internal organs. A brown crystal, on the other hand, needs to be brown both inside and outside. A brown book is brown if its cover, but not necessarily its inner pages, are mostly brown, while a newspaper is brown only if all its pages are brown. For a potato to be brown it needs to be brown only outside, ... Furthermore, in order for a cow or a bird to be brown the brown color should be the animal's natural color, since it is regarded as being 'really' brown even if it is painted with all over. A table, on the other hand, is brown even if it is only painted brown and its 'natural' color underneath the paint is, say, yellow. But while a table or a bird are not brown if covered with brown sugar, a cookie is. In short, what is to be brown is different for different types of objects. To be sure, brown objects do have something in common: a salient part that is wholly brownish. But this hardly suffices for an object to count as brown. A significant component of the applicability condition of the predicate 'brown' varies from one linguistic context to another. (Lahav 1993: 76)

Quine's *red apple*

– *a red apple*	[red peel]
– *a sweet apple*	[sweet pulp]
– *a reddish grapefruit*	[reddish pulp]
– *a white room/ a white house*	[inside/outside]



A red apple?

What color is an apple?

Q_1 What color is its peel?

Q_2 What color is its pulp?

Architectural conclusions 1: Meaning and interpretation

The observation: Linguistically encoded information doesn't fully specify the truth conditions of a sentence.

- Katz & Fodor (1963): A full account of sentence interpretation has to include more information than that of syntactic structure and lexical meaning.

  a. *Should we take the lion back to the zoo?*

  b. *Should we take the bus back to the zoo?*

- Psycholinguistics: Mental models, situation structure,...
  The tones sounded impure because the hem was torn.

Architectural conclusions 2: Underspecification

The tones sounded impure because the hem was torn.

SCOTTISH PIPER

Theoretical Models

- **Kaplan's distinction between *character* and *intension***

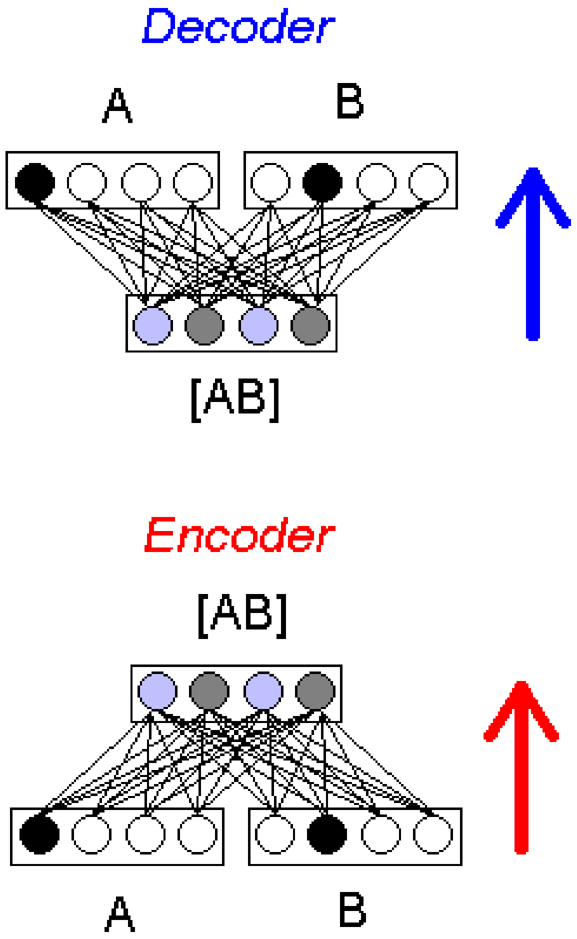
  intension = character(F)

- **Radical Underspecification View**

  Underspecified representations + contextual enrichment (Hobbs 1983, Alshawi 1990, Poesio 1991, Pinkal 1995, etc.)

  => Find optimal (energy-minimal!) enrichments!

