# What is an ontology

# Kinds of knowledge

Fido is black ← synthetic (assertional)

logical → Fido is black or Fido is not black ← analytic

terminological → If Jack is a bachelor, then he is not married ← analytic

Terminological knowledge is about relationships between *terms* and *concepts*

# Ontology and (natural language) semantics

- Strictly intertwined: ontology is about ***what there is***, semantics is about ***referring*** to what there is...
- Structural semantics vs. ***referential semantics***
- Different aspects of language, different roles of ontology
  - Language connectives (conjunctions, conditionals...)
  - Primitive sentences (predication)
  - Quantifiers and modifiers
  - Prepositions
  - Nouns and verbs
  - Discourse structure

**Increasing ontological commitment** (↓)

# Ontological commitment

- Every natural language (or maybe every contextualized sentence) *commits* to some ontology (i.e., makes assumptions on *what there is*), in two ways:
  - Through a *closed* system of grammatical features
  - Through an *open* system of lexemes
- "Ontological semantics" [Nirenburg & Raskin 2004]: the semantics is driven by an ontology.
  - Practical role of ontologies for NLP systems
- Every organization, every computer system
  - Adopts a certain lexicon, to which an ***intended semantics*** is ascribed.
  - Makes (implicit) ontologic assumptions

# What kinds of commitment?

- Commitment to ***existence***:
  - Quine: every (logical) theory commits to the class of entities it ***quantifies on***.
  - Problems:
    - Should every common noun correspond to an ontological category?
      - **Questionable entities**: Events, features, qualities, fictional characters...
    - Should different linguistic behaviors mark/reflect different ontological categories?
- Commitment to ***meaning***:
  - Problem: capturing ***meaning postulates***
- Ontologies are a way to specify *both* commitments.

# Philosophical ontologies

- ***Ontology***: the philosophical discipline
  - Study of ***what there is*** (being qua being...)
    ...a liberal reinterpretation for computer science:

    **content *qua* content**, *independently of the way it is represented*
  - Study of the ***nature*** and ***structure*** of “reality”
- ***A (philosophical) ontology:*** a structured system of entities assumed to exists, organized in categories and relations.

# Computational ontologies

Specific (theoretical or computational) artifacts
expressing the ***intended meaning*** of a ***vocabulary***
in terms of ***primitive*** categories and relations describing
the ***nature*** and ***structure*** of a ***domain of discourse***

**...in order to account for the competent use of vocabulary in real situations!**

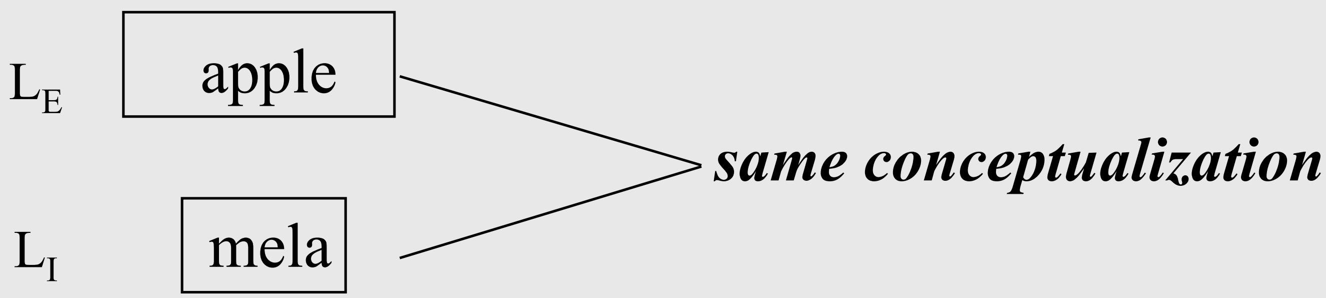
Gruber: "Explicit and formal specifications of a ***conceptualization***"

Computational ontologies, in the way they evolved, unavoidably mix
together philosophical, cognitive, and linguistic aspects.
Ignoring this ***intrinsic interdisciplinary nature***
makes them almost **useless**.

# What is a conceptualization

- Formal structure of (a piece of) reality *as perceived and organized by an agent*, *independently* of:
  - the ***vocabulary*** used
  - the actual occurence of a specific ***situation***

- Different situations involving same objects, described by different vocabularies, may share the same conceptualization.



