

# Creative Introspection and Knowledge Acquisition

*Learning about the world thru  
Introspective Questions and **Conceptual “Mash-Ups”***



**Tony Veale & Guofu Li**

**School of Computer Science  
and informatics**

**University College Dublin**

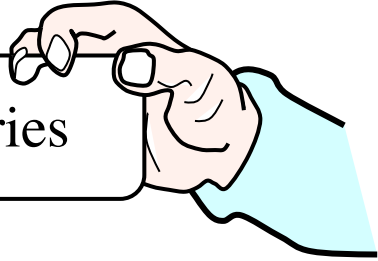
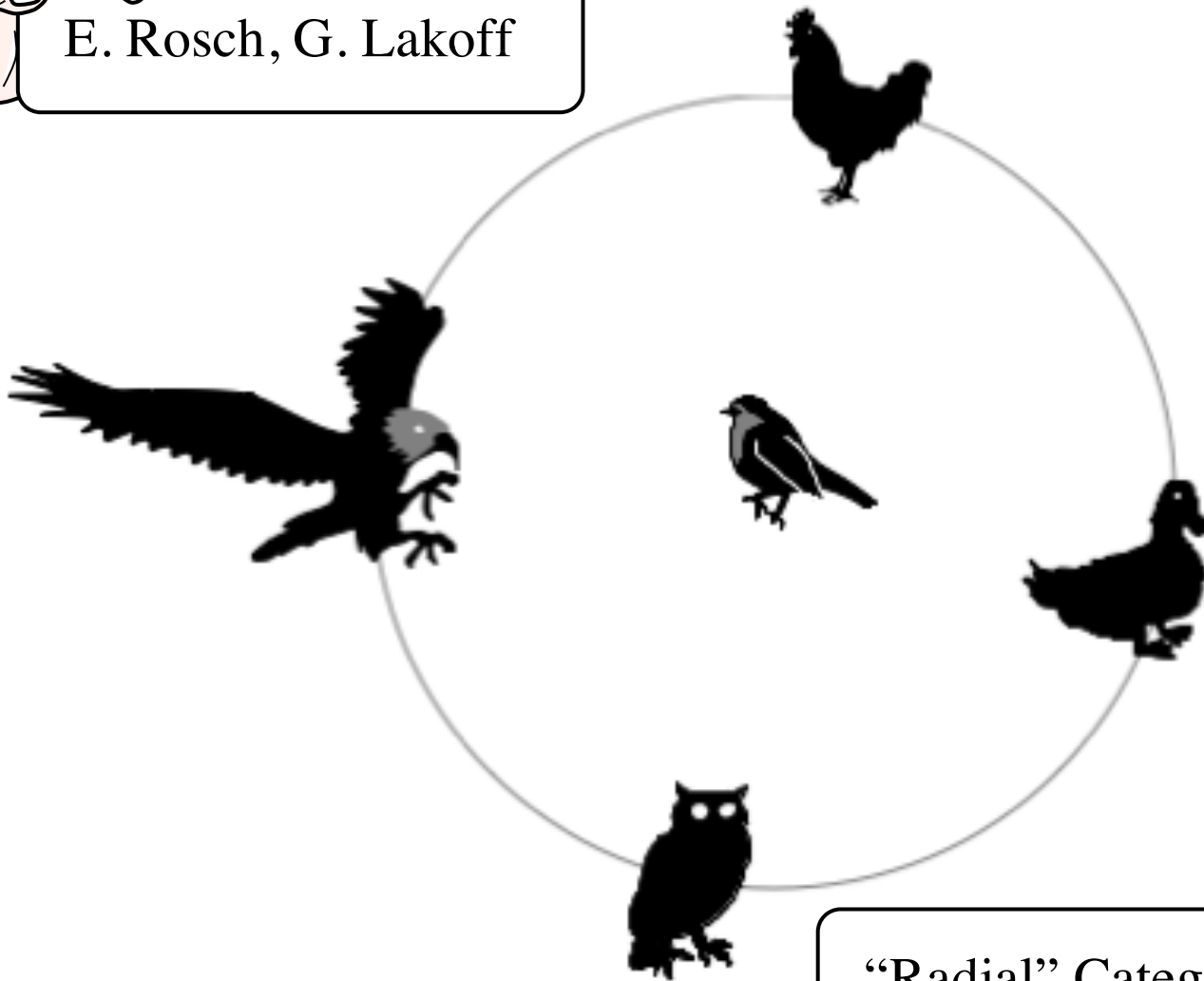
**[Tony.Veale@UCD.ie](mailto:Tony.Veale@UCD.ie)**