3 Compositionality of representations



Concatenative and functional compositionality

Rule-based systems employ concatenative compositionality

Representations involve the linking or ordering of constituents without alteration

Networks display functional compositionality

Components are fractionated within representations, but are recoverable at output (cf. RAAM)

(((ab)c)d): *raam2*(d, *raam2*(c, *raam2*(b, *raam2*(a, nil))))

Compositionality of meaning: property of meaning assignment ***m***

m(*raam2*(x,y)) = *raam2'* (***m***(x), ***m***(y)),

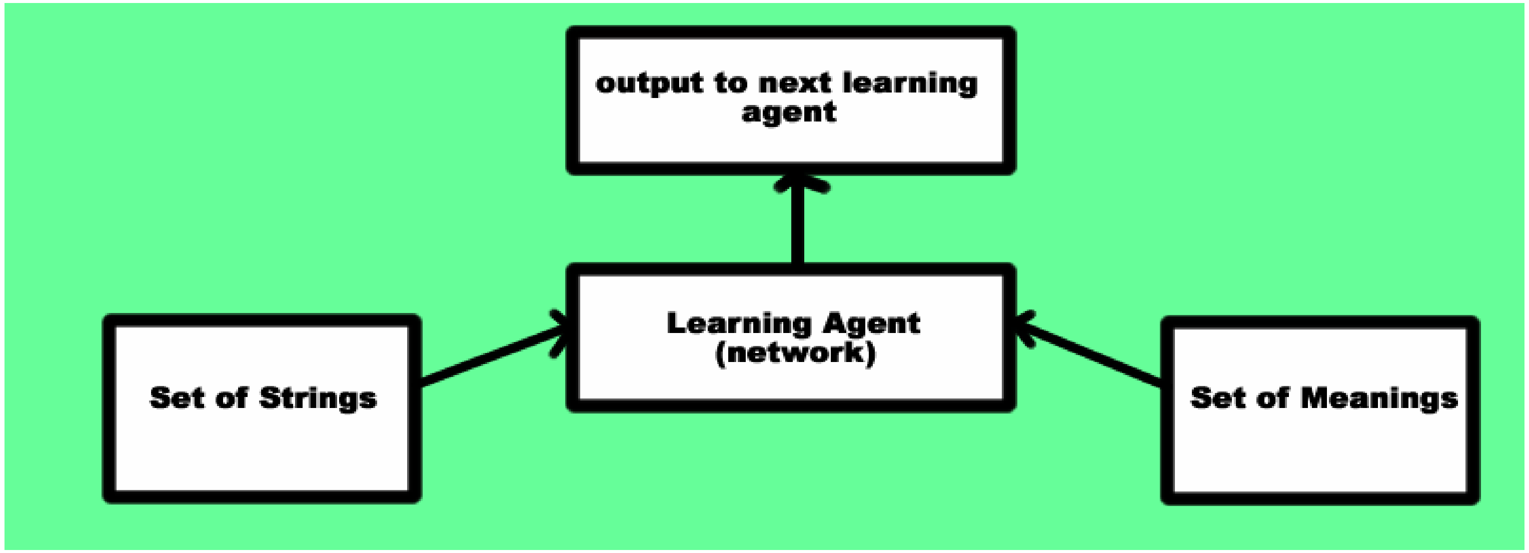
with *raam2'* : binary composition of semantic elements

Compositionality and cognitive architecture

- If I can entertain the notion of a *wug*, I can entertain the notion of a *big wug* or a *red wug* or *a wug that is on the table*
- According to Fodor & Pylyshyn, compositionality must follow from the basic nature of architecture alone, and not merely be compatible with the architecture.
- RAAM (and other mechanisms) show that connectionism is compatible with a compositional architecture.
- Why this is sufficient: There are non-compositional (holistic) modes of cognition, e.g. contextual strengthening in adjectival modification, idiom chunks (*kick the bucket*), metaphors. Hence, compositionality should **not** be a general consequence of the basic architecture!

The emergence of compositionality

- The iterated learning model (ILM) was developed by Simon Kirby (1999; 2000) to test a number of hypotheses concerning the role of learnability in language evolution
- The ILM is a series of learning agents (nets) occupying a problem domain, e.g. mapping strings to meanings.
  A subset of the mappings becomes the input to the next learning agent, which then repeats the process.



The emergence of compositionality, cont.

