EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Part I
On the very idea of an LoT

Fausto Carcassi

Introduction & disclaimers

Me: Fausto Carcassi

Various practicalities

- For content / slides: https://github.com/thelogicalgrammar/pLoT_workshop
- You can run labs on Google Colab 
- You can give me feedback on the website.

Disclaimers

- This is a mostly *informal/imprecise* introduction
- I'll be mostly Fodorian – but other options exist
- I'll assume little and explain a lot

Introduction & disclaimers

Who has heard about…

- Probabilistic context-free grammars
- Lambda calculus
- Bayesian inference

The overall plan

Part I	**Introduction: On the very idea of an LoT**
Part II	Technical background
Part III	Bayesian program induction (LOTlib3)
Part IV	Case studies
Part V	Summary & Future prospects

The plan for the session

- Some observations about the mind
- The Language of Thought (LoT)
- How the LoT makes sense of the observations
- The probabilistic LoT (pLoT)

Learning a rule

Robert Feldman → Dr Feldman

Ruth Millican → Dr Millikan

Joanna Newsom → ??

- Dj Newsom

"Dr <*last name*>" **or** "D<*first letter of first name*> <*last name*>"?

6 @ 2 = 12

3 @ 4 = 12

10 @ 2 = ??

- 12

@ = multiplication **or** return 12?

Learning a rule

Humans can learn a huge amount from a single instance.

Lake et al (2015)

- Based on just this one instance, we can do loads.
- E.g., classify new examples:

Learning a rule

Humans can learn a huge amount from a single instance.

Lake et al (2015)

- Based on just this one instance, we can do loads.
- E.g., generate new examples:

Learning a rule

Humans can learn a huge amount from a single instance.

Lake et al (2015)

- Based on just this one instance, we can do loads.
- E.g., parse the object into parts:

Learning a rule

Humans can learn a huge amount from a single instance.

Lake et al (2015)

- Based on just this one instance, we can do loads.
- E.g., generate new concepts:

Learning a rule

Humans can learn a huge amount from a single instance.

Lake et al (2015)

- Based on just this one instance, we can do loads.

What allows us to do this?

Pills of Fodorianism

Productivity

We can **think** indefinitely many thoughts

E.g., for every natural number *n*, "My favorite number is *n*"

"But we are finite beings – we can't think infinitely many thoughts!"
Finite performance, but competence for unboundedly many representations

Cf English: Finitely many sentences in a lifetime, but not in principle

- "Mary thinks that John are the apple"
- "John thinks that Mary thinks that John ate the apple"
- Etc.

Systematicity

The ability to represent (a) and (b) are related:

(a) ‘John is close to the fish’
(b) ‘The fish is close to John’

(a’) ‘It rains and Mary sleeps’
(b’) ‘Mary sleeps and it rains’

We say thoughts are *systematically* related:

- They differ in arrangement
- Same building blocks

Inferential coherence

If we can draw inference (a), we can draw inference (b)

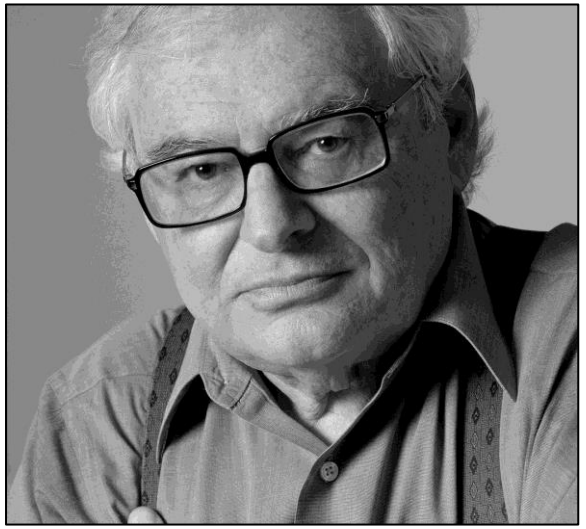
(a) It rains **and** it is wet –> It rains
(b) John sleeps **and** the cat purrs –> The cat purrs

And any inference with a similar structure!

(c) If blorgs zorg and bixes bon –> Blorgs zorg

Questions?