**<http://Afflatus.UCD.ie>**

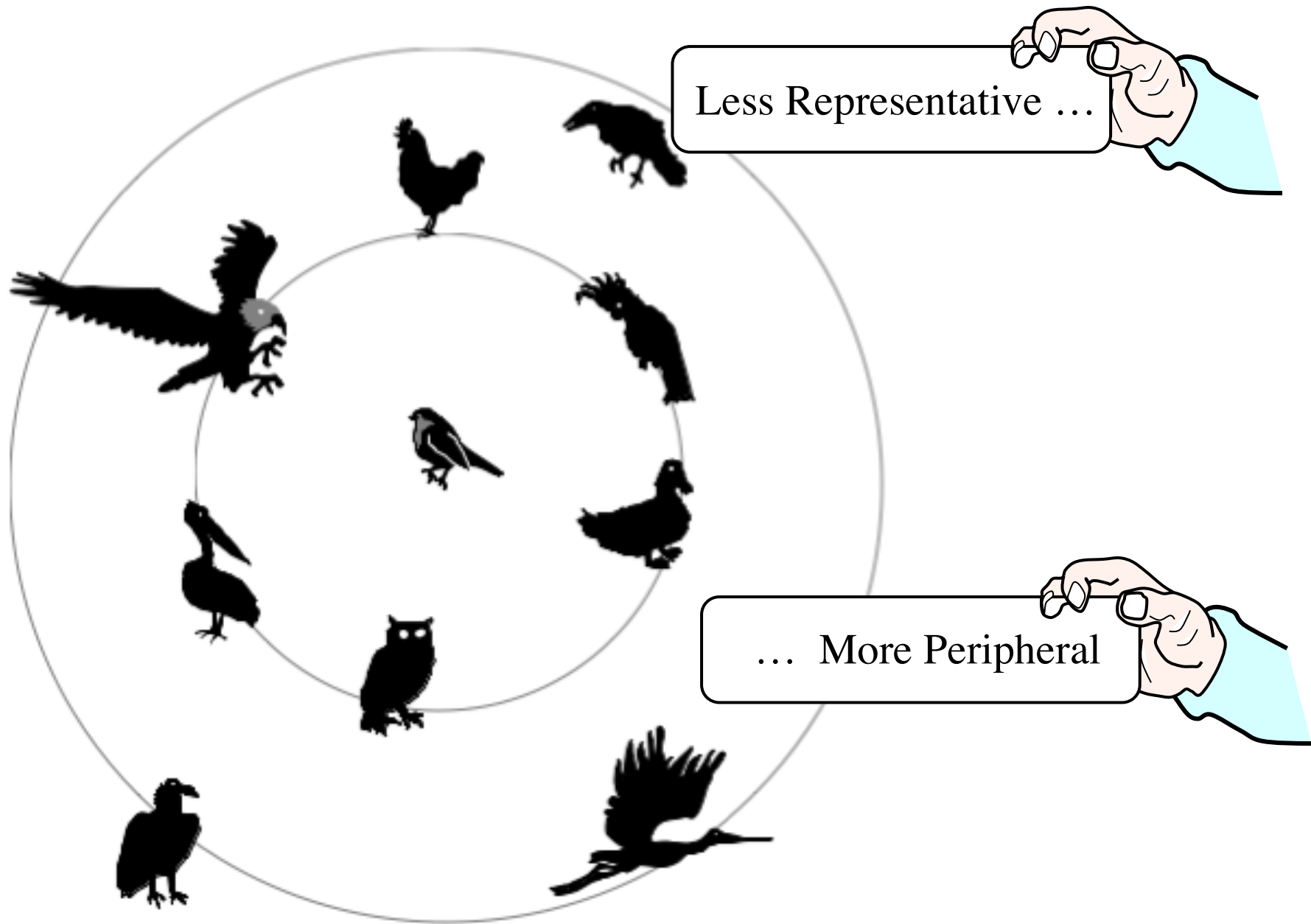
***AAAI'11, San Francisco, August 2011***