- In the ILM, simulations are initialized with a random language

  Initial agent productions are holistic and unstable

- Subsequent agents are exposed to an **insufficient** subset of the previous agent's output

  i.e. stimulus is impoverished

- Holistic languages are unlearnable

  Why?

- Eventually, some part of the language stabilizes

  i.e. regular, compositional representations sporadically emerge.

The emergence of compositionality, cont.

- As simulations continue, the amount of compositionality accrues
  - Compositional mappings are more stable in transmission from agent to agent
  - The language thus becomes more learnable
- After n-generations of learning agents, a compositional language emerges
- Only a compositional language can pass through the information bottleneck known as the ‘Poverty-of-Stimulus’
  - The poverty of stimulus is thus a feature of the evolutionary environment that selects for communication systems displaying compositionality

4 Inferential coherence

Systematicity of inference

- Fodor & Pylyshyn single out the systematicity of inference as a key component of the wider phenomenon of systematicity
- It is, roughly, the idea that the ability to make some inferences is intrinsically connected to the ability to make other, logically related inferences
- No precise definition is given, only anecdotally illustrations
- Example: You don't find minds that are prepared to infer *John went to the store* from *John and Mary and Susan and Sally went to the store* and from *John and Mary went to the store* but not from *John and Mary and Susan went to the store*. (p.48).

  Fodor &Pylyshyn: A logical rule is involved

Example: *Modus Tollens*

- Consider modus tollens: $A \supset B, \sim B \Rightarrow \sim A$
- Note that *modus tollens*, taken as a logical rule, entails the following systematicity sub-hypothesis (Van Gelder & Niklasson):

  Let $M_{O,MT}$ be the set of inferences by substituting into the modus tollens schema ($A \supset B, \sim B \Rightarrow \sim A$) any symbol in the set of symbols available to O. Then we have:

  Systematicity of Modus tollens (SMT): Any organism O capable of performing any instance of $M_{O,MT}$ is capable of performing every instance of that set.

SMT is empirically questionable

Study by Kern, Mirels & Hinshaw (1983). They presented subjects with abstract and concrete forms:

- Do $P \supset Q$ and *not-Q* imply *not-P*?
- Do *If Rex is a terrier, then he likes apples*, and *Rex does not like apples*, imply *Rex is not a terrier*?

Referents	Psychologists	Biologists	Physicists	Overall % correct collapsed across disciplines
abstract	33	33	58	41
concrete	50	83	75	69

Conclusions

- Assuming representativeness, this suggests that around 30% of people do not perform identically on structurally identical inferences, i.e., directly violate hypothesis SMT
- The clusters that these people's cognitive capacities come in are *not* the clusters entailed by the hypothesis that they have a classical architecture
- The critical role of content in conditional inference has been confirmed repeatedly in one of the most-studied tasks in the psychology of inference, the Wason selection task (Wason, 1966)
- Good solutions have to integrate the role of content. [theme 4, Bechtel: Natural deduction in connectionist systems]

5 General conclusions

- Connectionismus as providing mere implementation?
- The integrative view as an alternative

Connectionism as providing mere implementation

Assume that's all right: Connectionism lacks a combinatorial syntax etc., what is the consequence for connectionism then? To eliminate it?

- Connectionism is merely an account of the *medium* within which the symbolic representational system is implemented
- Only the analysis at the level of symbolic processing is relevant to cognitive theorizing, and this level is *nonconnectionist*
- Fodor is a *functionalist* (multiple realization!): A single function may be implemented in any one of a number of lower-level mechanism (like *money* that can be physically realized in paper, metal, stones)

My Footnote: Chomsky doesn't believe in multiple realization and takes (neuro)biology serious as a science that may restrict cognitive architecture.

The integrative view as an alternative?

- Do away with Putnam's multiple realization!
- Connectionism and symbolism as two different perspectives of the same system (like the wave and the particle picture in QM)
- The connectionist level can create restrictions for higher-level, symbolic architectures. Example: Optimality Theory
- However, this is not the only way to pose restrictions. The evolutionary perspective is another one (Kirby, Hurford etc.)

Symbolic grounding as most important aspect of embodiment (the content of a symbol cannot be explained by referring to other symbols)