The LoT

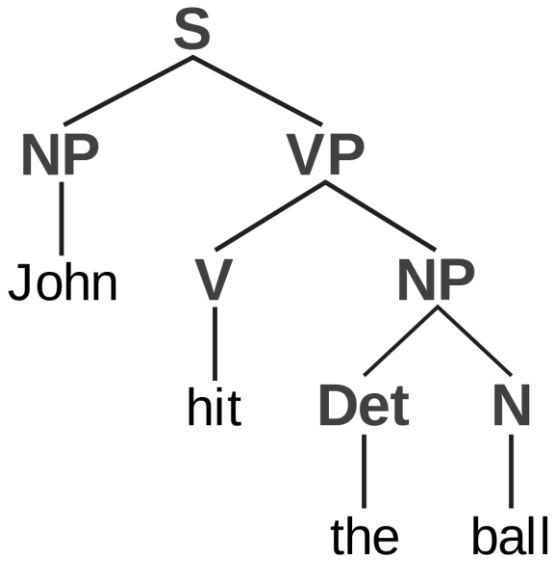


The main idea

Main claim: Thinking happens *in a language*.

"In what sense is it like English?"

- **Combinatorial grammar**
  - Basic building blocks ("atoms of thought")…
  - …combined into hierarchical structures (sentences)…
  - …with a grammar: only some combinations allowed!
- **Compositional semantics**
  - Sentences have a meaning, which depends on…
  - …the meaning of the building blocks and…
  - …the way they are combined.

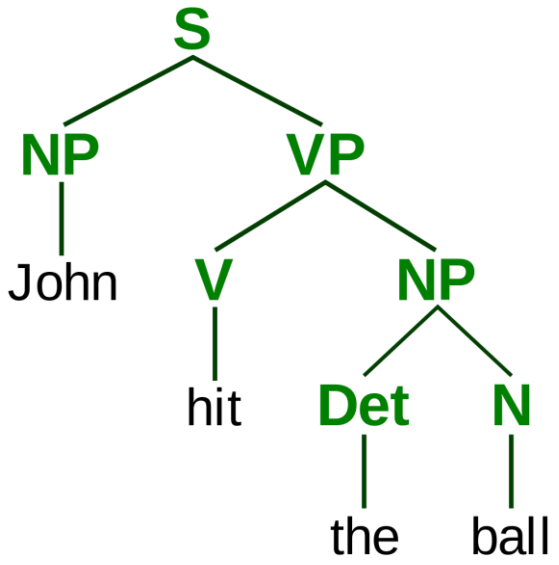


The main idea

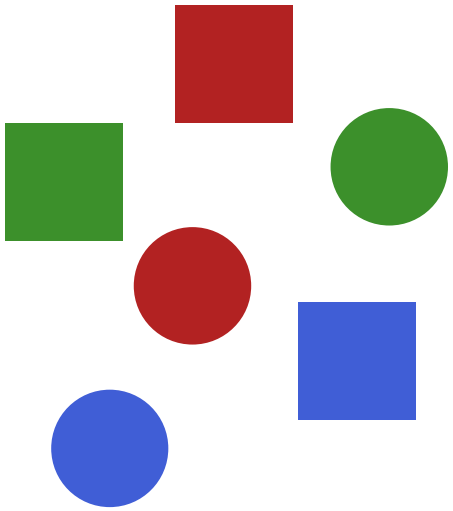
Main claim: Thinking happens *in a language*.

"In what sense is it *not* like English?"

- **No phonetics** or writing system
- **Not usable for communication** with others
  - Though maybe *within* the mind?
- Meanings are **not represented** (Fodorian point!)
  - Thinking happens *in* a language
  - Cf machine language vs interpreted language



