Ontologies

A Beginner's Guide *for* Coffee-House Philosophers

Ontologies

Organizing Knowledge using Categories and Relations

Ontologies in Computer Science has a very specific function and interpretation:

1. **"Formal Specification of a Conceptualization"**

   This (vague) definition is due to **Thomas Gruber**, introduced in the 1990s

   An Ontology in this sense is the **explicit** (*logical, axiomatic*) representation of a conceptualization (our domain knowledge for a given problem)

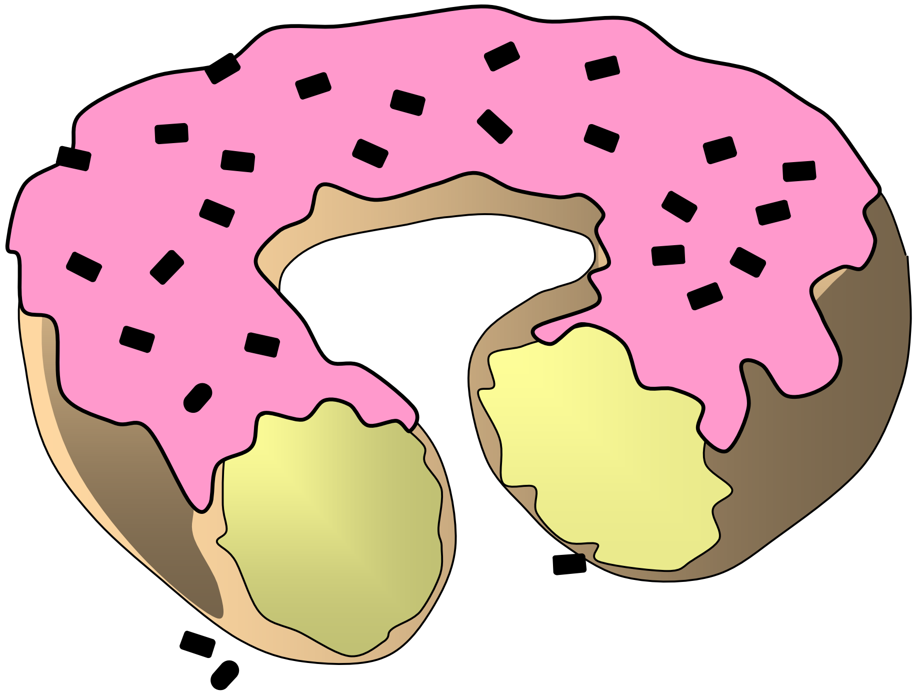
   The ontology allows us to reason within a domain using an explicit formalism

2. **"an engineering artifact, constituted by a specific vocabulary"**

   This definition, introduced by Guarino (1995), views an ontology as the formal definition of our logical/conceptual vocabulary for a given domain/problem.

   Guarino claims an ontology merely describes a "certain reality" (perspective)

A Simple Task (Simple enough for *Homer Simpson*)



Can you provide a good logical definition of a **DONUT**?

(One that matches all donuts but no non-donuts)

“A **DONUT** is simply an **Edible Torus**, that’s all”

(Eminent AI Philosopher)

How can a ***static structure*** keep up with our creative urge to innovate, to ***challenge expectations*** and ***stretch category boundaries***?

The essential tension in the philosophy of categories and ideas:
to impose top-down order or ***respond to bottom-up diversity***?

Ontology in Philosophy

Following **Gruber** and **Guarino**, “Ontologies” (plural) is oft mentioned in CS / AI

But the term “Ontology” has a very long history in Philosophy, dating to antiquity.

Since ancient times in **Philosophy**, **Ontology** has been a branch of **Metaphysics** that concerns itself with the nature of **Being** and **Reality**:

Ontology concerns *what exists, what doesn't exist*, *what can/cannot possibly exist*.

In **logic**, Ontology describes the set of commitments to reality made by a theory.

The foundations of Ontology (an activity, not an artifact!) lie in **Greek Philosophy**:

Parmenides: reality is limitless, eternal, uniform, unchanging (*change is illusion*)

Plato: distinction between *ideal forms* (Platonic) and *imperfect instances*

Aristotle: Ontology is the study / organization of *commonalities* among objects

Plato, Aristotle, Universals

Plato argues that even *abstract concepts* really exist, in a **Platonic** *realm of ideas*.

New ideas are not ***created/invented***, but ***discovered*** in a world of pre-existing forms

Aristotle also argues for the existence of abstract categories that serve to organize our experience of the world.

But, Aristotle suggests that these have **no independent existence** from us, but are intimately linked to the objects/entities in the world that they serve to organize.

E.g., Aristotle argues that **Properties** cannot exist independently of the types of objects they are used to qualify in the real world (e.g., Tasty and Food)

Aristotle thus has a **CS/AI-friendly view** of Ontology – a set of categorizations that organizes those objects/entities/processes in which we are interested.

Plato vs. Diogenes on Universals and Ideal Forms

Plato was discoursing on his theory of ideas and, pointing to the cups on the table before him, said while there are many cups in the world, there is only one `**idea'** of a cup, and this *cupness* precedes the existence of all particular cups.

"***I can see the cup on the table***," said Diogenes, "***but I can't see the `cupness"***'

"***That's because you have the eyes to see the cup***," said Plato, "***but***", tapping his head with his forefinger, "***you don't have the intellect to comprehend `cupness'***."

Plato argues that all real-world objects are imperfect realizations of ideal, ***Platonic forms***; these forms exist independently of their real-world instances, and understanding of the world follows from perceiving the forms behind the instances.

Aristotle argues *instead* that real-world objects conform to **generalizations**/ **categories**, but such categories are functions of real-world objects, not vice-versa.

Plato vs. Diogenes (continued)

Diogenes walked up to the table, examined a cup and, looking inside, asked, "***Is it empty?***"

Plato nodded.

"***Where is the `emptiness' which proceeds this empty cup?***" asked Diogenes.

Plato allowed himself a few moments to collect his thoughts, but Diogenes reached over and, tapping Plato's head with his finger, said "***I think you will find that the `emptiness' is here***."

<u>This anecdote reveals some key challenges for Ontological analysis</u>:

1. What can be meaningfully analyzed/reduced, where should we stop?

2. ***How authoritative is an ontology? A picture of logical reality, or a guide?***

The Ontological Argument (in Philosophy)

Is Ontology simply a description of reality, or can it possess real logical force?

Consider the medieval "<u>Ontological Argument</u>" for God's existence:

1. **(*by common definition*): God is the most perfect of beings, the greatest of all beings, the sum of all perfections.**
2. **It is more perfect to exist than not to exist. It is greater to exist in reality than merely in someone's imagination.**
3. **By (1) then, God must exist, otherwise God would be less perfect than it is possible to be.**

Cf. Avicenna, St. Anselm, Descartes

This is called the Ontological Argument because it concerns the ontological categorization of God (does God belong to the category of existing things?)

It is an *<u>A Priori</u>* argument, based on reason alone: it attempts to will God into existence by force of ontology-based reasoning.

Key Questions in Ontology (w.r.t. Comp-Sci / AI)

Some key issues in the study of Ontology, in Philosophy & Computer Science:

1. **Which entities / categories are universal?**
2. **What are the foundational categories on which an ontology should be built *(in CS/AI terms, what is the ideal Upper-Ontology Model)***
3. **How do the properties of an object relate to the object / define an object?**
4. **What properties are truly necessary/essential/definitional of a concept/ object and not simply *accidental* (or *contingent*) ?**
5. **How does one uniquely identify a category / entity (identity criteria)**
6. **When does an object go out of existence, rather than simply changing (e.g., *Lincoln's axe*, the *Ship of Theseus*, etc.)**
7. **Can an object change its ontological categorization over time?**

Ontologies versus Taxonomies

Like an Ontology, a Taxonomy is a means of dissecting reality into a organized system of related categories.

Just as "**Ontology**" is the philosophical study of "**Being**", "**Taxonomy**" is the study (i.e., *practice and science*) of **Classification**.

Since Ontology crucially depends on ***Classification***, CS/AI Ontologies naturally contain Taxonomies. That is, a Taxonomy is a key part of most Ontologies.

In Practice, a Taxonomy provides the structural Backbone of a CS/AI Ontology.

A **Taxonomy** is ***NOT*** an **Ontology**, just a structure on which an Ontology hangs.

A Taxonomy is a ***hierarchical system*** of **subsuming categories** (a **tree** or **forest**)

The *key relationship* in a Taxonomy is usually **IS-A** (***Subset-of*** and ***Instance-of***)

Taxonomies are often called **IS-A Hierarchies** in *Artificial Intelligence*

The Linnaean Taxonomy (after Carolus Linnaeus)

Common Name:	Human	Nine-Banded Armadillo
Kingdom:	Animalia	Animalia
Phylum:	Chordata	Chordata
Class:	Mammalia	Mammalia
Order:	Primata	Cingulata
Family:	Hominidae	Dasypodidae
Genus:	*Homo*	*Dasypus*
Species:	*Homo sapiens*	*Dasypus novemcinctus*

The Linnaean Taxonomy organizes all life into a series of **Taxa** (sing. **Taxon**)

From the highest (broadest) Taxon **Kingdom** to the lowest (most specific) **Species**

Above is a comparison of the Linnaean classifications of **Humans** and **Armadillos**

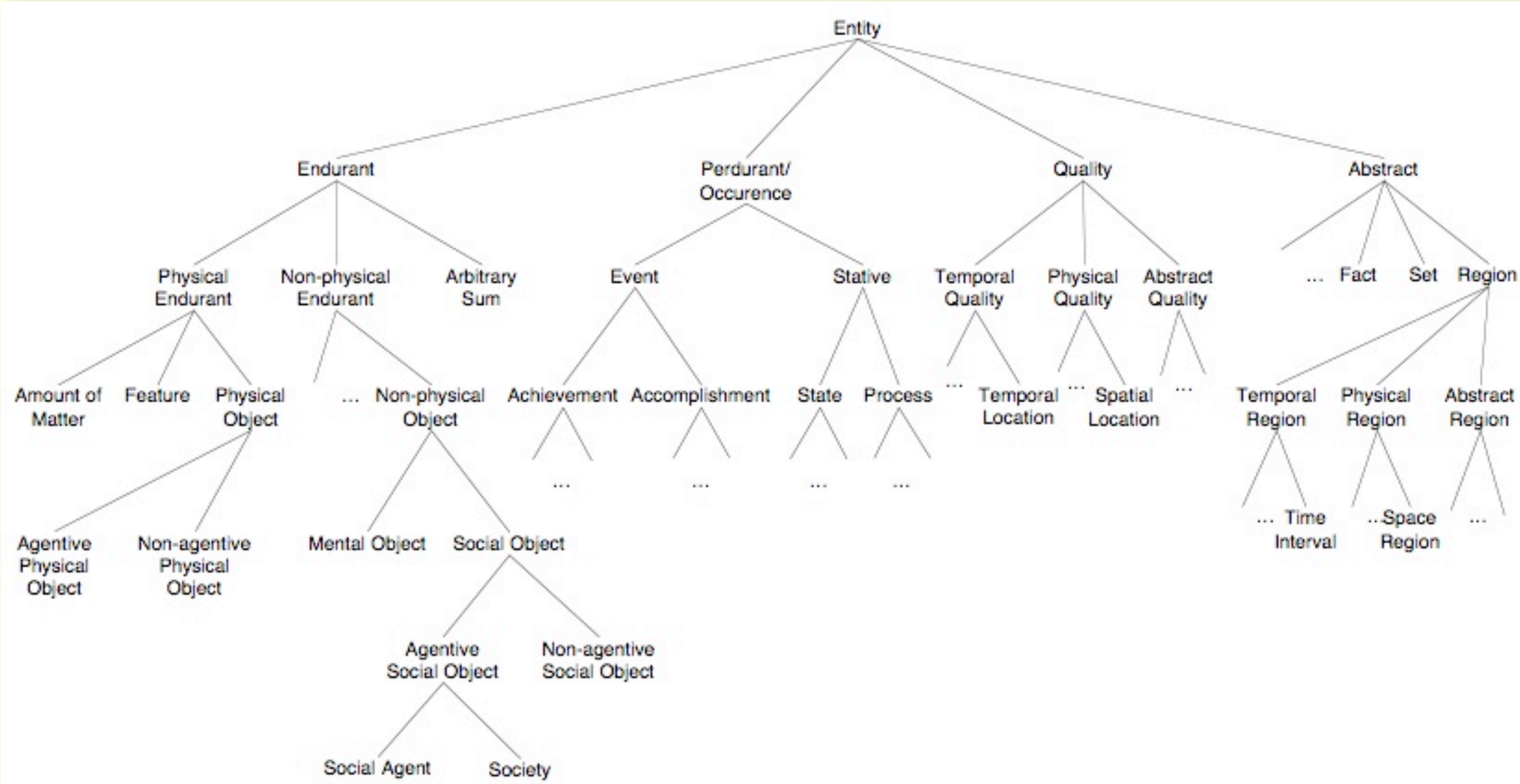
The Dewey Decimal Taxonomy

001-099	Generalities
100-199	Philosophy
200-299	Religion
300-399	Social Science
400-499	Languages
500-599	Natural Science
600-699	Applied Science
700-799	Arts and Recreation
800-899	Literature
900-999	Geography and History

Library Scientists are particularly keen on the use and study of Taxonomies, and more recently, ontologies (in connection with Semantic Mark-up Schemes)