# What is a *conceptualization?* A cognitive approach

- Humans isolate ***relevant invariances*** from physical reality (quality distributions) on the basis of:
  - Perception (as resulting from evolution)
  - Cognition and cultural experience (driven by actual needs)
  - (Language)

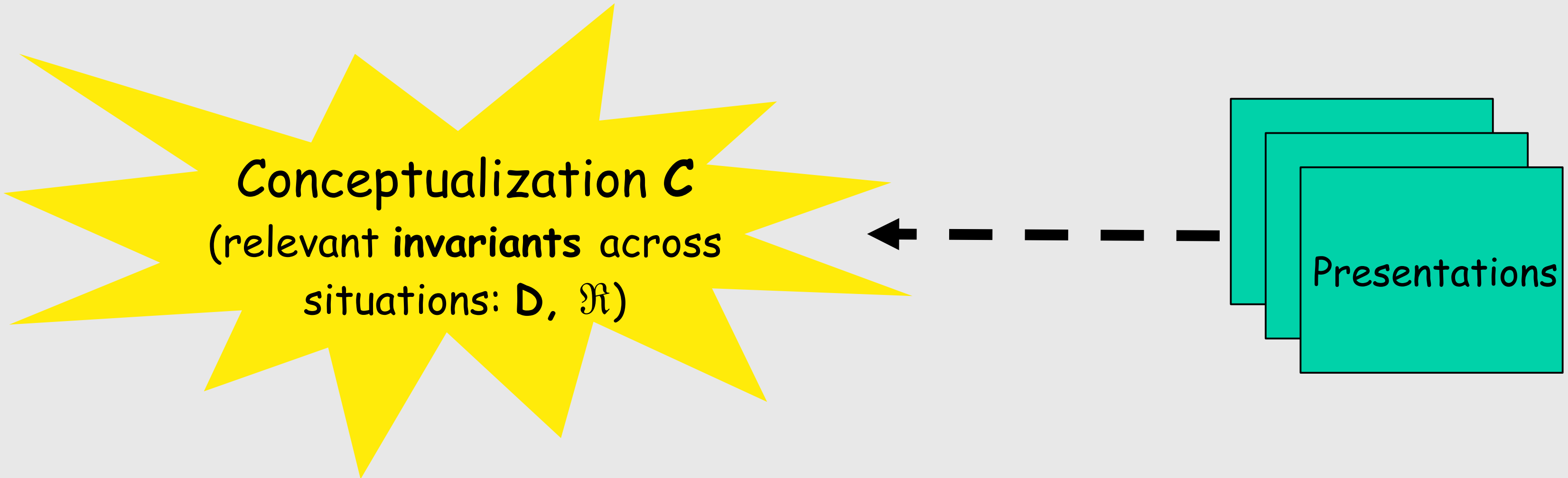
- ***presentation***: atomic event corresponding to the perception of an external phenomenon occurring in a certain region of space (the *presentation space*).

- ***Presentation pattern*** (or *input pattern*): a pattern of ***atomic stimuli*** each associated to an atomic region of the presentation space. (Each presentation tessellates its presentation space in a sum of atomic regions, depending on the granularity of the sensory system).

- Each atomic stimulus consists of a bundle of ***sensory quality values*** (qualia) related to an atomic region of timespace (e.g., *there is red, here; it is soft and white, here*).

- Domain elements corresponds to invariants ***within and across*** presentation patterns