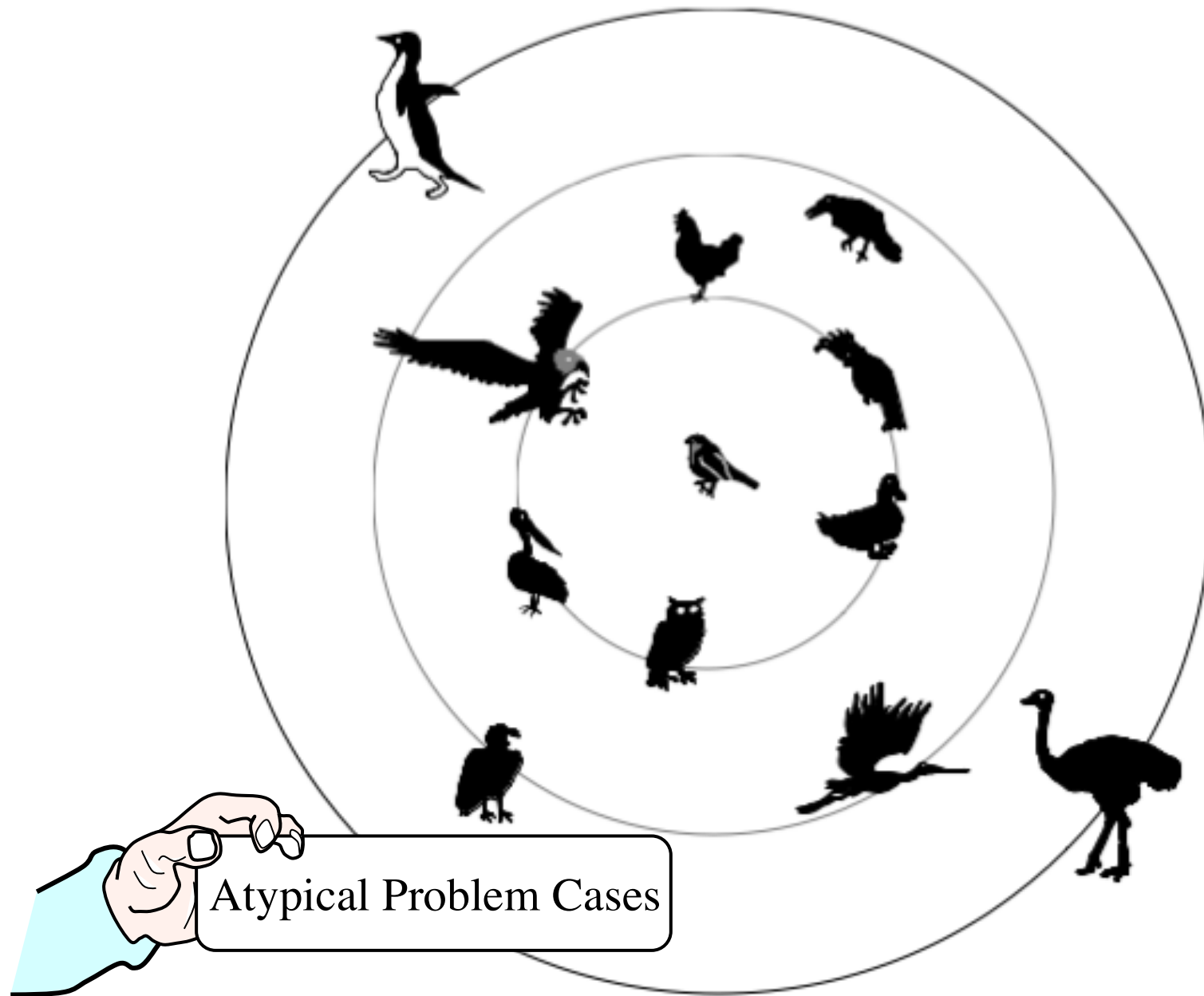


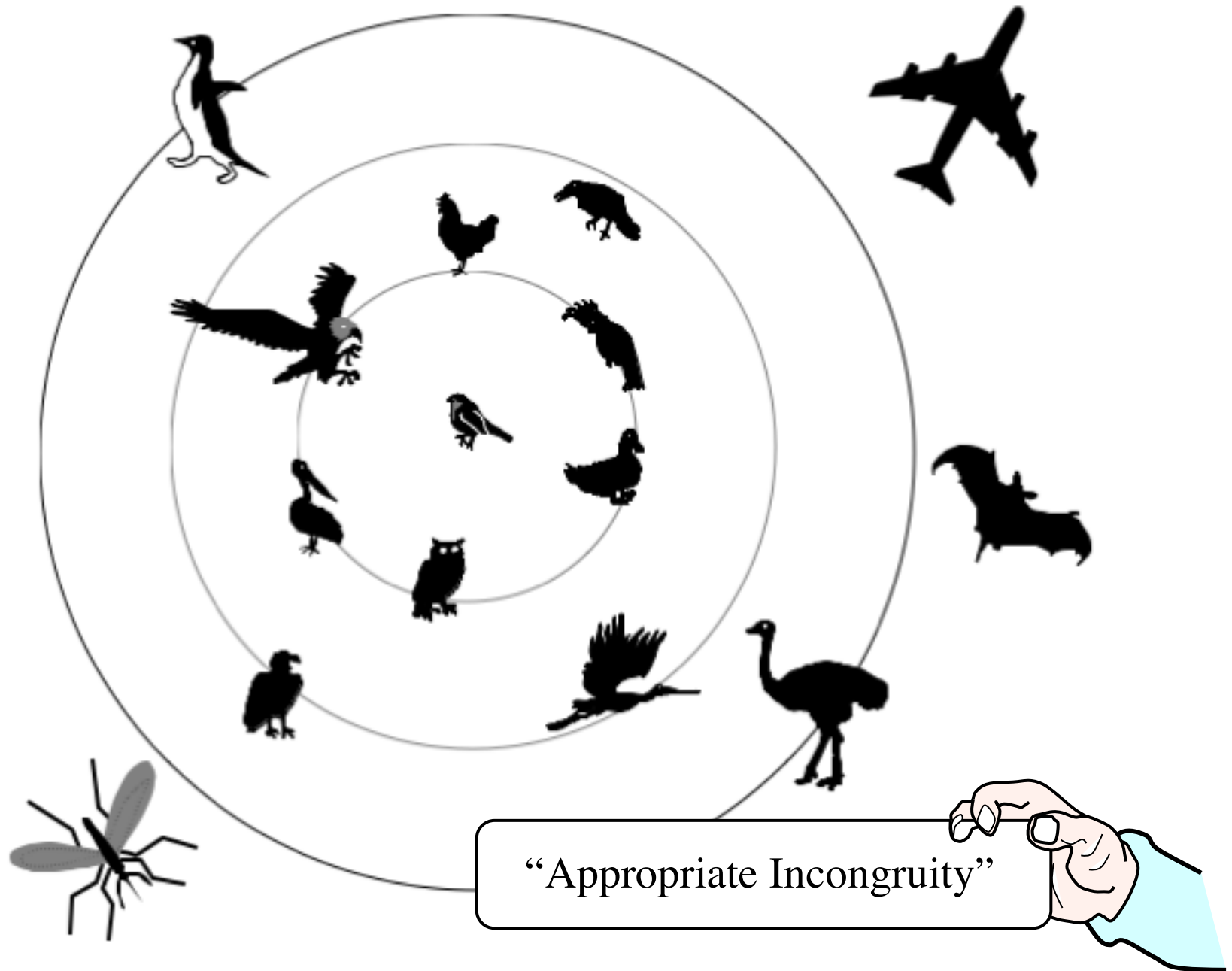
E. Rosch, G. Lakoff



“Radial” Categories





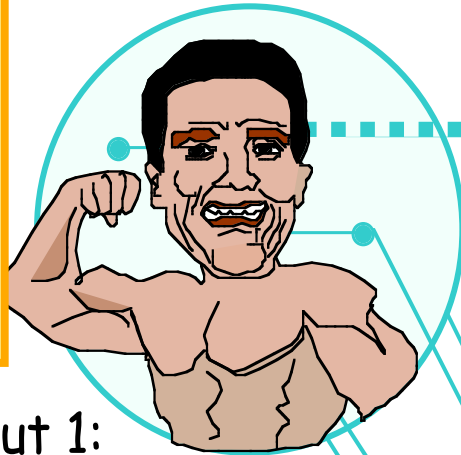


## Conceptual Blending: Creative reuse of existing knowledge

**Fauconnier & Turner**

Knowledge elements  