The main idea – example



LoT representation

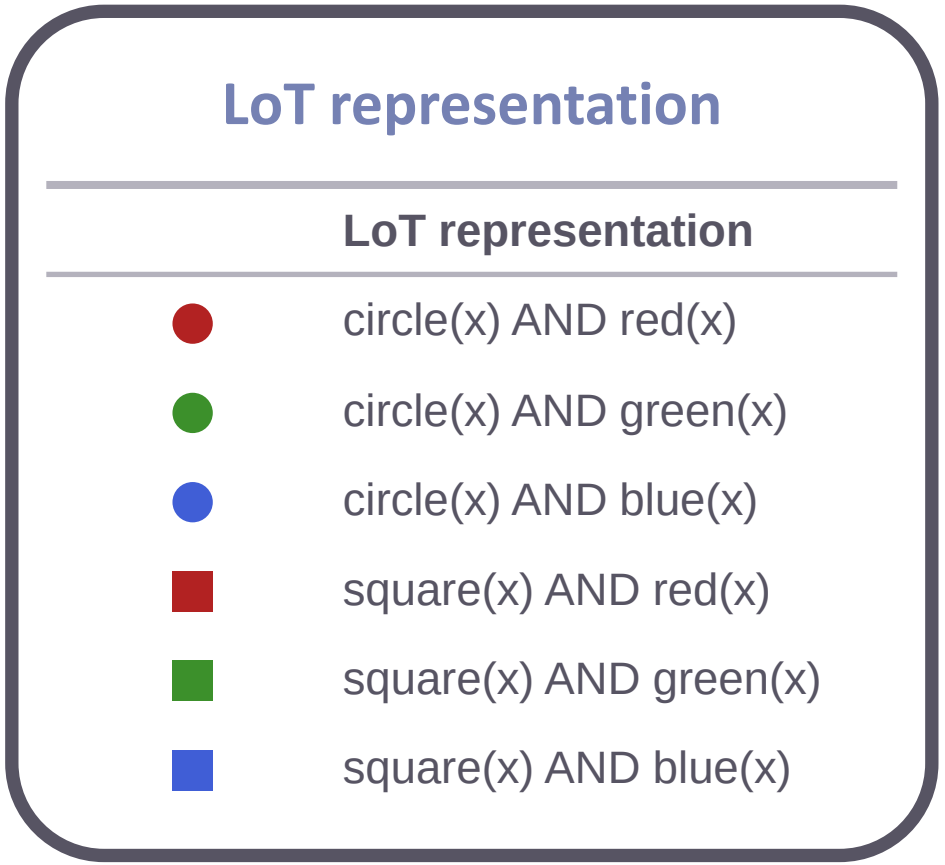
	LoT representation
●	circle(x) AND red(x)
●	circle(x) AND green(x)
●	circle(x) AND blue(x)
■	square(x) AND red(x)
■	square(x) AND green(x)
■	square(x) AND blue(x)

Properties - Quilty-Dunn, Porot, & Mandelbaum (2022)