# From experience to *conceptualization*

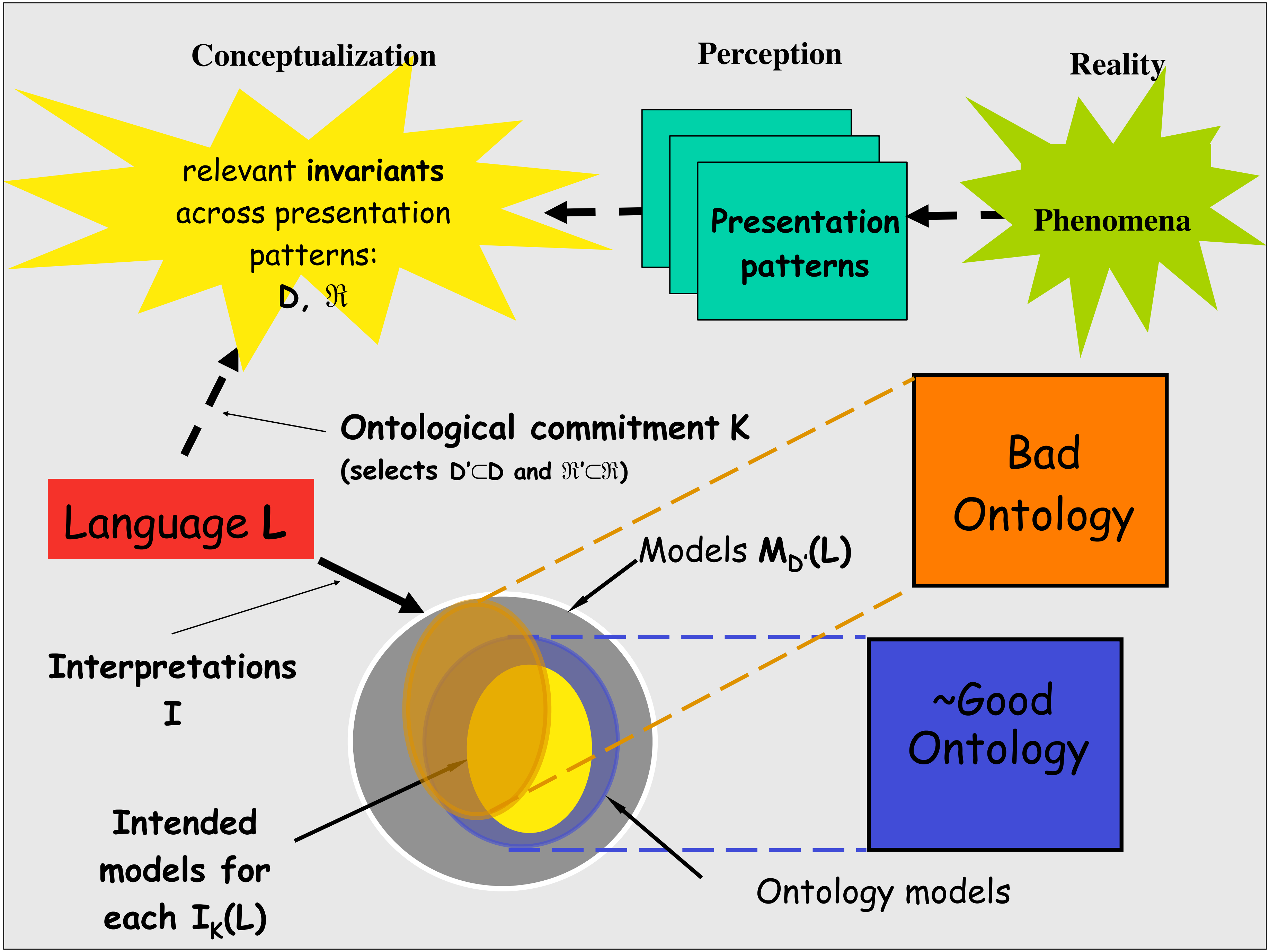



**D** : cognitive domain

$\Re$ : set of *conceptual relations* on elements of D

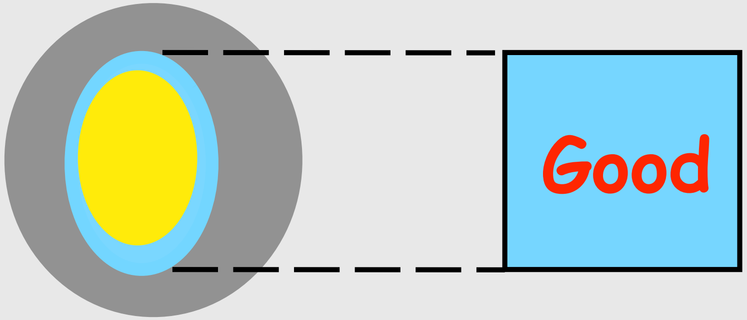
# The basic ingredients of a conceptualization
## (simplified view)