400 Language
410 Linguistics
420 English & Old English
430 Germanic languages German
440 Romance languages French
450 Italian, Romanian languages
460 Spanish & Portuguese languages
470 Italic languages, Latin
480 Hellenic languages, Classical Greek
490 Other languages

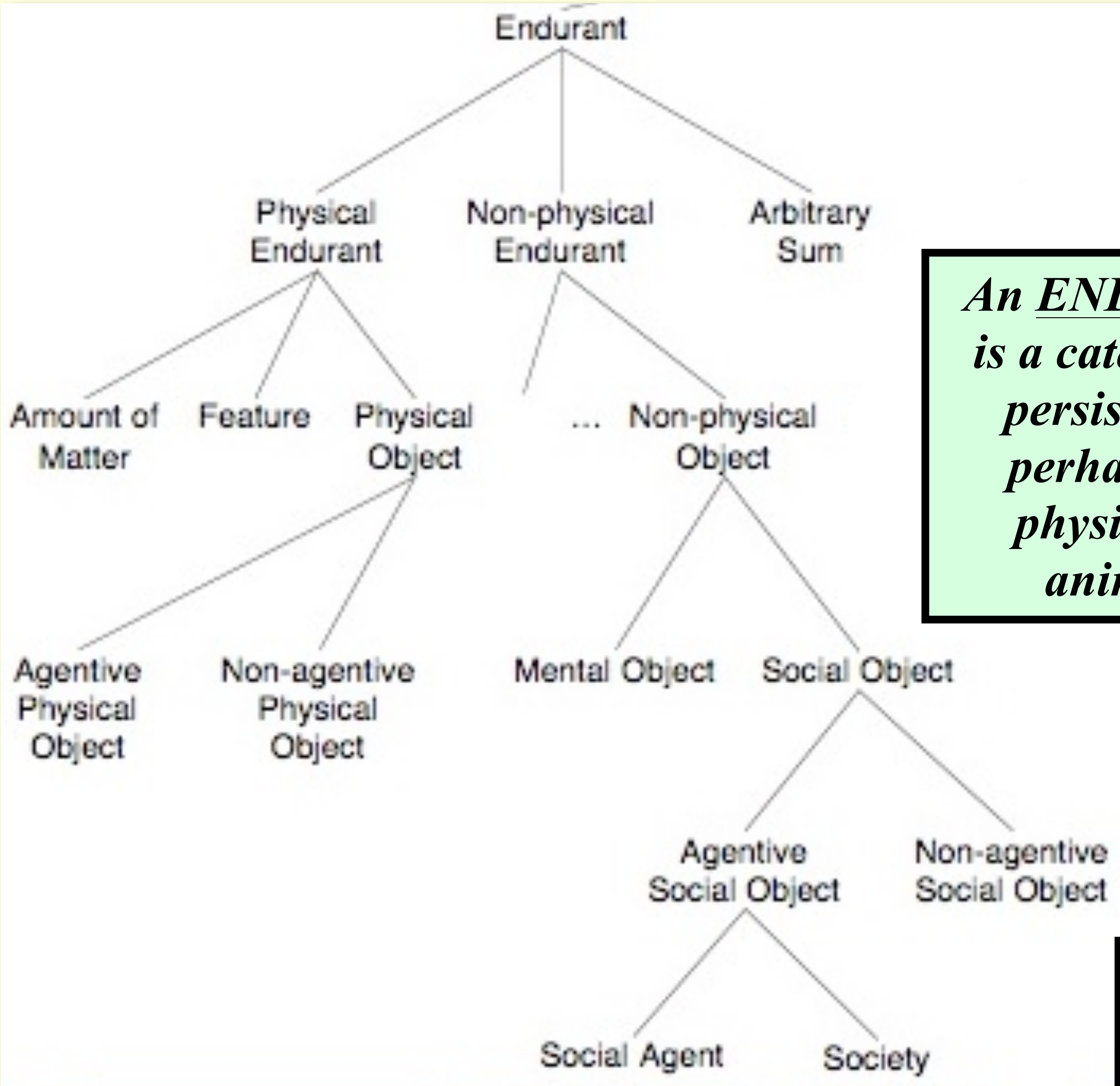
Upper Ontological Models: DOLCE



DOLCE: **D**escriptive **O**ntology for **L**inguistic and **C**ognitive **E**ngineering

[Gangemi and Guarino]

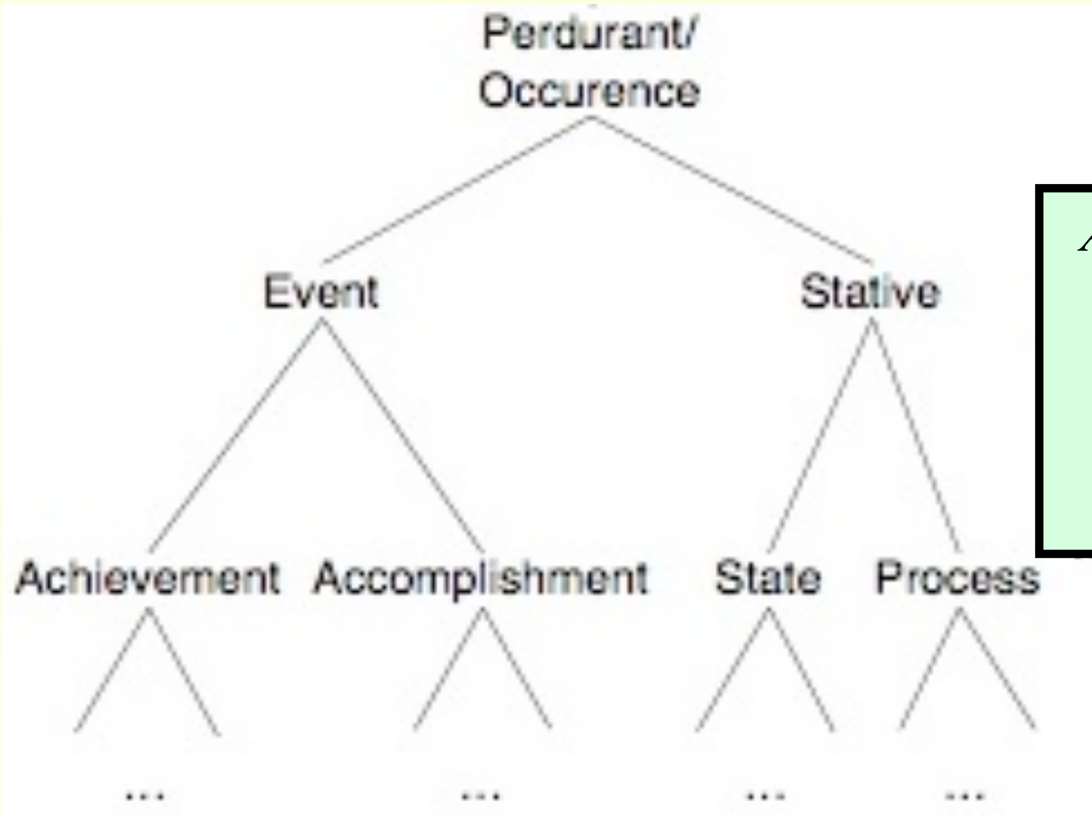
Upper Ontology: DOLCE Endurants



An ENDURANT in DOLCE is a category of entities that persist across Time (and perhaps Space), such as physical things, people, animals, even ideas!

Contrasts with a PERDURANT →

Upper Ontology: DOLCE Perdurants

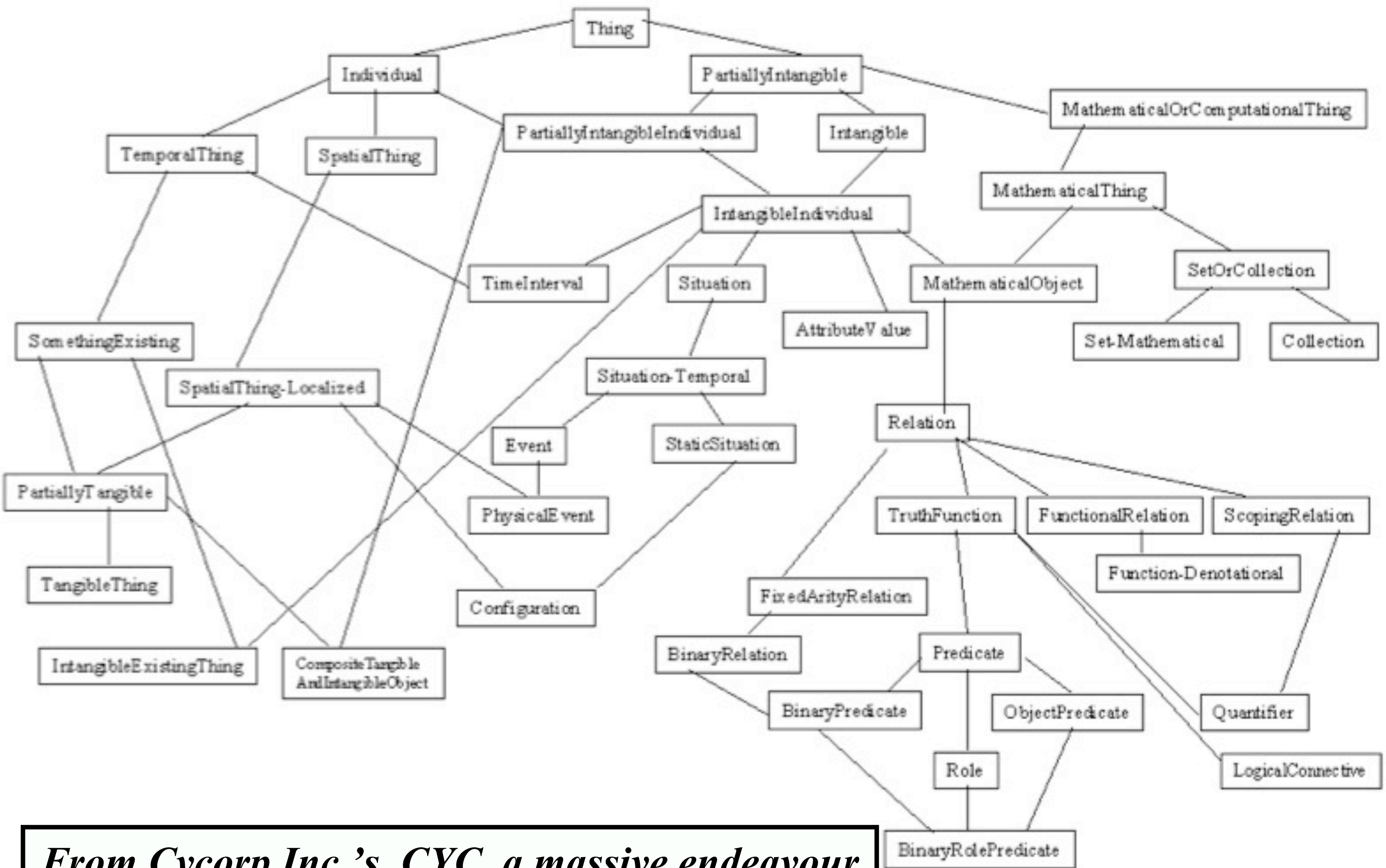


A <u>PERDURANT</u> in DOLCE is an event, process or situation that involves ENDURANTS and their goals/achievements

An Upper Ontology provides a logically and philosophically-sound discrimination of the key categories in the world at large.

The Upper Ontology thus acts as a sound foundation on which to hang more detailed, domain/application-specific categories and ideas.

Upper Ontologies: The CYC Upper-Model



From Cycorp Inc.'s CYC, a massive endeavour

Knowledge Engineering (KE) in Cyc

- Knowledge Engineering is a process of ***Ontologization*** and ***Axiomatization***:

```
(#$forAll ?PER
   (#$thereExists ?FANCLUB
     (#$implies  (#$and (#$isa ?PER FamousPerson)
                        (#$isa ?FANCLUB (#$MobFn Person))
                        (#$groupMembers ?FAN ?FANCLUB))
                 (#$feelsTowardsObject ?FAN ?PER
                                 #$Admiration #$Positive))))
```

- Axioms are associated with concepts *(collections or individuals)* in microtheories.
- Implication Axioms *(rules)* can be designated as *forward*- or *backward*- firing.

Rule-Bound Reasoning

- At the Core of CYC is an **Ontology** of Concepts (Taxonomy + Relationships) that informs and underpins all axioms in the KB.
- Unfortunately, these concept representations do not reflect current thinking in the **cognitive psychology of category structure** (e.g., radial, fuzzy, prototype-based).

For Example, consider how Cyc combines concepts for Noun-Noun compounds:

```
(#$nnRule "potato gun"
    (#$and  (#$genls :NOUN1 #$PartiallyTangible)
            (#$genls :NOUN2 #$ProjectileLauncher)
            (#$not
                (#$genls :NOUN1 #$Organism-Whole)))
    (#$isa  :NOUN
            (#$SubcollectionOfWithRelationToTypeFn
                          :NOUN2 #$launchesProjectile :NOUN1))
```

There are many problems with this account of NN-Compounding: (concepts should combine as a matter of definition and meaning; rules are easily defeated; too top-down).