1. Discrete constituents
   circle, AND, square, etc.

2. Role-filler independence
   AND means the same across objects

3. Predicate-argument structure
   circle, red, etc.: attributes *of* objects

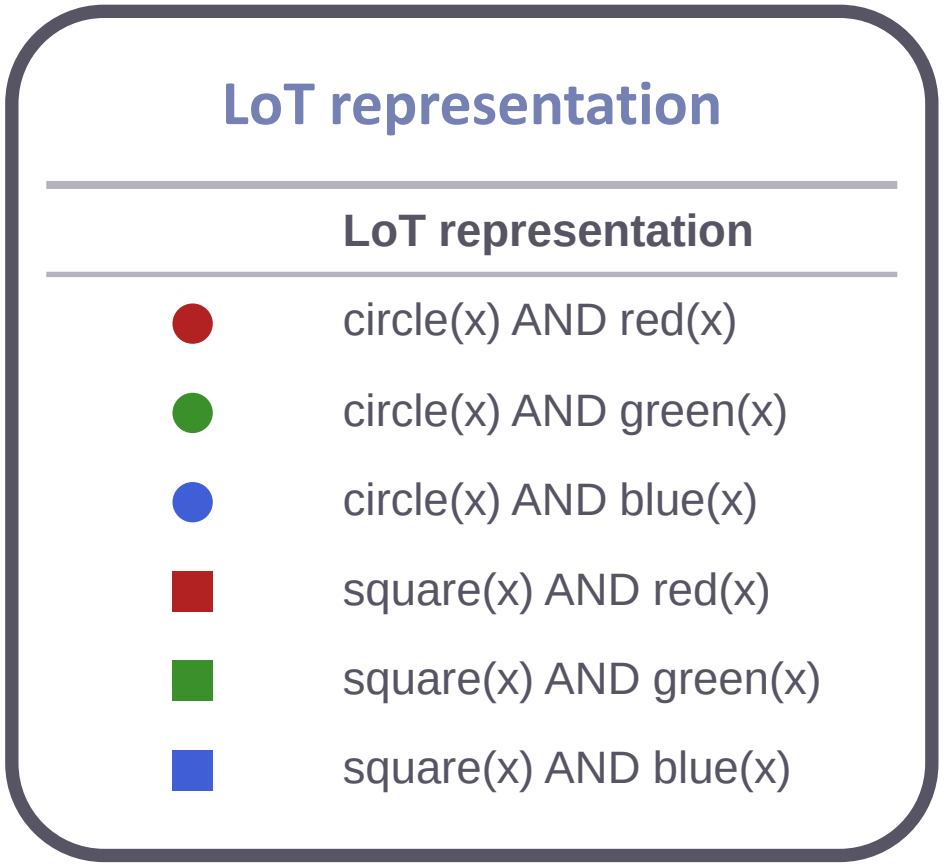


Properties - Quilty-Dunn, Porot, & Mandelbaum (2022)

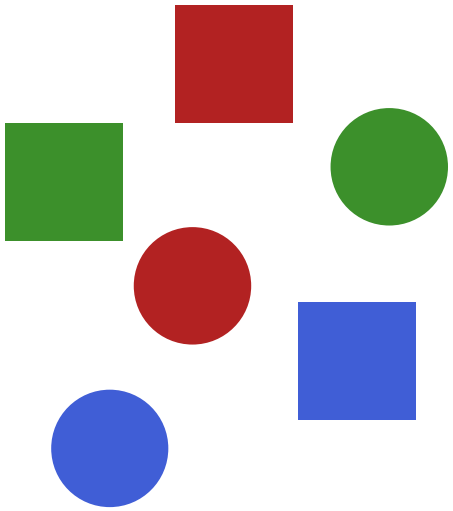
4. Logical operators
   LoT contains word for AND etc.

5. Inferential promiscuity
   Inference proceeds automatically & is sensitive only to form.
   circle(x) AND red(x) –> red(x)
   circle(x) AND blue(x) –> blue(x)