- ***cognitive objects (and events)*:** mappings from (sequences of) presentation patterns into their ***parts***
  - for every presentation, such parts constitute the ***perceptual reification*** of the object.
  - multiple objects in a single presentation: equivalence relationship among parts based on ***unity criteria***
- ***concepts and conceptual relations*:** functions from (sequences of) presentation patterns into ***sets of (tuples of) cognitive objects***
  - if the value of such function (the concept's ***extension***) is not an empty set, the correponding perceptual state is a (positive) ***example*** of the given concept
  - ***Rigid concepts*:** same extension for all presentation patterns (possible worlds)

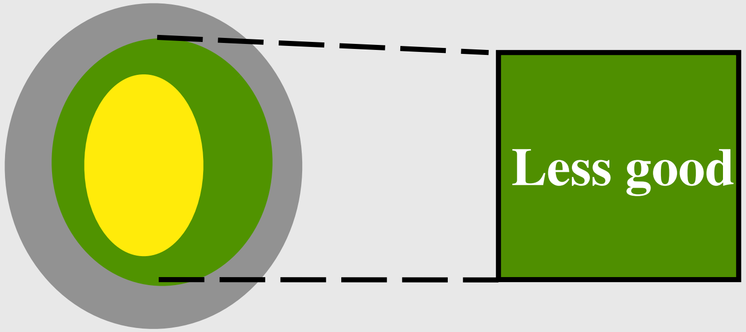


Conceptualization
Perception
Reality
relevant invariants across presentation patterns: D, ℜ
Presentation patterns
Phenomena
Ontological commitment K
(selects D'⊂D and ℜ'⊂ℜ)
Language L
Bad Ontology
Models $M_{D'}(L)$
Interpretations I
~Good Ontology
Intended models for each $I_K(L)$
Ontology models

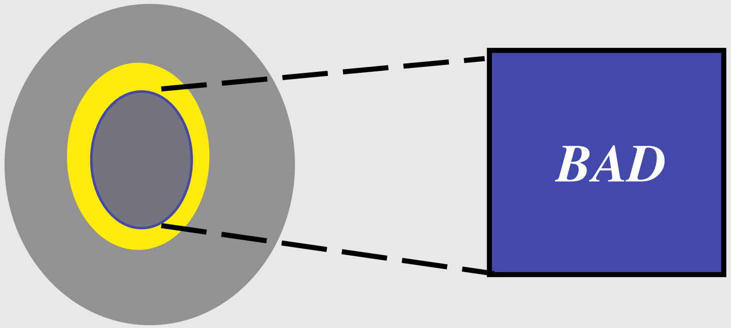
# Ontology Quality: Precision and Correctness



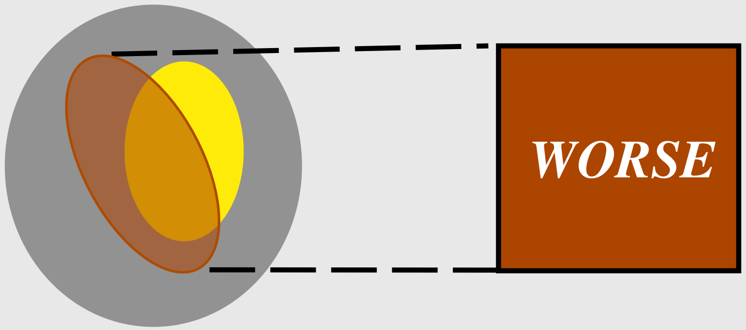

High precision, max correctness




Low precision, max correctness




Max precision, low correctness




Low precision, low correctness

# Levels of Ontological Precision

tennis
football
game
field game
court game
athletic game
outdoor game

Glossary

Catalog

game
  athletic game
    court game
      tennis
    outdoor game
      field game
        football

Taxonomy

game
NT athletic game
  NT court game
    RT court
    NT tennis
      RT double fault