Over-Specification

- Excessive (and *obsessive*) **Ontologization** can lead to **hair-splitting**.
- For example, Cyc discriminates among many different senses of "**in**" :

E.g.,

in *(full submerged)*	**– like an olive *in* a martini**
in *(partially submerged)*	**– like a toothpick *in* the olive**
in *(surrounded by)*	**– like a man *in* a field**
in *(membership)*	**– like a man *in* a club**

But strangely, not:

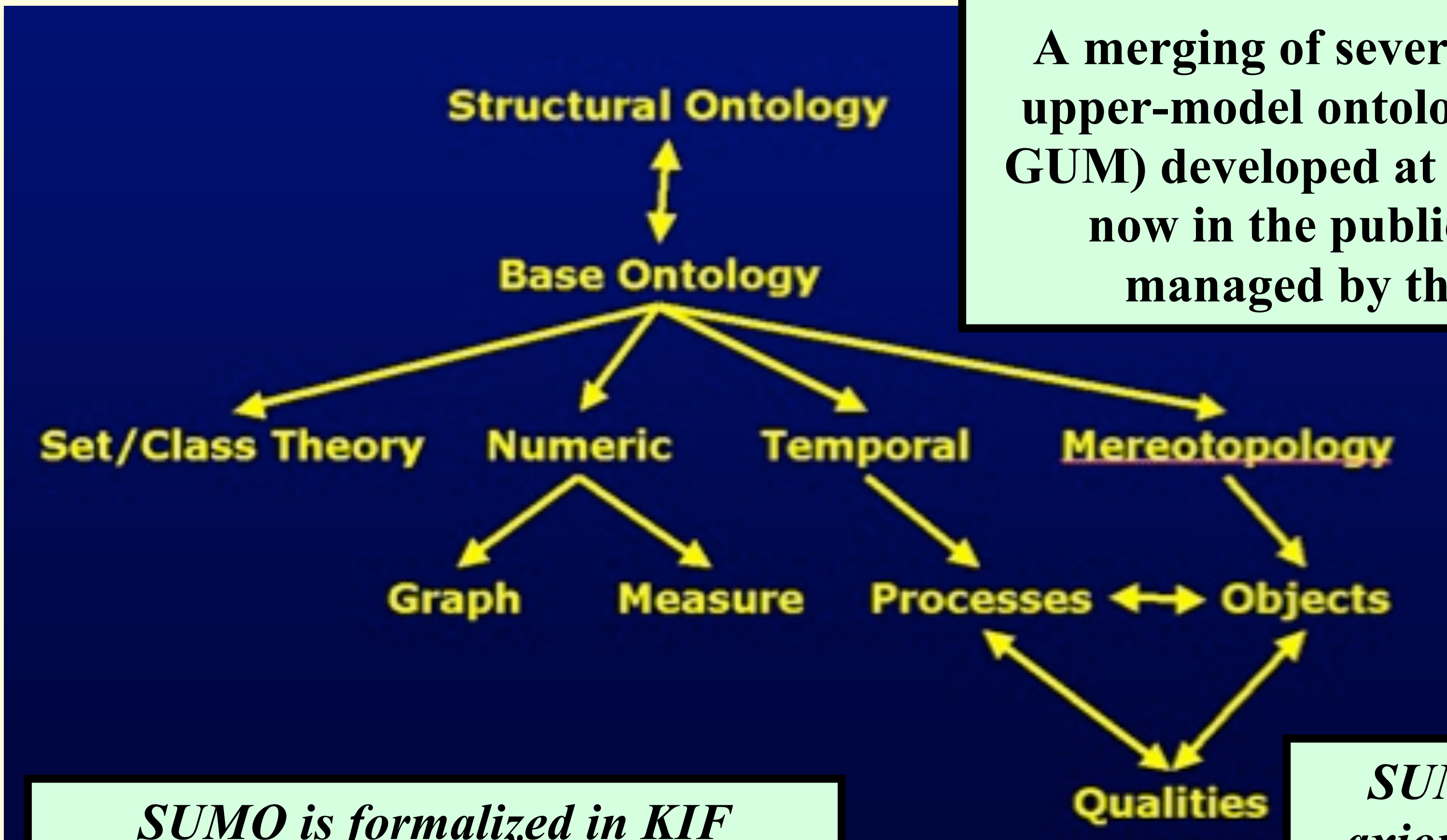
in *(abstract situation)*	**– like a woman *in* love**
in *(content area)*	**– like an academic in a research field**

These copious *(and uneven)* discriminations cause a combinatorial explosion for natural language parsing systems, *yet fail to capture the true essence of "in"*.

Upper Ontologies: The SUMO Upper-Model

Suggested Upper Merged Ontology

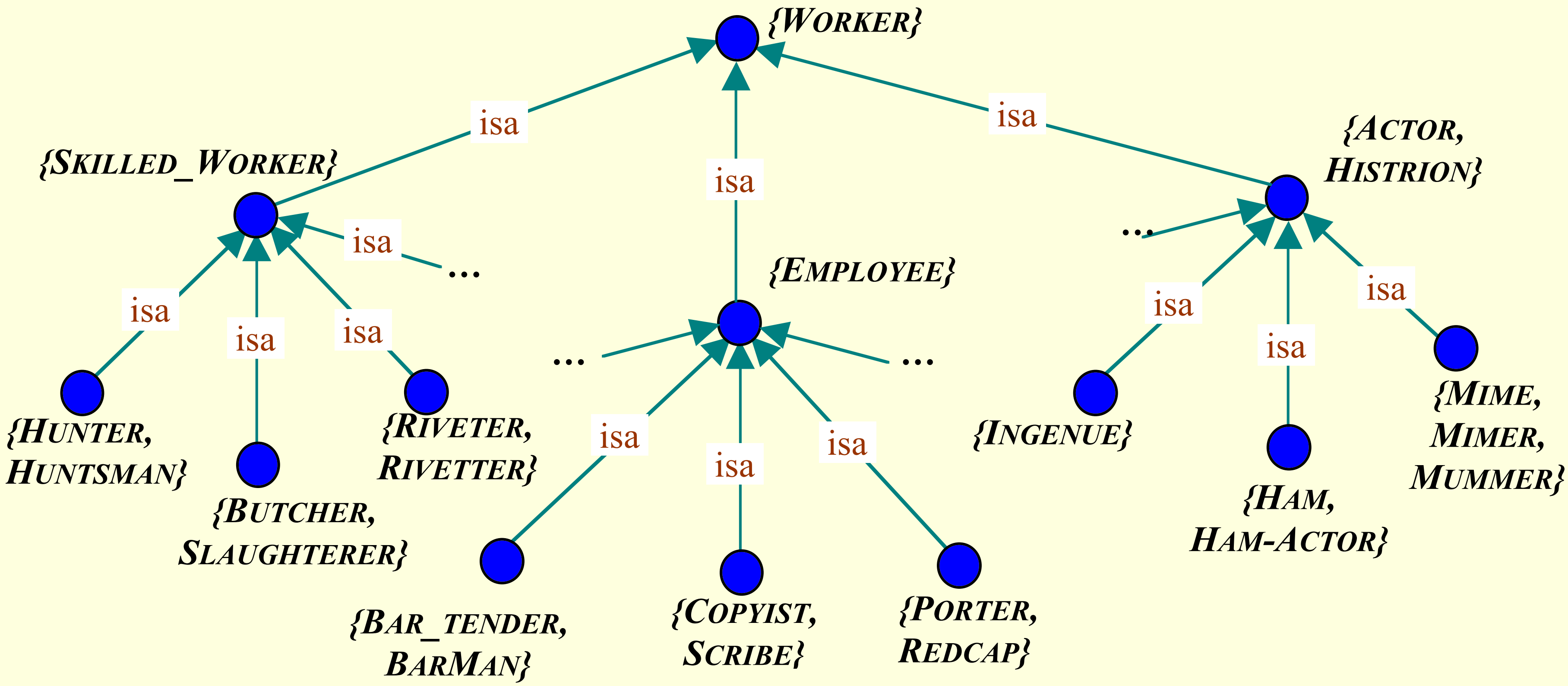
A merging of several suggested upper-model ontologies (such as GUM) developed at *Teknowledge*, now in the public domain, managed by the IEEE.



SUMO is formalized in KIF (Knowledge Interchange Format) and is used with a variety of more specific **middle ontologies**

SUMO contains axiomatizations of generic concepts like Competition, Business, etc.

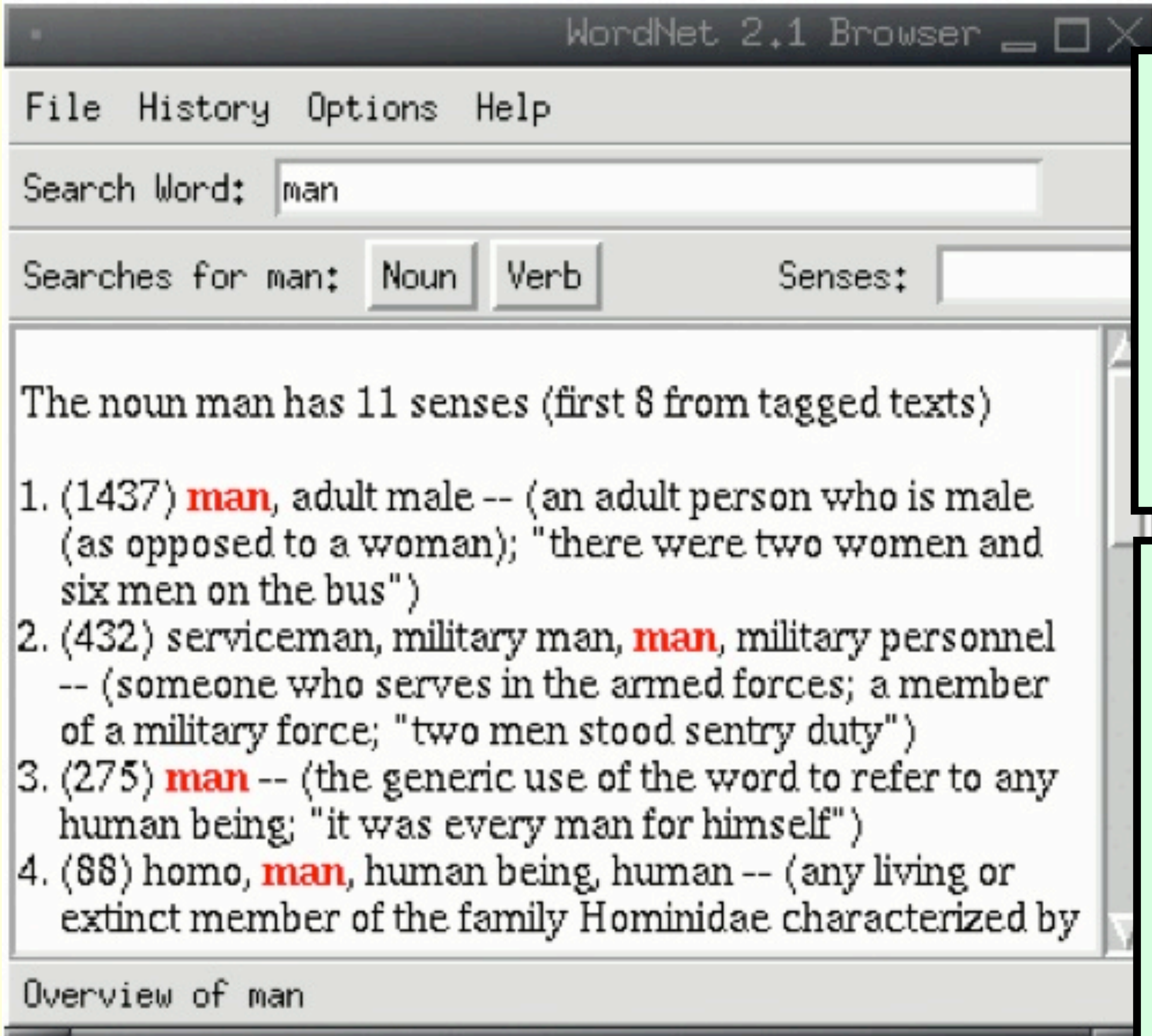
Lexical Ontologies: Princeton WordNet



Words are grouped into **Synsets** (<u>**set**</u>s of <u>**syn**</u>onyms) that *uniquely denote* <u>**senses**</u>

Synsets (denoting senses) are then structured via <u>**IS-A**</u>, <u>**Part-Whole**</u> relations