6. Abstract content
   'square' does not specify a color



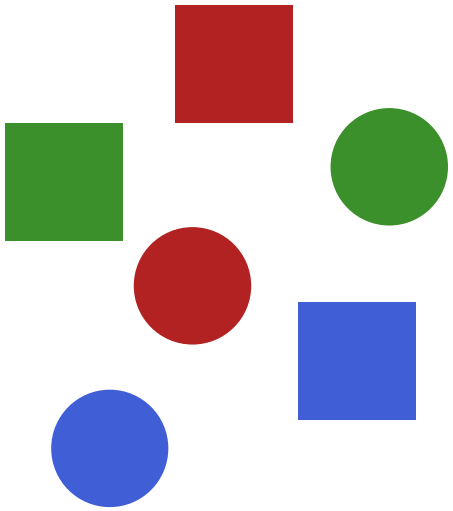
Alternative representational formats



Disentangled bit encoding

	n1	n2	n3	n4
●	1	1	0	0
●	1	0	1	0
●	1	0	0	1
■	0	1	0	0
■	0	0	1	0
■	0	0	0	1

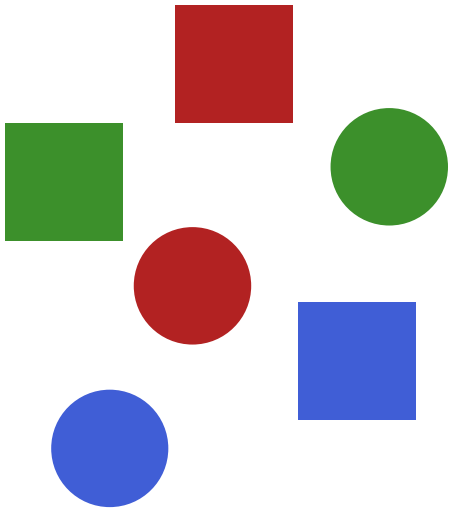
Alternative representational formats



Holistic bit encoding

	n1	n2	n3	n4
●	0	1	0	0
●	0	1	0	1
●	0	0	1	0
■	0	0	1	1
■	1	0	0	1
■	1	0	0	0

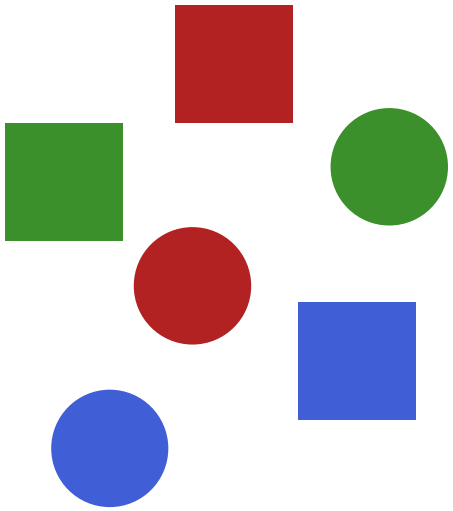
Alternative representational formats



Reals intervals

	LB	UB
●	12	18
●	18	19
●	- ∞	0
■	0	6
■	6	12
■	19	∞

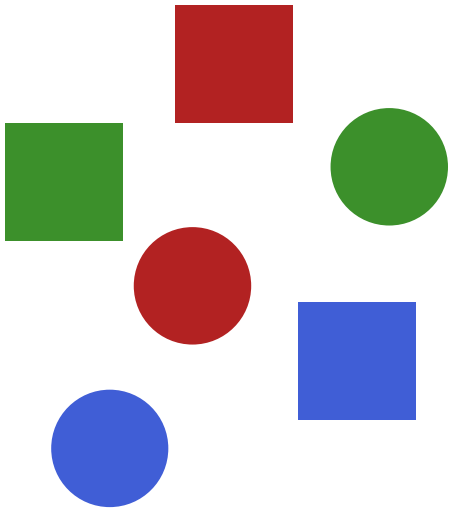
Alternative representational formats

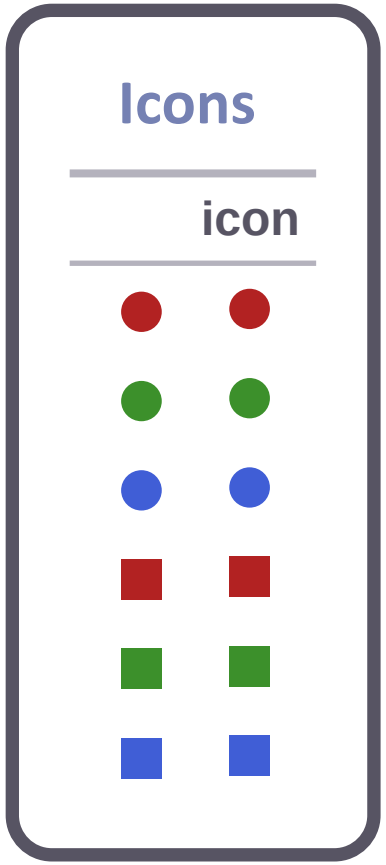


Embeddings

	n1	n2	n3
●	-19	0.3	54.4
●	3	-10.2	23
●	-54.3	0	81
■	-7.3	4	34
■	-776	9.7	34.6
■	13	0.01	123

Alternative representational formats





A rough sketch of failures

	Discrete constituents	Role-filler independence	Pred-arg structure	Logical ops	Inferential promiscuity	Abstract concepts
Disentangled bits	✓	✓		✓	✓	✓
Holistic bits	✓					✓
Intervals of reals						✓
Embeddings				✓	✓	✓
Icons / maps		✓	✓			

The Fodorian challenges

Productivity
Primitive symbols + combination rules which we can apply iteratively

Systematicity
John is close to the fish is a sentence in the LoT w/ expressions for *John*, *being-close-to*, etc.
W/ some plausible assumptions, you can reshuffle to get systematically connected thoughts.
Question: How can we exclude ill-formed thoughts like 'Is close John'?

Inferential coherence
Reasoning consists of…
- computations over mental representations (which are expressions in the LoT)
- that only depends on *form*

Block (1995) *The Mind as the Software of the Brain*

(A simplified version of) Daniel Dennett

- We have infinitely many beliefs!
- For instance, '2 is smaller than 3', '2 is smaller than 4', etc.
- Implausible that we store infinitely many LoT expressions

Ned Block

- Distinguish *ordinary* and *scientific* notions of belief.
- In ordinary sense, we have infinitely many beliefs.
- In scientific sense, we only have finitely many beliefs.
- *Causally active beliefs*: they cause behaviour or other mental states, or they are caused by perception or other mental states.