from multiple input  
spaces are selectively  
combined in a *blend*

Input 1:  
"Governor of California"



Input 2:  
President  
of the  
United States



Blend: "President of  
12% of the U.S."



**Governor of California**

**Governor of 12% of the U.S.**

**Leader of 12% of the U.S.**

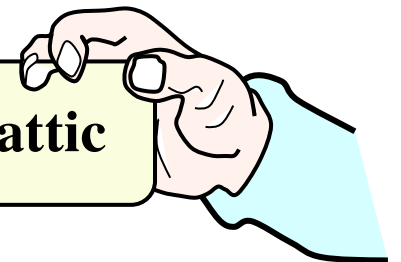
**President of 12% of the U.S.**

**President of 100% of the U.S.**

**President of the United States**



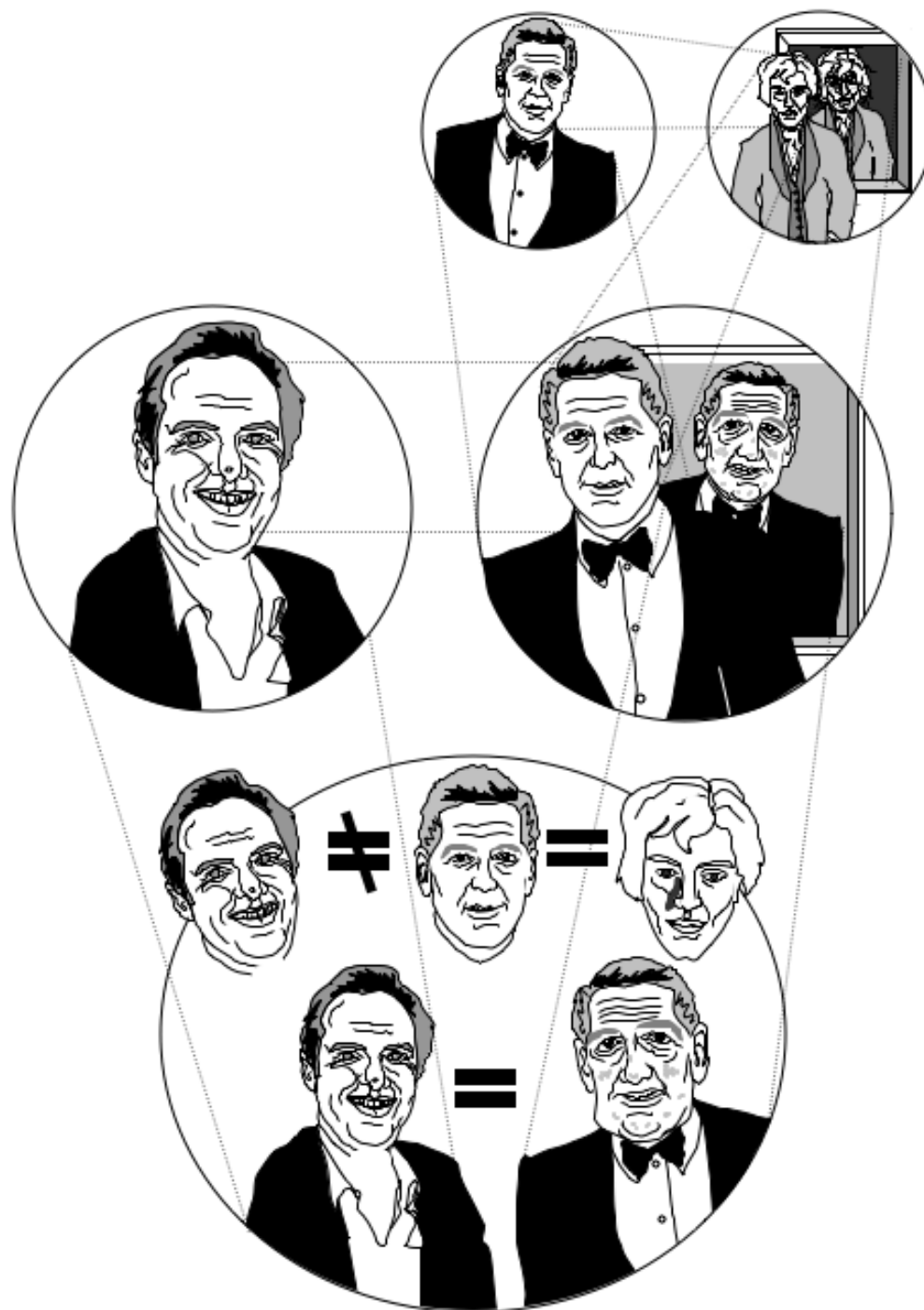
**Sam Mendes *[looks like]* the picture in George Clooney's attic**