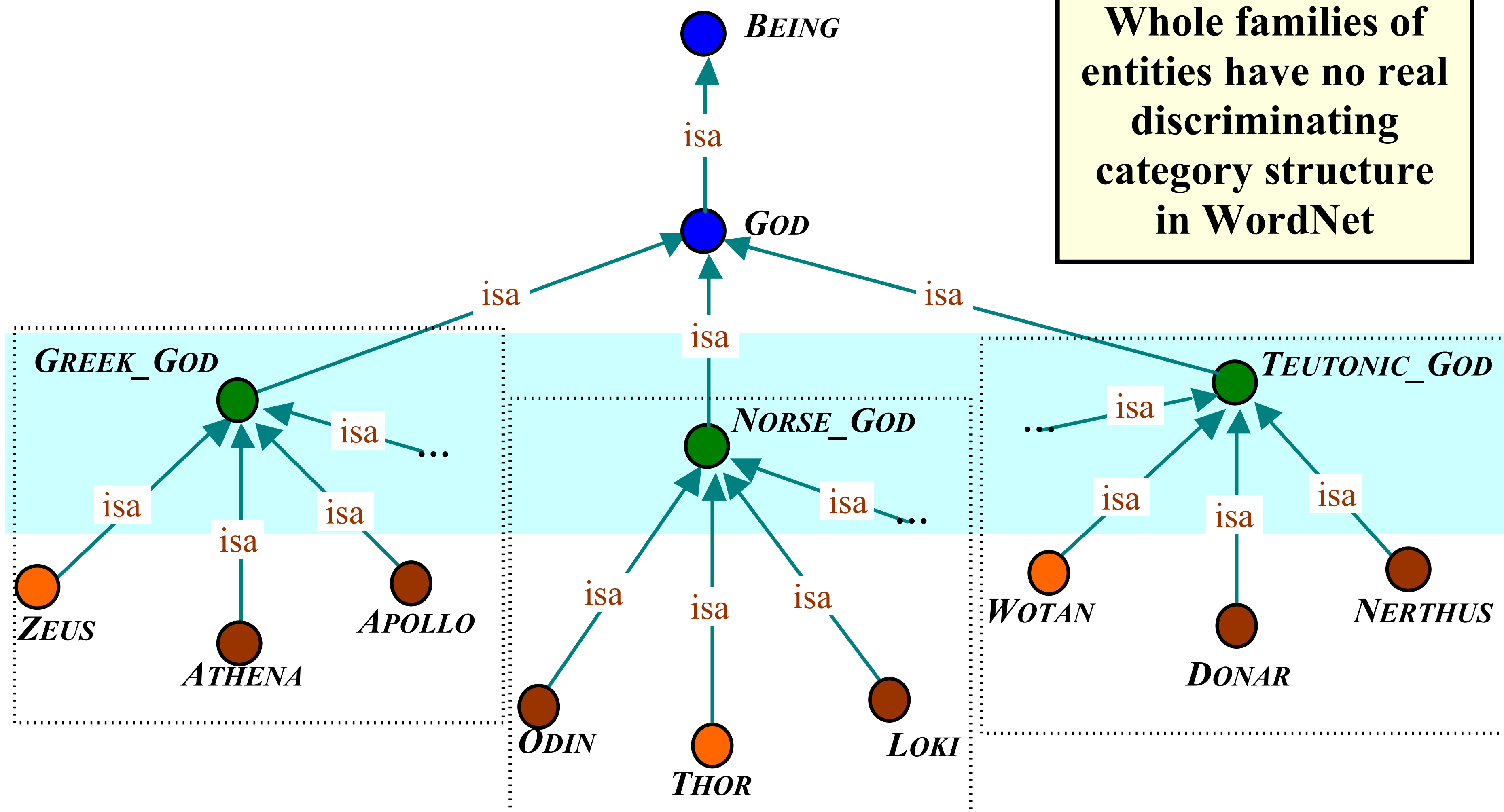
Lexical Ontologies: Accessing WordNet



WordNet is a hierarchically-structured lexical database that can be accessed on the web, or locally (via a simple lexicographer-oriented file-structure)

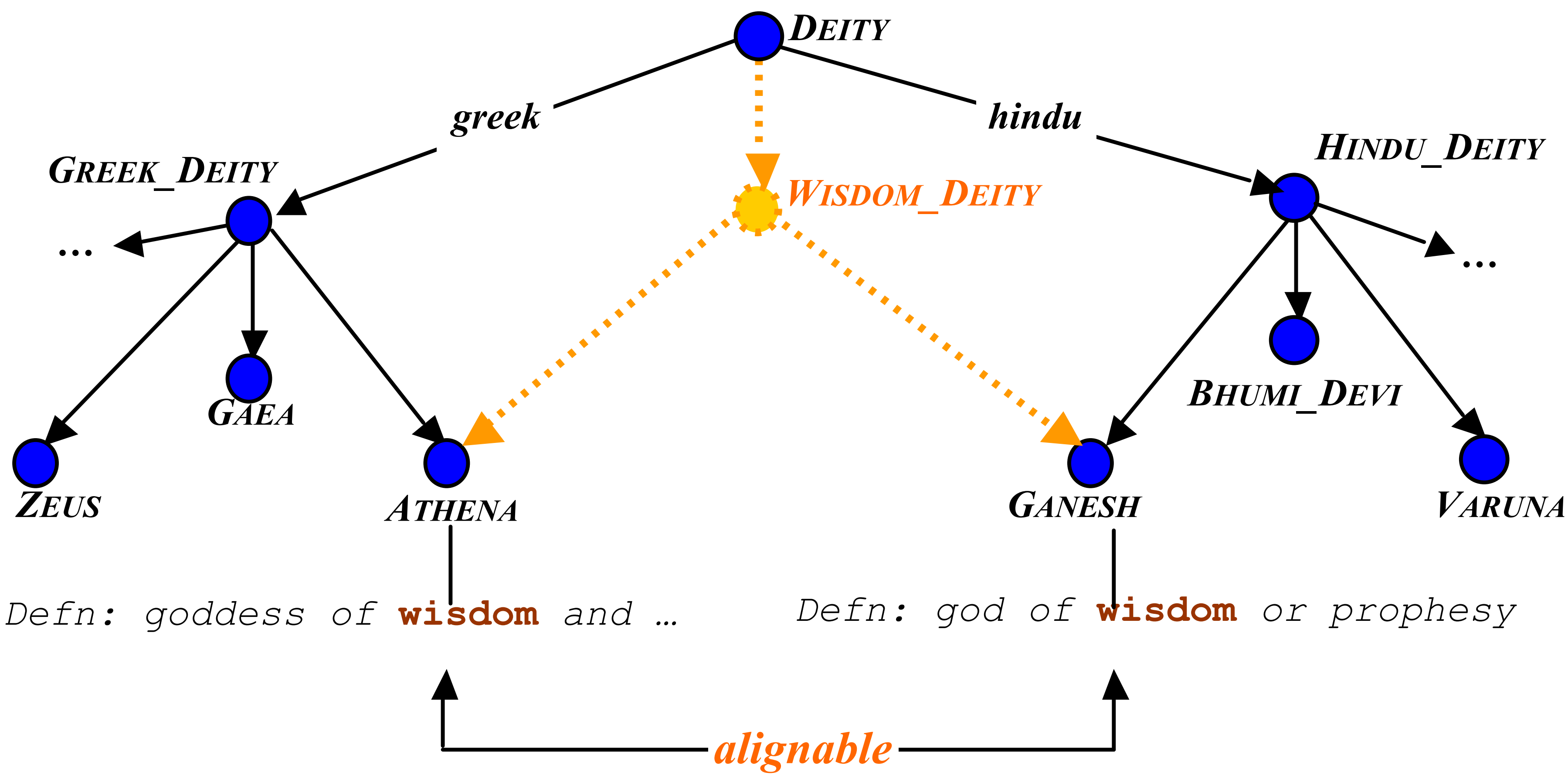
Because it is comprehensive (more that 70,000 noun definitions) and free (unencumbered), WordNet is widely used in NLP and IR applications, as a lexicon or as a source of simple world-knowledge

Taxonomic Weakness: Lack of Discriminating Structure in WordNet



Whole families of entities have no real discriminating category structure in WordNet

Automatically Enrich WordNet With Categories from Glosses



Empirical Evaluation: Analogical Retrieval with WordNet

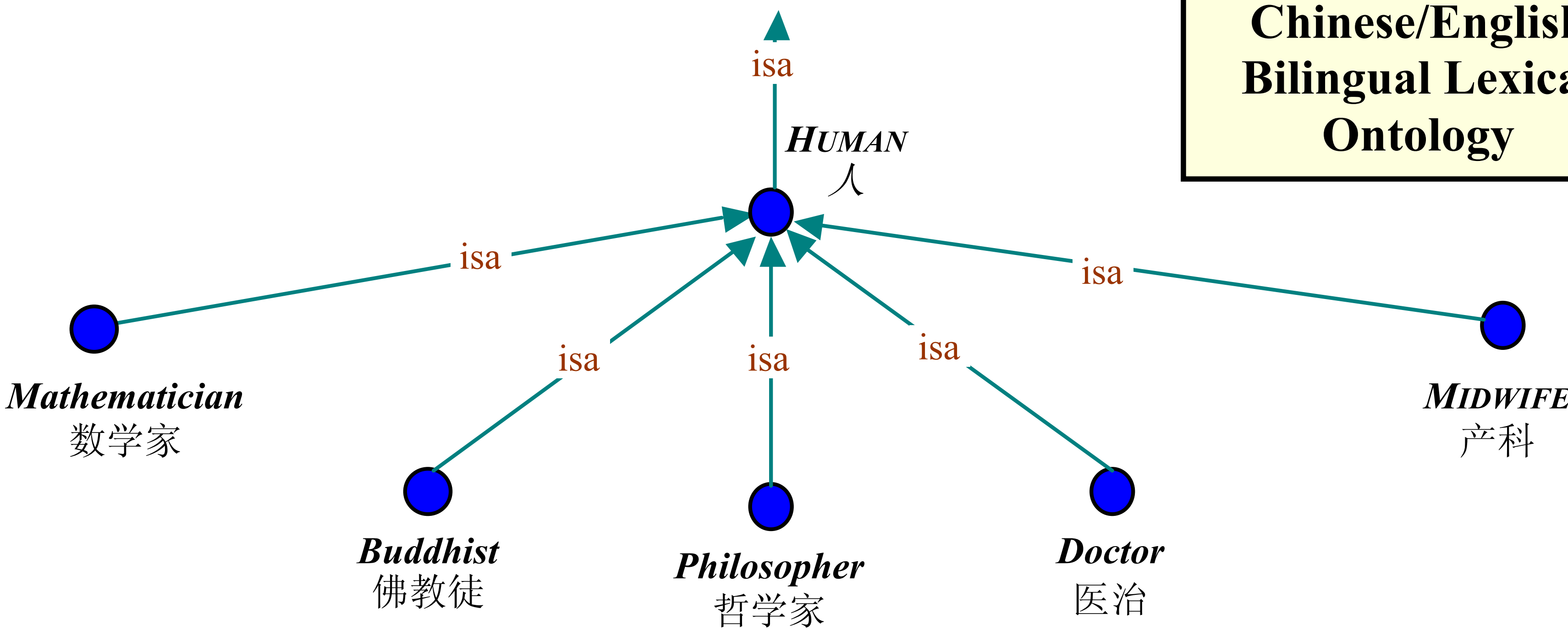
Deity to Deity Task	Precision	Recall
Static WN representations	***0.115***	***0.34***
Dynamic WN representation *(+ gloss-category reification)*	***0.935***	***0.61***

Letter to Letter Task	Precision	Recall
Static WN representations	***0.04***	***0.98***
Dynamic WN representation *(+ gloss- category reification)*	***0.96***	***0.98***

E.g., Given Zeus (Greek) retrieve Odin (Norse) and Jupiter (Roman) etc.
E.g., Given Alpha (Greek) retrieve Aleph (Hebrew)

Ontological / Taxonomic Structure in HowNet

HowNet is a Chinese/English Bilingual Lexical Ontology



It has a weaker taxonomy than WordNet but logical definitions of each class

HowNet Semantics: A Bilingual Constructivist Ontology

HowNet is an English / Chinese ontology whose entries are semantically defined. E.g.

surgeon|医生

{human|人:HostOf={Occupation}, domain={medical}},

*{**cure**|医治:agent={~}}*

…

resume | 恢复

cure | *医治*

repair | *修理*

amend | 改正

…

HowNet Semantics: Underspecified and Often Unstructured

Some lexical concepts in WordNet have more substantial logical defn.'s than others

cyclist|自行车手

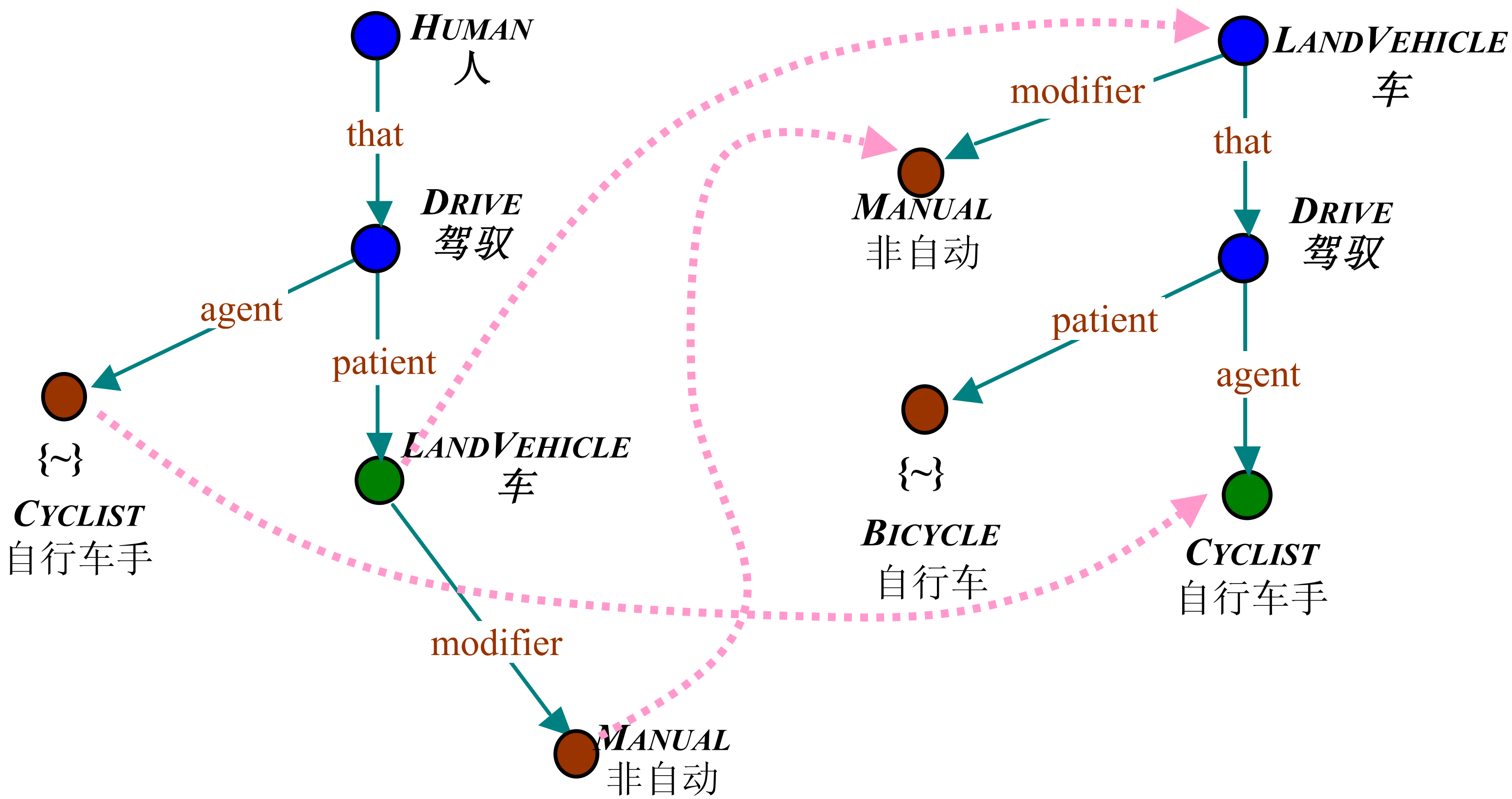
{human|人: {drive|驾驭:
agent =\{~\},
patient=\{LandVehicle|车:modifier=\{manual\}\}\}\}