Block (1995) *The Mind as the Software of the Brain*

Dennett

- We ascribe beliefs that are not explicitly represented.
- E.g., guitarist: “I have to switch to an A-”
  - Despite lacking any such explicit representation!
- Pre-theoretical notion of belief ≠ LoT notion.

Block

- Again: LoT concerns scientific, *cognitive* notion of belief

Block (1995) *The Mind as the Software of the Brain*

Dennett's thought experiment:

- Surgeon inserts LoT expression "I have a pet turtle" as a belief in your brain.
- You wake up & are asked whether you have pets
- You say "Indeed, I have a pet turtle"
- But! When asked how old it is and what its name is, you can't answer, etc.
- Intuitively: You don't really believe it!

Block

- The functional role for belief includes integration with other beliefs, other mental states, and behavior. An isolated belief isn't a belief.

The pLoT

The “p” in “pLoT”

‘pLoT’ stands for ‘*probabilistic* LoT’

Recent (~20 years) extension: LoT enriched w/ probabilistic inference

Examples of category

In category?

Much more in the rest of the course!

Taking stock – a family of claims

What the LoT is like:

- Compositional language w/ words
- Some logical operators and predicates

What the LoT can do:

- Content of beliefs/desires/intentions/… are LoT expressions
- Reasoning -> manipulation of LoT expressions
- Learning -> finding LoT expressions that match requirements

Conclusions

We saw a conceptual picture of the mind

It relates to various empirical claims

We want to make it more formal / quantitative

- Computational modelling!

Next session: Formal tools

Part I	**Introduction: On the very idea of an LoT**
Part II	Technical background
Part III	Bayesian program induction (LOTlib3)
Part IV	Case studies
Part V	Summary & Future prospects

Questions?

If there’s time left…

Action - Fodor's LoT (p.28-31)

Claim: Organisms have a representational system

1. The agent finds themselves in situation S
2. The agent *believes* that (in S) they can only do B_1, B_2, …, B_n
3. The probable consequence of performing each are predicted
   - i.e., agent computes a set of hypotheticals of roughly the form:
     "If B_i is performed in S, then, with a certain probability, C_i"
4. A preference ordering is assigned to the consequences.
5. Action is determined based on preferences and probabilities assigned.

'The notion that the agent can represent to himself salient aspects of the situations in which he finds himself presupposes that such familiar semantic properties as truth and reference are exhibited by formulae in the representational system' (p.32)

Learning

Claim: Organisms have a representational system

Experiences of *x* which are *F* cause the conclusion that 'all *x*s are *F*'

- Rather than some other property *G*!
- E.g., why do observations of white swans (rather than brown ducks) lead to the conclusion that 'All swans are white'?

Fodor thinks there's only one story:

1. The organism represents the relevant experiences **as being of** *x*s which are *F*.
2. One of the hypotheses that the organism entertains is that "all *x*s are *F*".
3. The organism employs a rule that says that observations of *x*s that are *F* is ground for the belief that all *x*s are *F*.

Perception

Claim: Organisms have a representational system

1. The organism somehow infers a task-relevant environment description *from* a physical description
   E.g., sensorial input –> 'it's time for tea'
2. Perception typically involves hypothesis formation and confirmation
   Inference to the best explanation
3. There is typically no *intrinsic* conceptual connection between sensorial input and description
4. The only plausible solution to appeals to the computational capacities of the organism

The big argument in the LoT (p.27)

1. The only psychological models of cognitive processes that seem even remotely plausible represent such processes as computation.
2. Computation presupposes a medium of computation: a representational system.

→ Today's topic!

3. Remotely plausible theories are better than no theories at all. → Self evident!
4. We are thus provisionally committed to attributing a representational system to organisms. 'Provisionally committed' means: committed insofar as we attribute cognitive processes to organisms and insofar as we take seriously such theories of these processes as are currently available. → Follows from 1-3
5. It is a reasonable research *goal* to try to characterize the representational system to which we thus find ourselves provisionally committed.
6. It is a reasonable research *strategy* to try to infer this characterization from the details of such psychological theories as seem likely to prove true.
7. This strategy may actually work: It is possible to exhibit specimen inferences along the lines of item 6 which, if not precisely apodictic, have at least an air of prima facie plausibility.

→ To justify in practice