Thesaurus

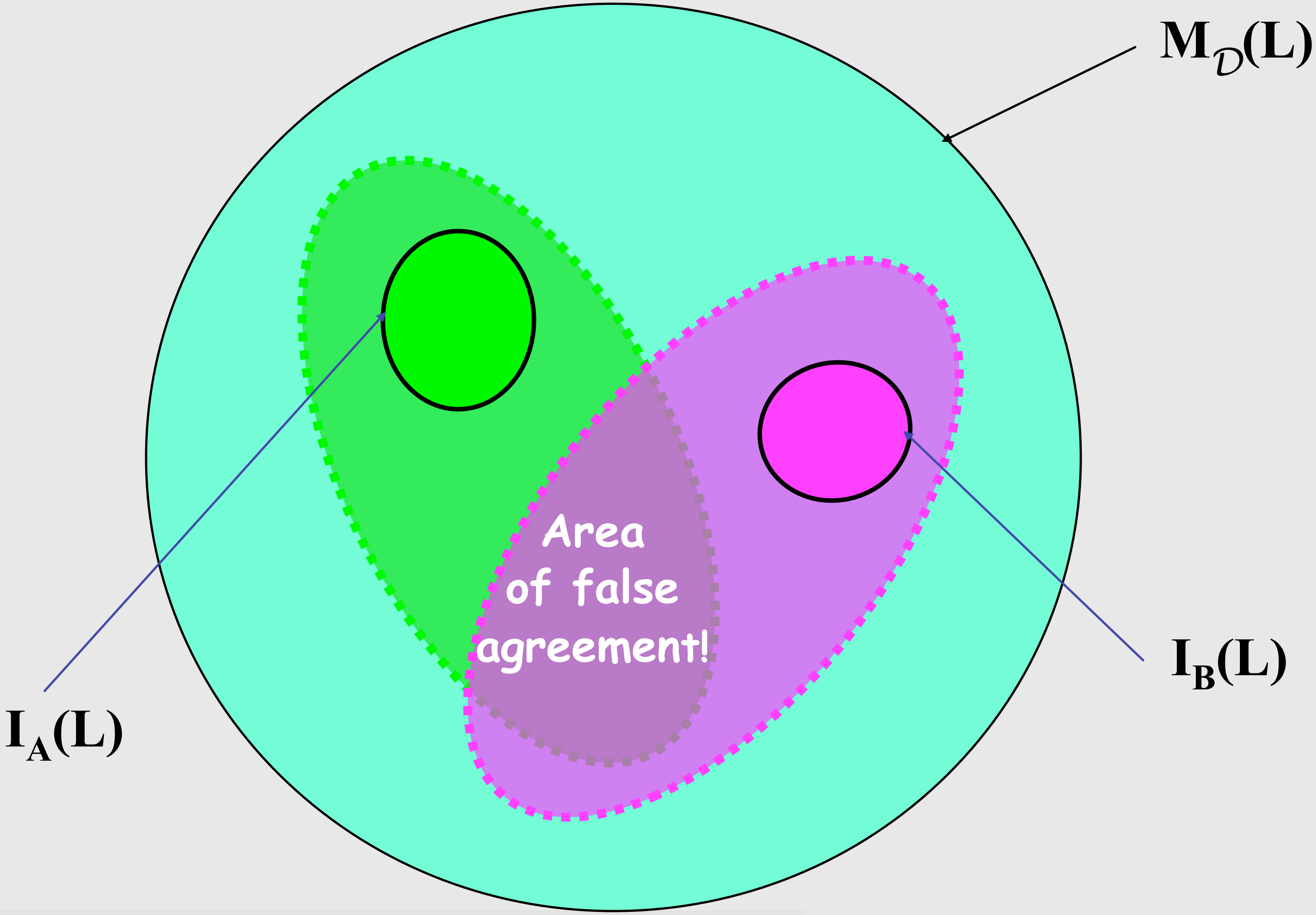
game(x) → activity(x)
athletic game(x) → game(x)
court game(x) ↔ athletic game(x) ∧ ∃y. played_in(x,y) ∧ court(y)
tennis(x) → court game(x)
double fault(x) → fault(x) ∧ ∃y. part_of(x,y) ∧ tennis(y)

Axiomatic theory

DB/OO scheme

Ontological precision →

# Why precision is important



# When precision is not enough

Only one binary predicate in the language: on

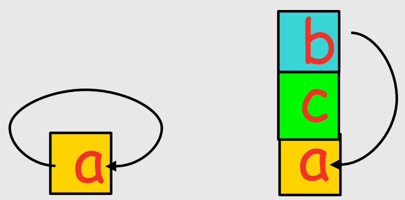
Only three blocks in the domain: a, b, c.