bicycle|自行车

\{LandVehicle| 车: modifier=\{manual|非自动\}\}

N.B. **cyclist = (bicycle)自行车 + (person good at job)手**

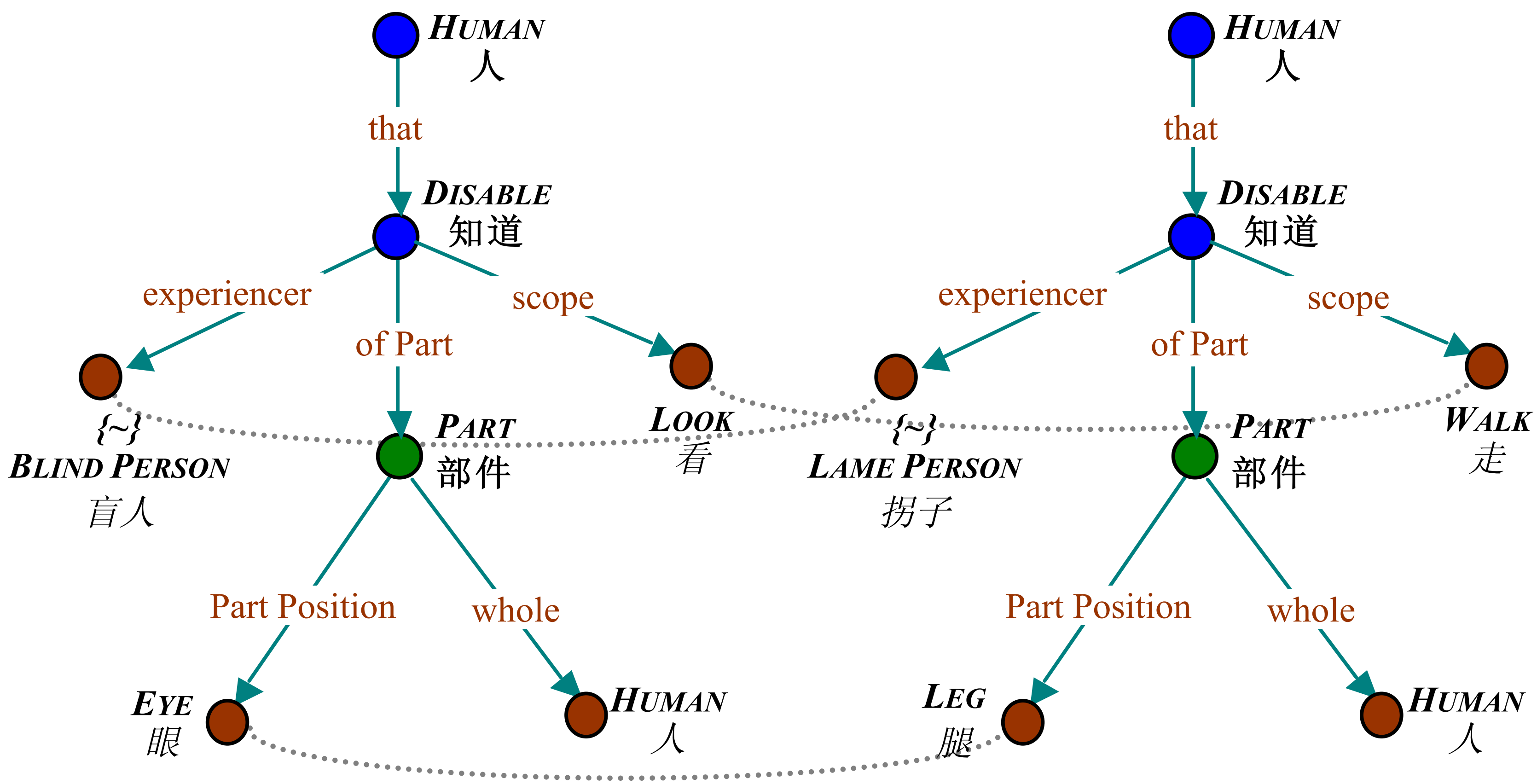
Structural-Inversion: Converting one definition into another



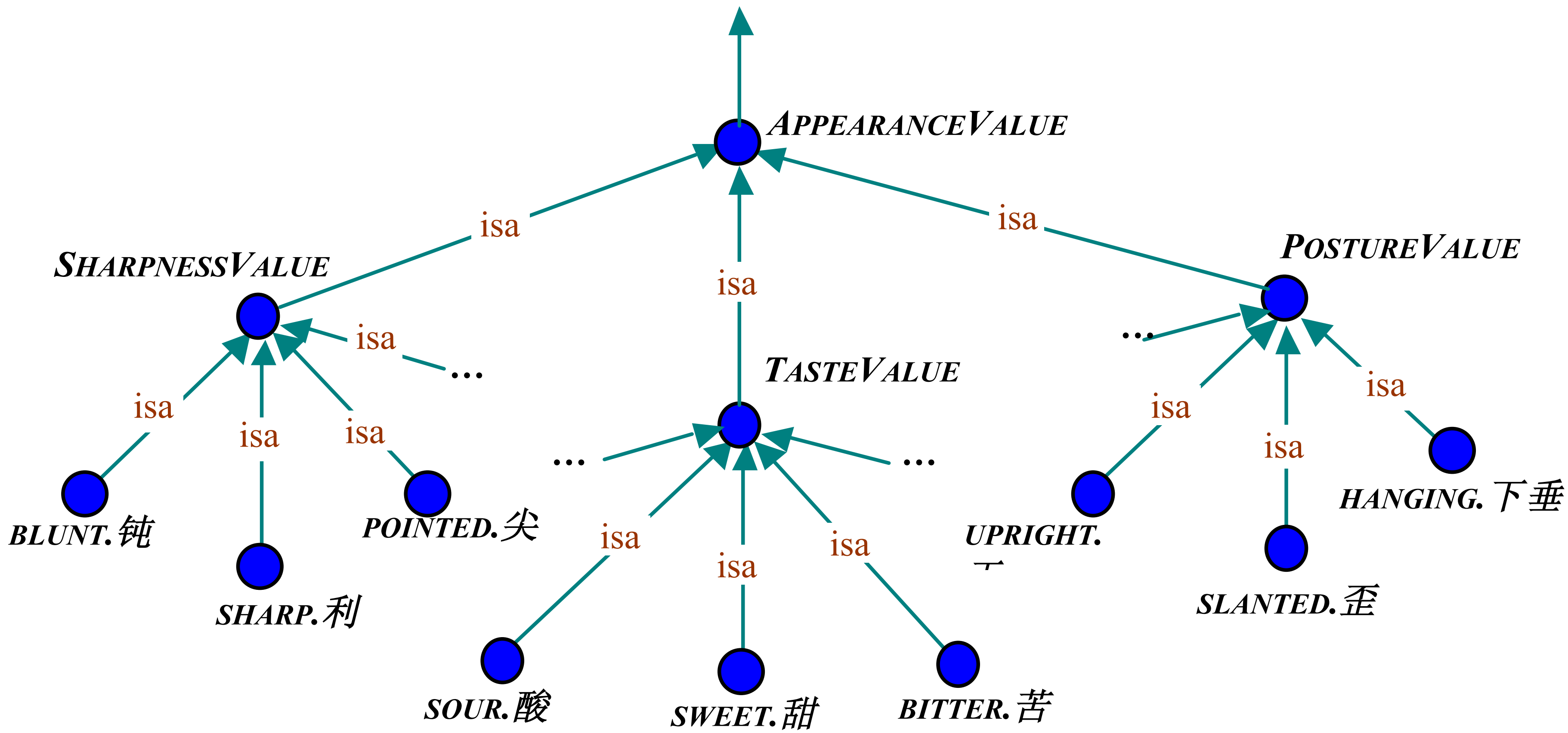
➔ **Bicycle:** ***{LandVehicle: {drive:agent={cyclist}, patient={~}}}***

Structure-Mapping in HowNet

Structure Hash: {?:{ill|病态:OfPart={?},experiencer={~},scope={?}}}



HowNet: Property / Modifier Taxonomy



E.g., (samurai)武士 ⇒ Samurai:CourageValue=valiant 武

Lexical Decomposition Can Unlock Implicit World Knowledge

brown rice 糙米 is already a phrasal decomposition in English

However: 糙 = (rough)糙 + (rice)米 ∴ brown rice : *smoothness* = rough

(Blueprint) 草案 = (rough)草 + (draft)案 ∴ blueprint : *carefulness* = rough

(estimate)概算 = (rough)概 + (count)算 ∴ estimate : *accuracy* = rough

(robot)机器人 = (mechanical)机器 + (person)人 ∴ robot *like a* person

(**bone-joint**) 骨节 = (skeleton)骨 + (knot)节 ∴ a joint *is like a* knot

Kenning Riddles

(breast)乳房 = (milk) 乳 + (house) 房

(sky)天宇 = (celestial)天 + (house) 宇

Exploiting Semantic Transparency in other Languages/Ontologies

Espresso =

浓咖啡

Deep (isa HueValue)
Strong (isa IntensityValue)
Rich (isa TasteValue)
Concentrated (isa ConcentrationValue)
Thick (isa DensityValue)

coffee (isa drinks)

Decomposition enables Construction of Slot-filler Frame-Structures

WordNet headword
as frame-name

HowNet parent of
orthographic modifier
yields a
Slot-Name

espresso	
IS-A:	***coffee***
Taste:	***rich***
Hue:	***deep***
Density:	***thick***
Intensity:	***strong***
See-also:	**cappuccino**
Patient-of:	***drink***

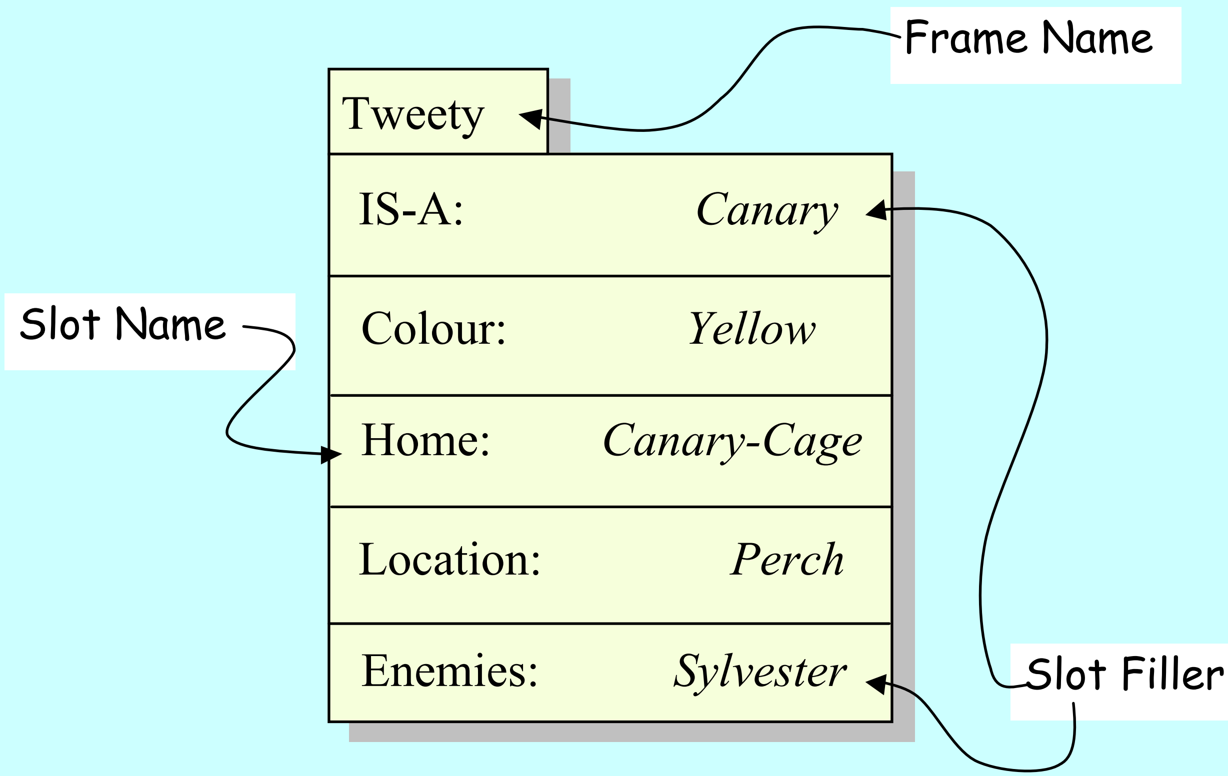
Translation of
HowNet modifier
yields a
Slot-Filler