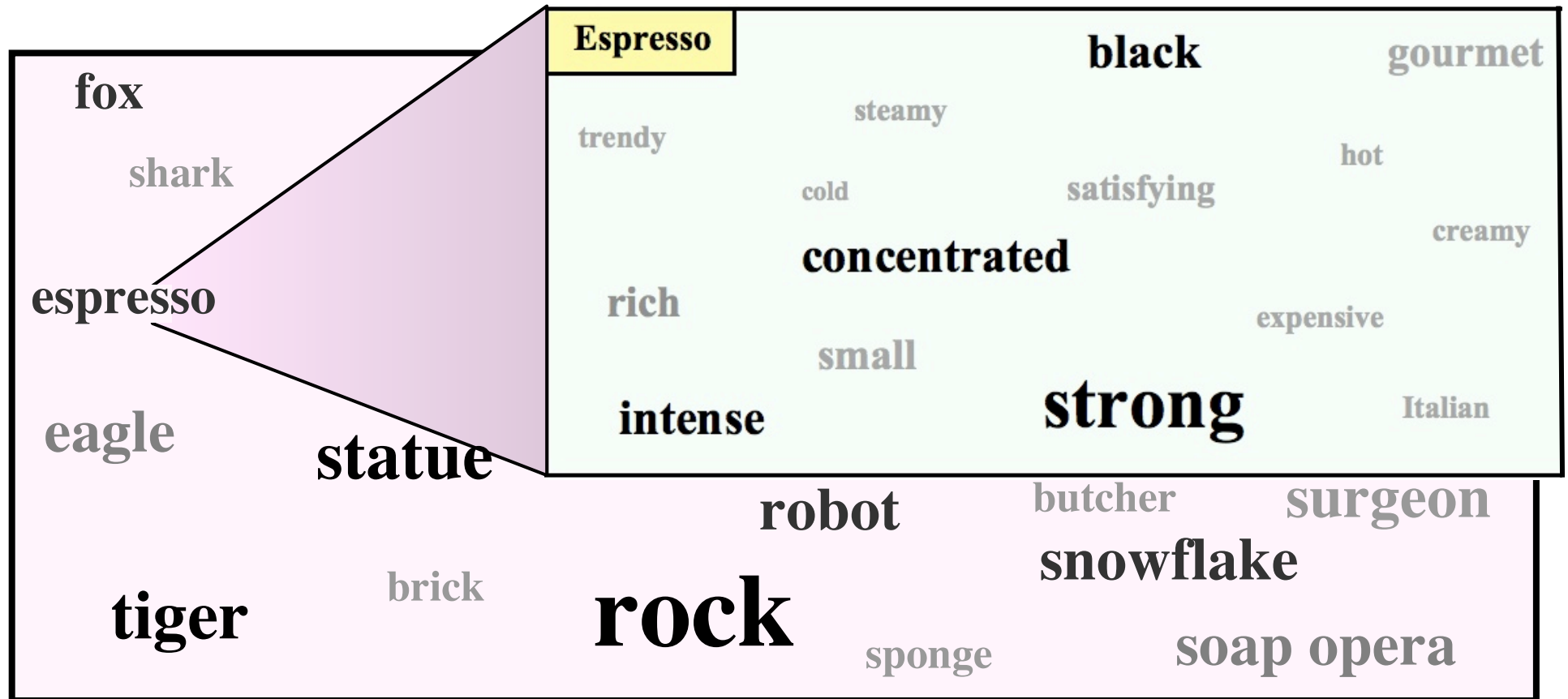








## Stereotypical Associations: Mine *Simile* patterns from the WWW



Use Web query pattern “as \* a | an as \* ” to harvest 1000’s of similes

## *Collocations in Corpora* provide *Pragmatic Category Structures*

### 5-grams

*Roger Federer , Tiger Woods*  
Rafael Nadal and Roger Federer  
Roger Federer, Andy Roddick  
*Thierry Henry , Roger Federer*  
*Tiger Woods , Roger Federer*  
David Beckham, Thierry Henry  
Tom Cruise & David Beckham  
Tom Cruise and Katie Holmes  
Steven Spielberg, Tom Cruise  
Tom Hanks / Steven Spielberg  
Dan Brown and Tom Hanks  
*Tiger Woods vs. Thierry Henry*  
: : : : :

### 4-grams

*tennis and golf players*  
tennis / squash players  
soccer and hockey moms  
polo and tennis teams  
squash and tennis courts  
soccer and rugby fields  
tennis and soccer fans  
*soccer and tennis players*  
polo and lacrosse teams  
*soccer vs. golf players*  
TV and movie stars  
radio and TV stars  
: : : :

### 3-grams

*tennis and golf*  
polo and tennis  
artists and scientists  
apples and oranges  
players and fans  
coaches and players  
*golf vs. soccer*  
terrorism and extremism  
*soccer versus tennis*  
Hollywood / Bollywood  
radio and TV  
actors and directors  