Axioms (for all x,y,z):

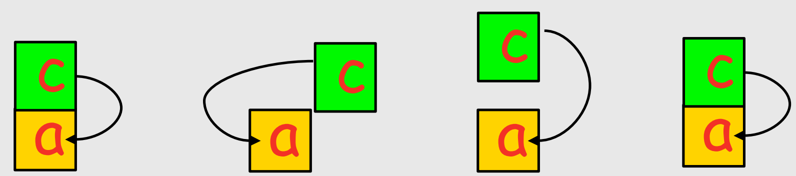
$$on(x,y) \rightarrow \neg on(y,x)$$

$$on(x,y) \rightarrow \neg \exists z\ (on(x,z) \wedge on(z,y))$$

**Non-intended models are excluded, but the rules for the competent usage of on in different situations are not captured.**




Excluded conceptualizations




Indistinguishable conceptualizations

# The reasons for ontology inaccuracy

- In general, a single intended ***model*** may not discriminate between positive and negative ***examples*** because of a ***mismatch*** between*:*
  - Cognitive domain and domain of discourse: lack of ***entities***
  - Conceptual relations and ontology relations: lack of ***primitives***

- Capturing all intended models is not sufficient for a "perfect" ontology

  *Precision*: non-intended *models* are excluded

  *Accuracy*: negative *examples* are excluded

# When is a precise and accurate ontology useful?

1. When *subtle distinctions* are important
2. When *recognizing disagreement* is important
3. When *general abstractions* are important
4. When *careful explanation and justification* of ontological commitment is important
5. When ***mutual understanding*** is more important than interoperability.

# Kinds of ontology change
## (to be suitably encoded in versioning systems!)

- Reality changes
  - Observed phenomena
- Perception system changes
  - Observed qualities (different qualia)
  - Space/time granularity
  - Quality space granularity
- Conceptualization changes
  - Changes in cognitive domain
  - Changes in conceptual relations
    - metaproperties like rigidity contribute to characterize them (OntoClean assumptions reflect a particular conceptualization)
- Logical characterization changes
  - Domain
  - Vocabulary
  - Axiomatization (Correctness and Precision)
  - Accuracy

# A quantitative metric for ontology correctness and precision

- Assumption: finite **D**, finite **W** *(examples)*

- Correctness = $card(I_k \cap O_k)/card(I_k)$
- Precision = $card(I_k \cap O_k)/card(O_k)$

# Measuring ontological accuracy
## (wrt benchmark examples)

- *Anomalous intended models* (set $A_k$): those that collapse intended and non-intended situations

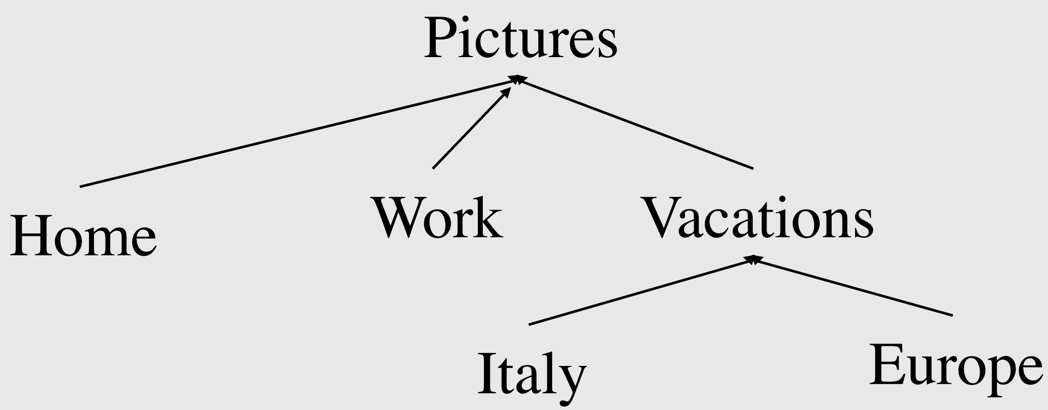
$$\text{Accuracy} = (\text{card}(I_k)-\text{card}(A_k))/\text{card}(I_k)$$

# Ontologies vs. classifications

- Classifications focus on:
  - ***access***, based on pre-determined criteria (encoded by ***syntactic keys***)

- Ontologies focus on:
  - ***Meaning*** of terms
  - ***Nature*** and ***structure*** of a domain

# A simple classification




What's the ***meaning*** of these terms?