CAUTION:
This enrichment relies on the sensibility of the HowNet ontology
+
transparency of orthographic modifiers

Frame-based Knowledge Representation

An Object-Oriented AI Knowledge-Representation Scheme

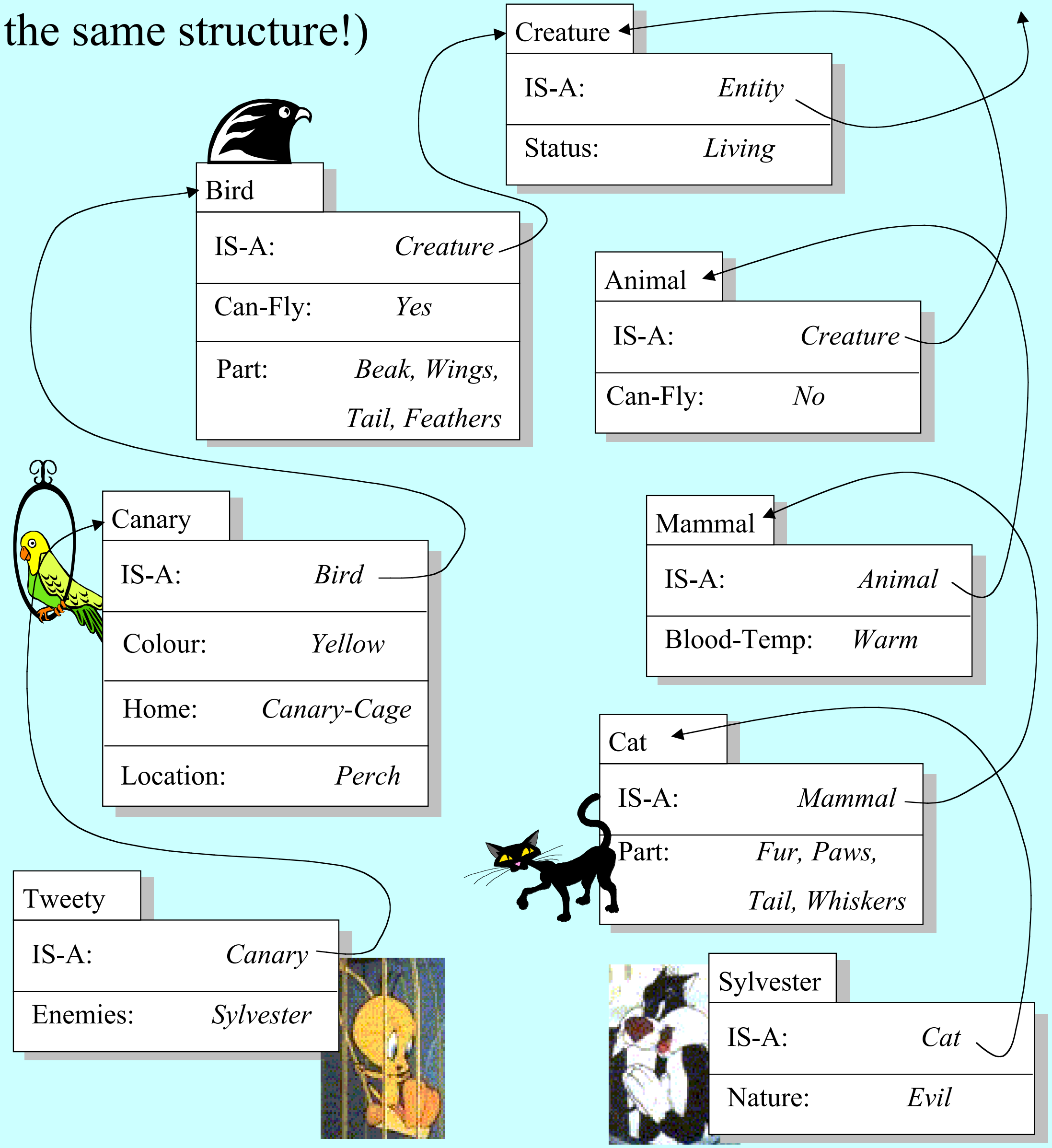
Introduced by ***Marvin Minsky*** in 1976, ***Frames*** combine a structured representation (*slots/fillers*) with procedural attachments (called ***demons*** or ***methods***):



A frame gathers all the relevant knowledge on a given topic (entity, place, event, etc.) into an directly accessible ***centralized*** structure. Frames used to represent **Concepts** in AI / CogSci.

Frame Hierarchies / Taxonomies

Frames are usually organised into an 'IS-A' hierarchy, or concept taxonomy, or ontology (basically 3 names for the same structure!)



Inheritance along IS-A Links

Tweety inherits
<Can-Fly: Yes>,
<Home: Canary-Cage>,
<Location: Perch>,
<Color: Yellow>,
and various body-parts
from parent concepts

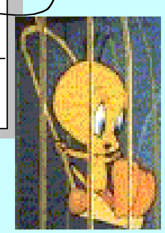
Creature	
IS-A:	*Entity*
Status:	*Living*

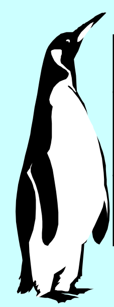
Bird	
IS-A:	*Creature*
Can-Fly:	*Yes*
Part:	*Beak, Wings, Tail, Feathers*

Canary	
IS-A:	*Bird*
Colour:	*Yellow*
Home:	*Canary-Cage*
Location:	*Perch*

Penguin	
IS-A:	*Bird*
Blood-Temp:	*Warm*
Can-Fly:	*No*
Part:	*Blubber*
Habitat:	*South-Pole*

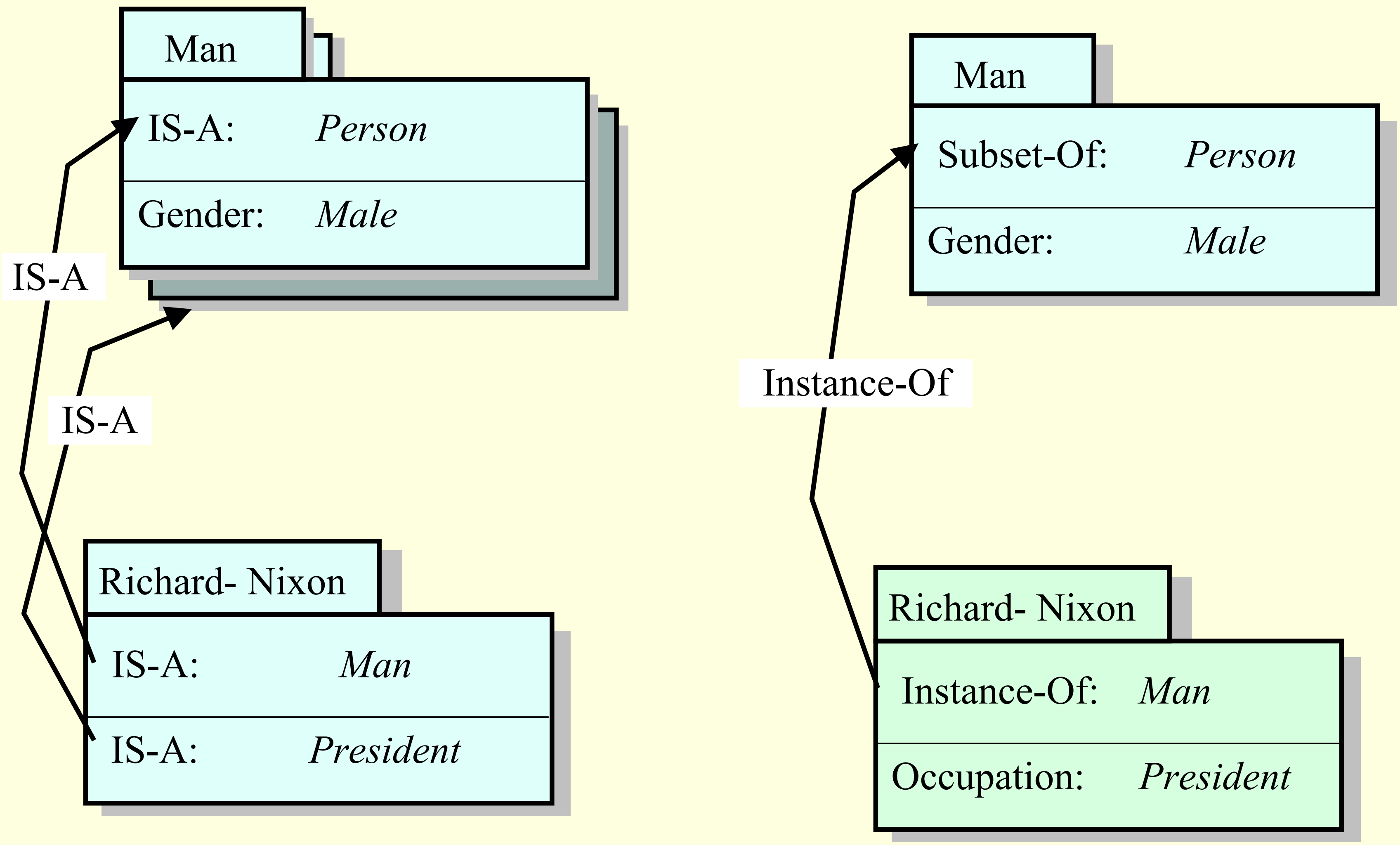
Tweety	
IS-A:	*Canary*
Enemies:	*Sylvester*



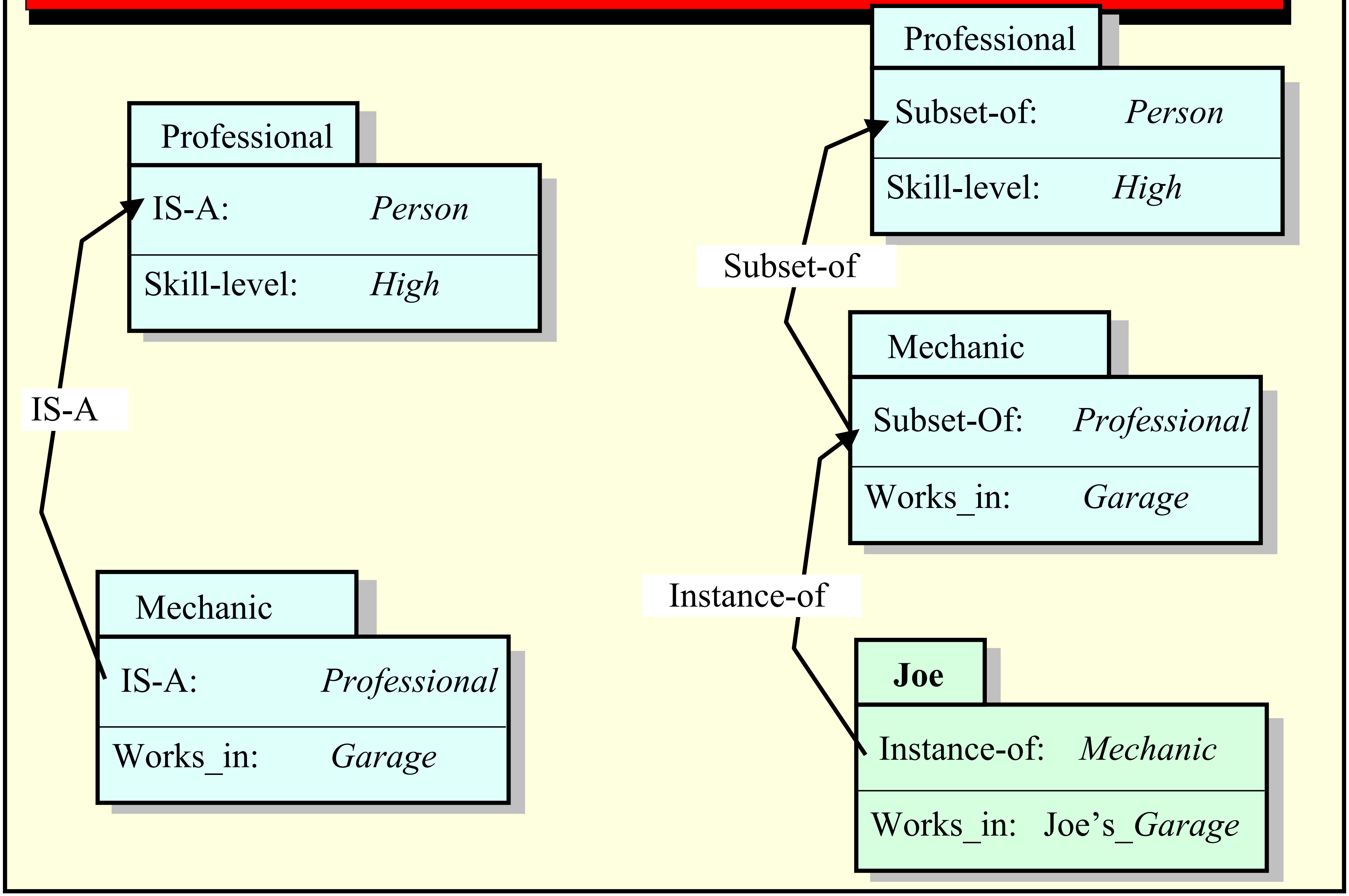


Fred	
IS-A:	*Penguin*
Enemies:	*WallyWalrus*

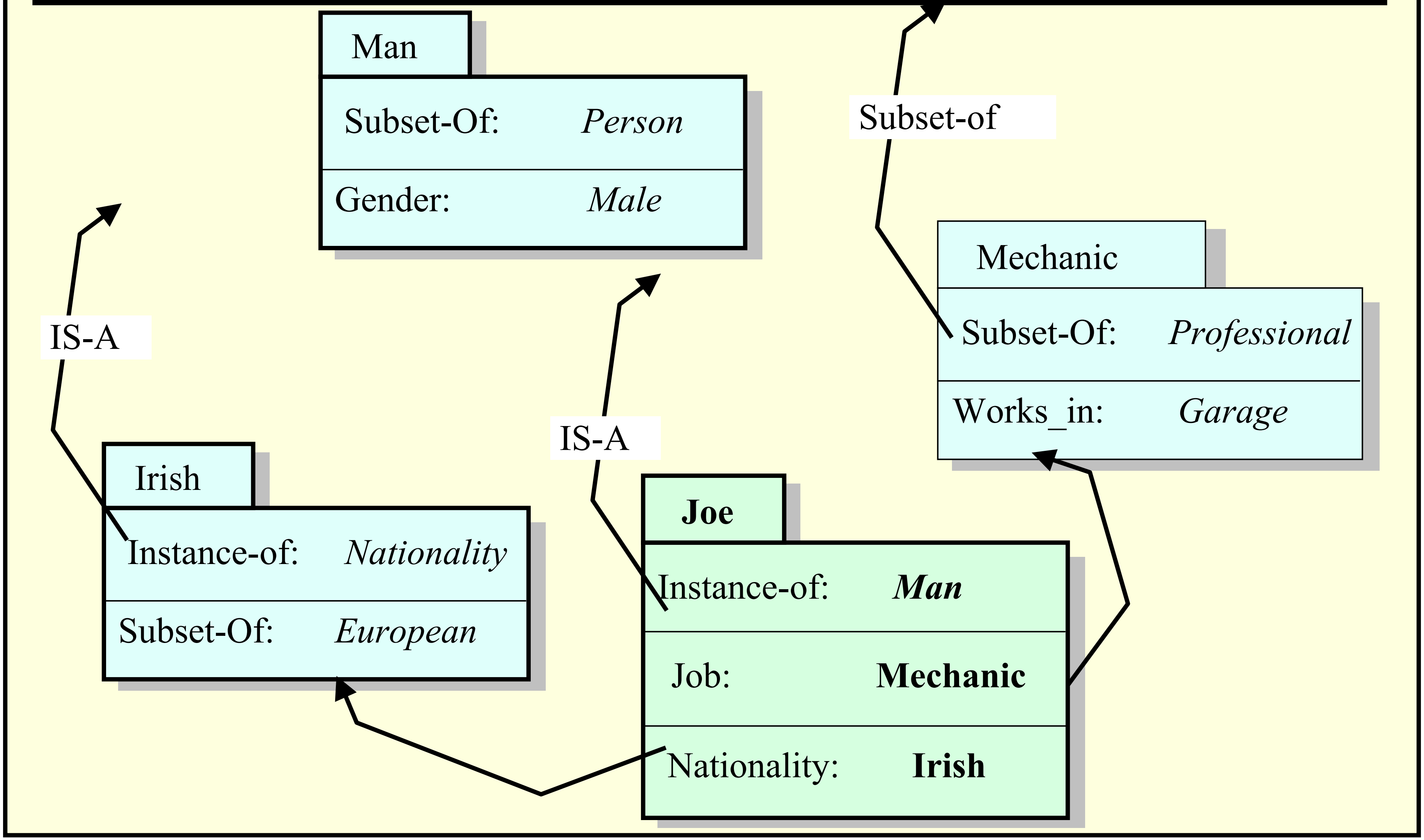
What's in an ISA-Link I: Instantiation



What's in an ISA-Link II: Class Subset Relations



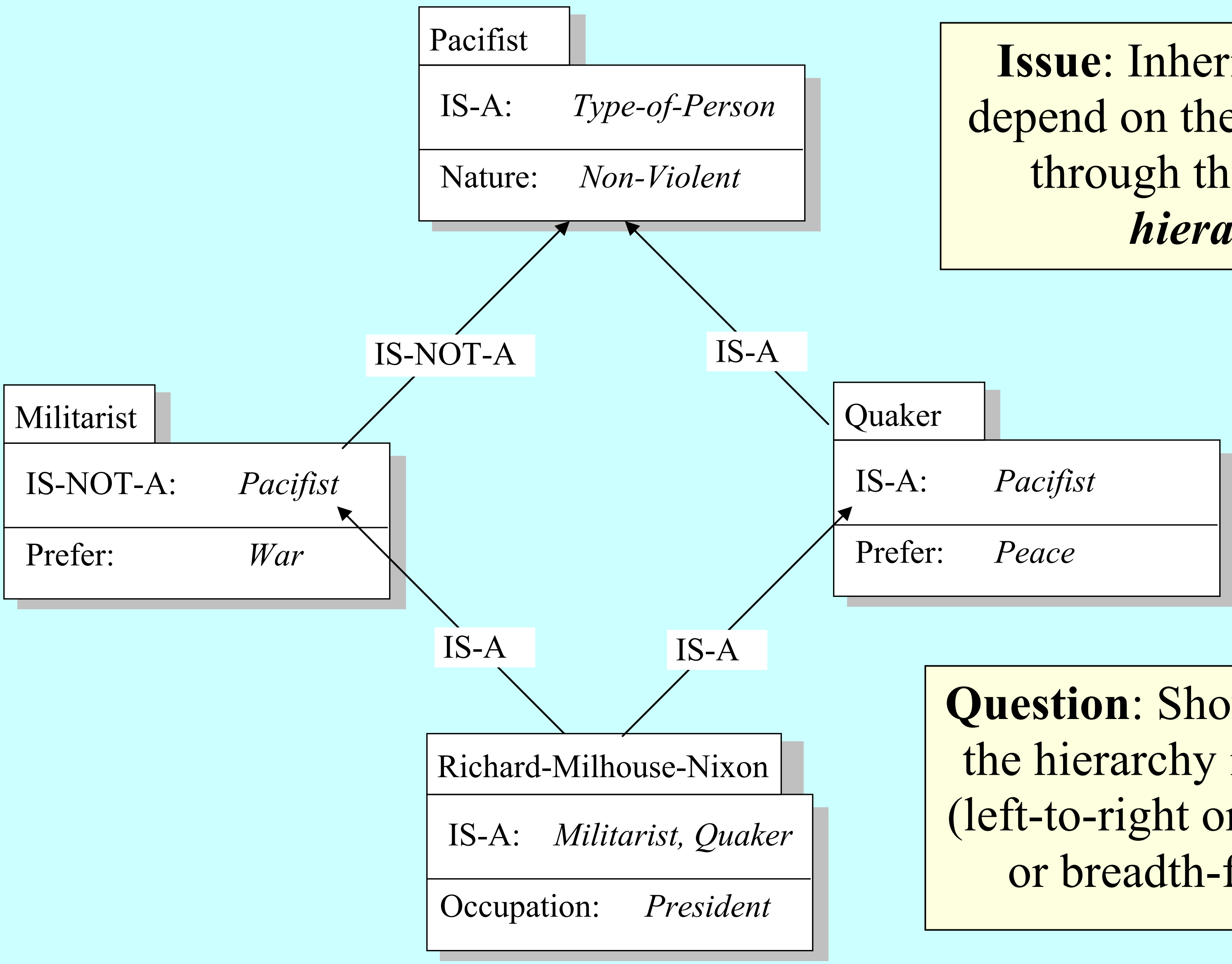
What's in an ISA-Link III: Role Relations



E.g., "Joe" is "Irish" not: "Joe" IS-A "Irish" IS-A "Nationality"

Multiple Inheritance can lead to Conflicts

A classic example of category conflict / inheritance conflict is the **'Nixon Diamond'** :



Issue: Inherited features depend on the path we take through the ***category hierarchy***.

Question: Should we search the hierarchy in depth-first (left-to-right or right-to-left) or breadth-first order?

Classical View: Intensional Definition of Categories

The categories in an Ontology can be specified/defined in two different ways:

Extensional Definition: *by exhaustively listing all members of the category*

Intensional Definition: *by specifying the logical criteria for membership*

Kingdom:	Animalia	Animals: Multicellular, heterotrophic, eukaryotic organisms.
Phylum:	Chordata	Chordates: animals with a notochord.
Subphylum:	Vertebrata	Vertebrates: Animals with bony spines.
Class:	Mammalia	Mammals: animals that have hair and give milk to their young.
Subclass:	Eutheria	Placentals: Mammals in which the young develop in a placenta inside the uterus.

Above is an *intensional definition* of the categories Mammal, Chordate, *etc*.

Classical View: Extensional Definition of Categories

The **Extension** of a Category is an ***exhaustive enumeration*** of its members.

An **Extension** is a ***set-theoretic*** perspective on category structure that allows for useful set-theoretic operations to create ***new complex categories***.

E.g.,

DOG = {Fido, Fred, Sam, Jupiter, Apollo, Spot, Rover, Lassie, Rintintin, …}

FEMALE = {Eve, Mary, Helen, Hera, Boudicea, Theresa, Lassie, …}

BLACK = {Fido, Jupiter, Apollo, Rover, Onyx, Tar, Soot, DarthVader, …}

"female dog" = FEMALE ∩ DOG = {Lassie, …}

"black dog" = BLACK ∩ DOG = { Fido, Jupiter, Apollo, …}

Classical Definitions: Necessary and Sufficient Conditions

A **Necessary** condition is one that ***must be obeyed*** to be a member of a category

E.g., ***edible, organic, tree-borne, internal-seeds***, *etc*. are necessary for an **Apple**

two-wheeled, mobile, rigid, manual, *etc*. are necessary for a **Bicycle**

warm-blood, ***live-birth, suckles-young***, *etc*. are necessary for a **Mammal**

A **Sufficient** condition is one that ***in itself*** *is* ***enough*** to be a member of a category

E.g., **having-womb** is sufficient for membership in the category **Female**

A **Set of Conditions is Sufficient** if possessing them all *guarantees* membership

E.g., ***warm-blood*** *and* ***suckles-young*** *are together sufficient* to be a **Mammal**

Nonetheless, Not all Category Members are Equal

"Knowledge is knowing that a tomato is a fruit"

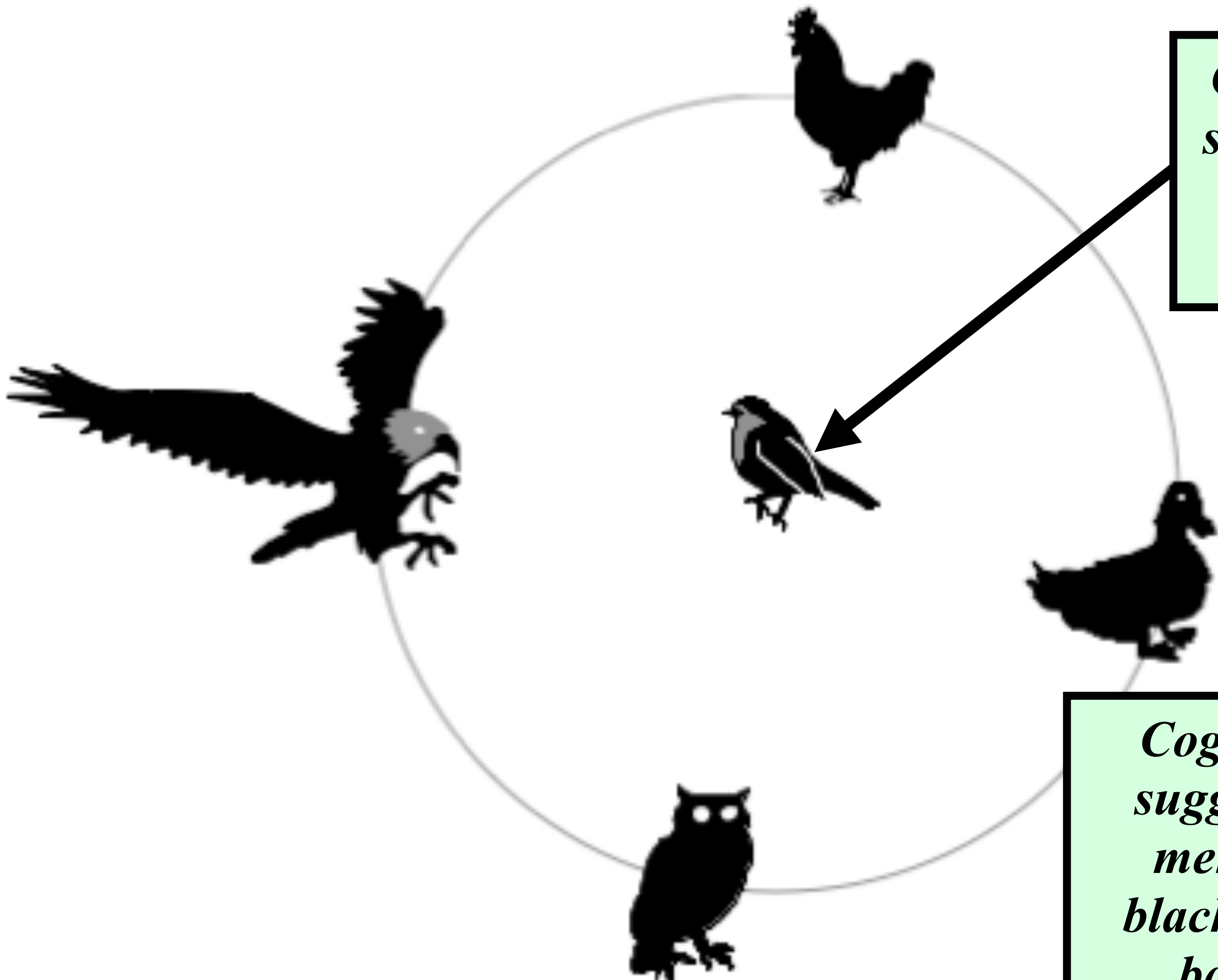
BUT

"Wisdom is knowing not to put it in a fruit salad"