What's the meaning of arcs?

# Ontologies vs. Knowledge Bases

- Knowledge base
  - Assertional component
    - **reflects *specific (epistemic) states of affairs***
    - designed for ***problem-solving***
  - Terminological component (*ontology*)
    - ***independent*** of particular ***states of affairs***
    - Designed to support ***terminological services***

**Ontological formulas are (assumed to be)**
**invariant, necessary information**

# Ontologies vs. Database Schemas

- Database schemas:
  - Constraints focus on ***data integrity*** (and on decoupling query language from data language)
  - Relationships and attribute values out of the DoD
  - Typically ***non-executable***

- Ontologies:
  - Constraints focus on ***intended meaning***
  - Relationships and attribute values first class citizens
  - Typically ***executable***

# Role of ontologies in information architecture

**(thanks to Dagobert Soergel)**

- ***Relate concepts to terms***. Clarify their meaning by providing a system of definitions.
- Provide a ***semantic road map*** and common conceptual reference tool across different disciplines, languages, and cultures
  - Make medical concepts clear to social science researchers and vice versa…
- Improve ***communication***. Support ***learning*** by helping the learner ask the right questions
- Support ***information retrieval*** and analysis
- Support the compilation and use of ***statistics***
- Support meaningful, well-structured ***display of information***.
- Support ***multilinguality*** and automated language processing
- Support ***reasoning***.

# A single, imperialistic ontology?

- An ontology is first of all ***for understanding each other***
  - ...among people, first of all!
  - not necessarily for thinking in the same way

- A single ontology for multiple applications ***is not necessary***
  - Different applications using different ontologies can co-exist and co-operate (not necessarily inter-operate)
  - ...if linked (and compared) together **by means of a general enough basic categories and relations (*primitives*)**.

- If basic assumptions are not made explicit, any imposed, common ontology risks to be
  - seriously mis-used or misunderstood
  - opaque with respect to other ontologies

# Formalizing conceptualizations

# Representing Intensional Relations

ordinary (extensional) relations are defined on a ***domain*** D:

$r_1 \subseteq D$ $\quad r_2 \subseteq D \times D$ $\quad r_n \subseteq D^n$ $\quad r_n \in 2^{D^n}$

intensional relations are defined on a ***domain space*** <D, W>

$\rho_n : W \rightarrow 2^{D^n}$ *(Carnap, Montague)*

**But what are possible worlds?**
**What are the elements of a domain of discourse?**

A ***conceptualization*** for D is a tuple **C** = <D, W, ℜ>, where ℜ is a set of conceptual relations on <D, W>

A ***model*** for a language **L** with vocabulary V is a structure

<**S**, I>, where **S** = <D, **R**> is a ***world structure*** and I: V→D∪**R** is the usual interpretation function.

A model fixes a particular extensional interpretation of the language. Analogously, we can fix an ***intensional*** interpretation by means of a structure

<**C**, ℑ>, where **C** = <D, W, ℜ> is a conceptualization and ℑ: V→D∪ℜ is an *intensional interpretation function*.

We call such a structure **K=<C**, ℑ> an ***ontological commitment*** for **L**.
**L** ***commits*** to **C** by means of **K**.
**C** is the ***underlying conceptualization*** of **K**.

# Possible worlds as presentation patterns (or sensory states)

**Presentation pattern**: unique (maximal) pattern of *qualia* ascribed to a spatiotemporal region tessellated at a certain granularity

...This corresponds to the notion of *state* for a sensory system (maximal combination of values for sensory variables)

**Possible worlds are (for our purposes)**
***sensory states***
**(or if you prefer, [maximal] sensory *situations*)**

# Possible worlds vs. models

- Models are combinations of ***meaning assigments***
- Worlds are - so to speak - combinations of ***things****!*

- Consider the ***model*** where there is a bachelor which is married.
- Is there a ***world*** where bachelors are not married?
  - ...in this world *bachelor* and *married* would have a different ***meaning***!!

# Situations vs. possible worlds

- Situations *hold* (in a world): they are states ***of*** worlds (i.e., *properties* of worlds – universals)
- Possible worlds (strictly speaking) do not *hold* (they are *particulars*)
- Situations are *partial* states of affairs
- Worlds are described by *maximal* states of affairs (sometimes they are confused with them)