In **Classical Logical** ***terms, all members are equal: each conforms to the*** **necessary and sufficient conditions** ***for membership. All are equally valid*.**

In **Cognitive psychology** ***terms, this is clearly not so: some members of a category are more typical/representative than others, and members are linked not by shared necessary features, but by a network of*** **family resemblances.**

The Non-Classical, Cognitive Perspective



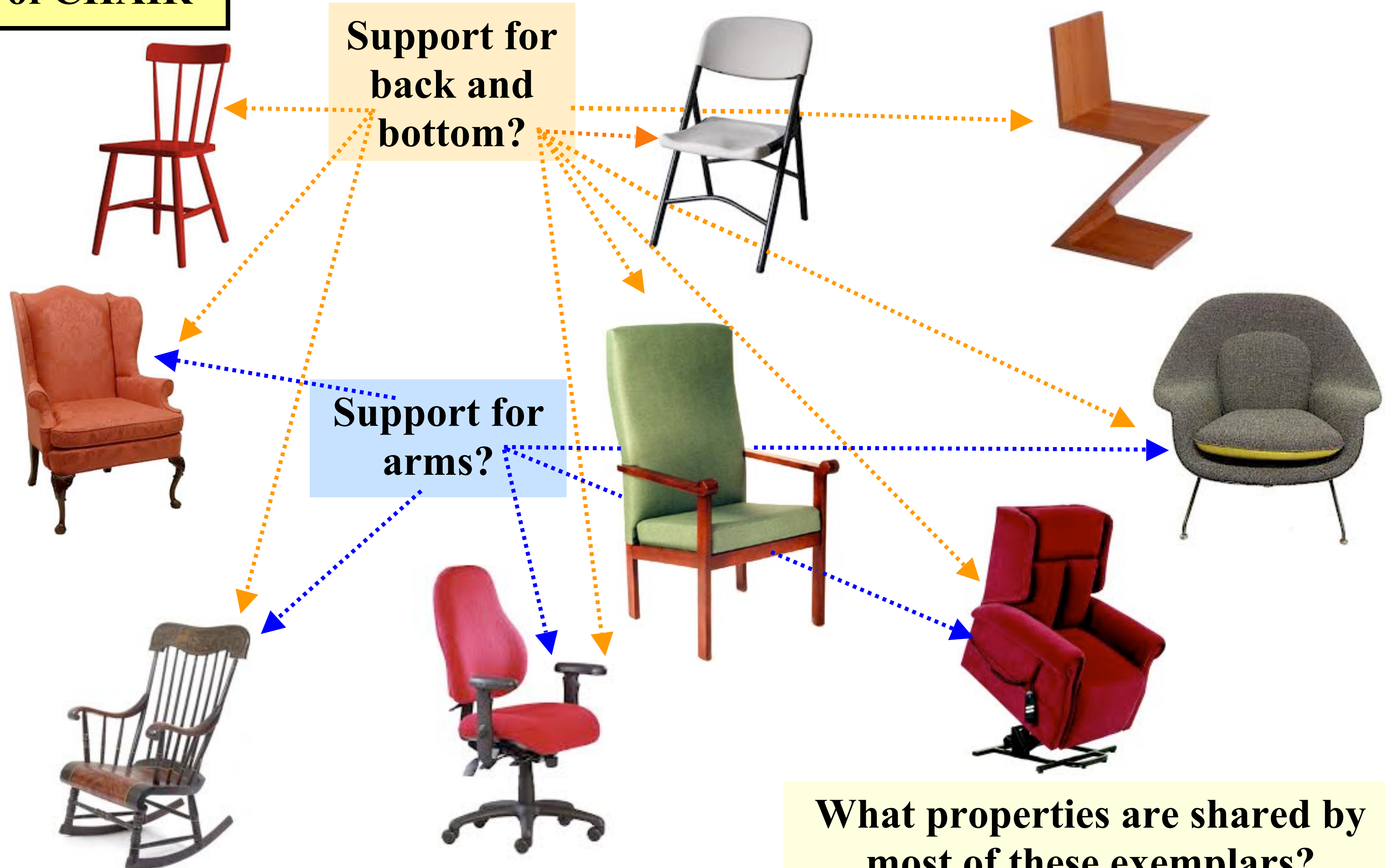
Categories are radially structured about highly central and evocative <u>prototypes</u>

Cognitive Psychology suggests that category membership is not a black-and-white affair, based on Boolean membership criteria

Categories have Prototypical Members:

What are the most accessible and typical examples of a category?

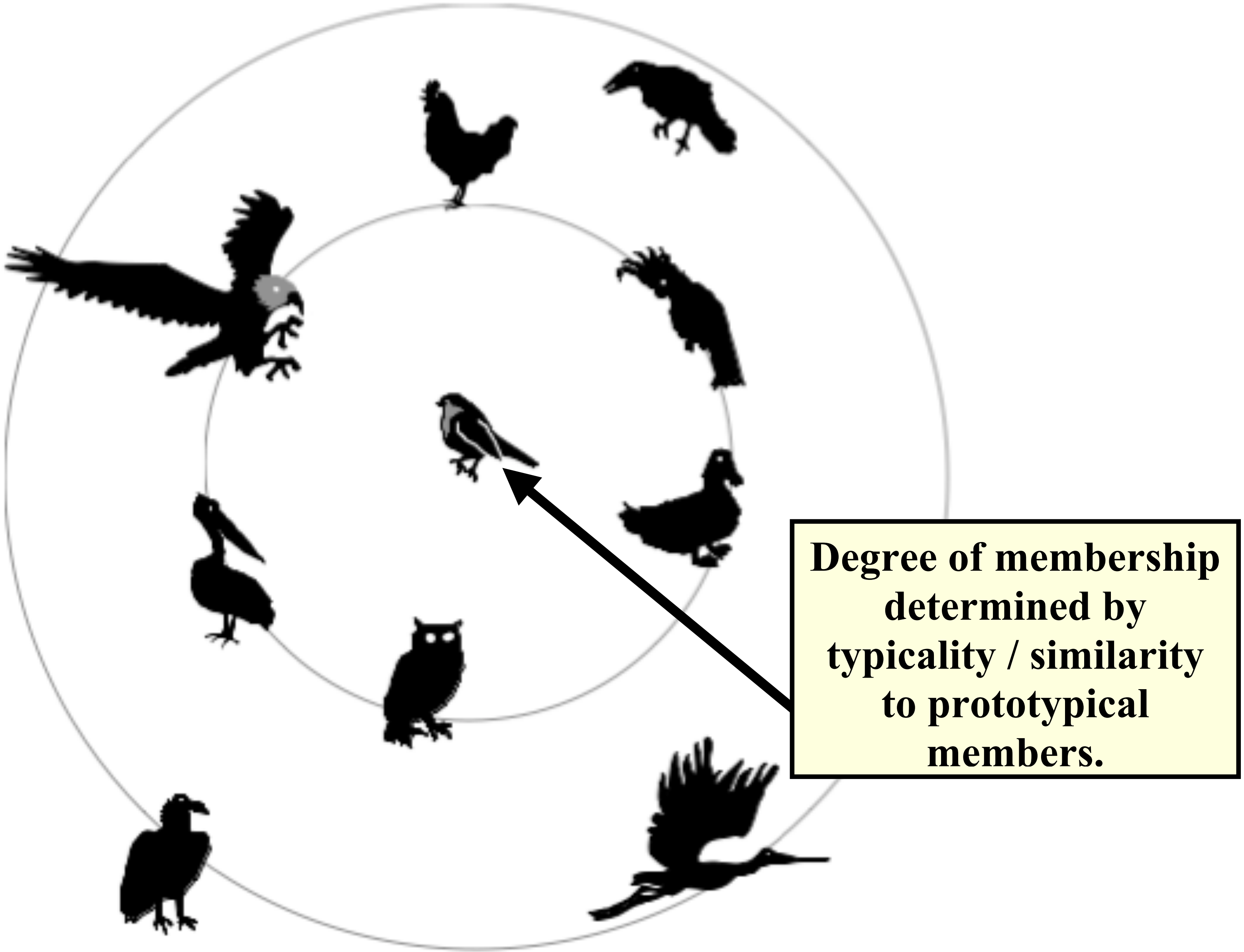
Prototypes of CHAIR



What properties are shared by most of these exemplars?

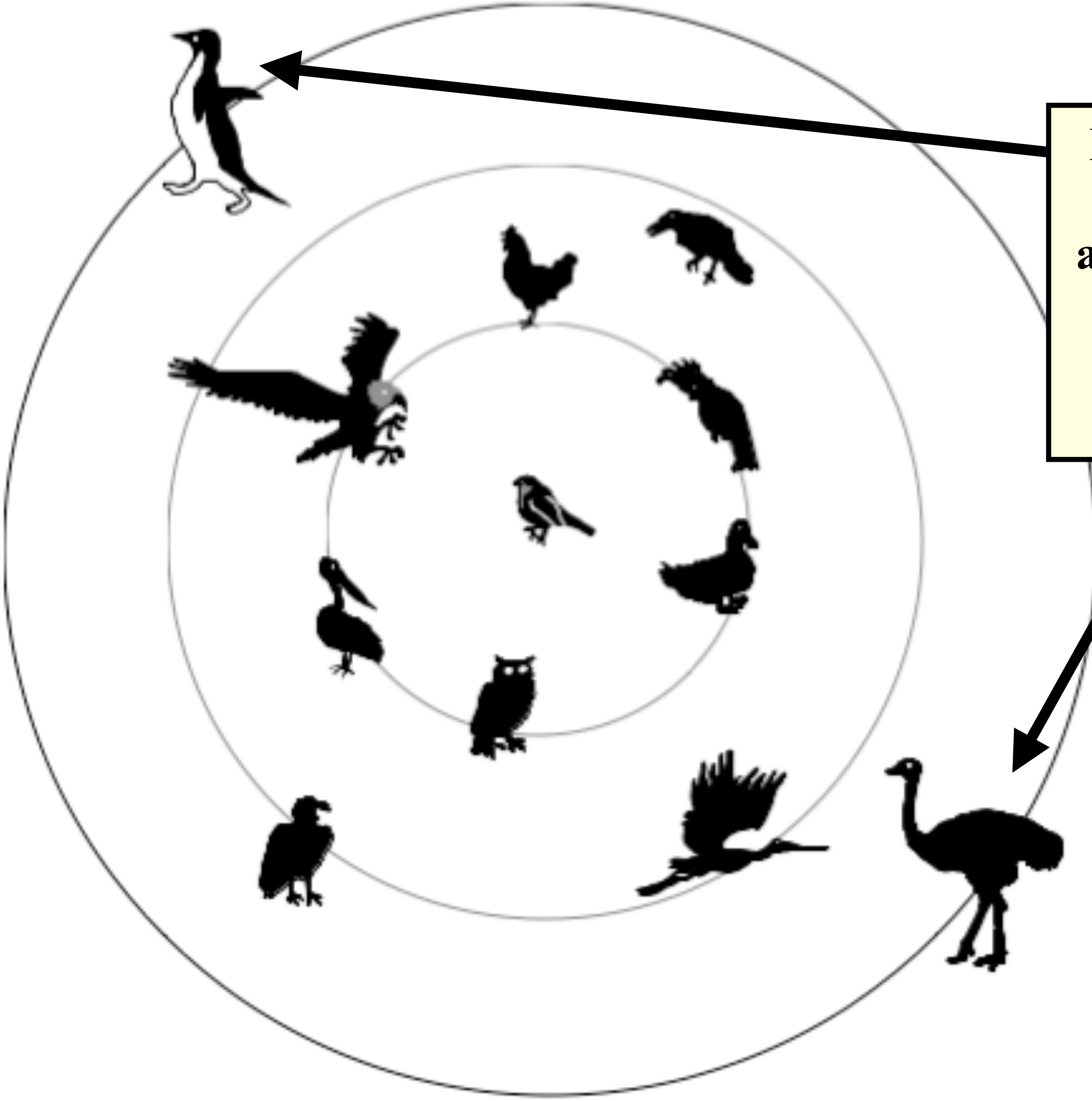
Radial Category Organization:

Some subclasses are more representative of a parent class than others



Radial Category Organization:

The Central Member is the Prototype, against which Typicality is Measured



Most categories have a shady area of almost or ‘strictly speaking’ members

Radial Category Organization: Humorous Cases

Some classes share core similarities without being properly subsumed