: : :

We use the *Google N-Grams 1T Web Corpus* ( $N \leq 5$ )

## Pragmatic Comparability Versus Semantic Similarity

<i>disaster</i>		<i>terrorist</i>		<i>beast</i>	
<i>tragedy</i>	99	<i>extremist</i>	90	<i>savage</i>	97
<i>catastrophe</i>	99	<i>radical</i>	88	<i>animal</i>	96
<i>calamity</i>	98	<i>anarchist</i>	83	<i>brute</i>	95
<i>destruction</i>	90	<i>subversive</i>	83	<i>wolf</i>	94
<i>famine</i>	89	<i>revolutionary</i>	82	<i>vulture</i>	86
<i>hardship</i>	89	<i>insurgent</i>	72	<i>pet</i>	83
<i>plague</i>	89	<i>separatist</i>	72	<i>plant</i>	73
<i>misfortune</i>	88	<i>guerrilla</i>	71	<i>thief</i>	73
<i>mishap</i>	85	<i>tyrant</i>	71	<i>bird</i>	70
<i>affliction</i>	84	<i>hacker</i>	70	<i>reptile</i>	64
<i>death</i>	80	<i>rebel</i>	70	<i>bandit</i>	63
<i>explosion</i>	70	<i>liberal</i>	70	<i>insect</i>	63
:		:		:	

Calculate WordNet-based semantic similarity for each coordination

## Common Questions On the Web: A Source of World Knowledge

why do cats **purr** × Search

why do cats **purr**  
why do cats **eat grass**  
why do cats **sleep so much**  
why do cats **cry**  
why do cats **have whiskers**

About 1,790,000 results (0.26 seconds)

[Go to Google.com](#) [Advanced search](#)

pourquoi les chats **ronronnent** × Rechercher

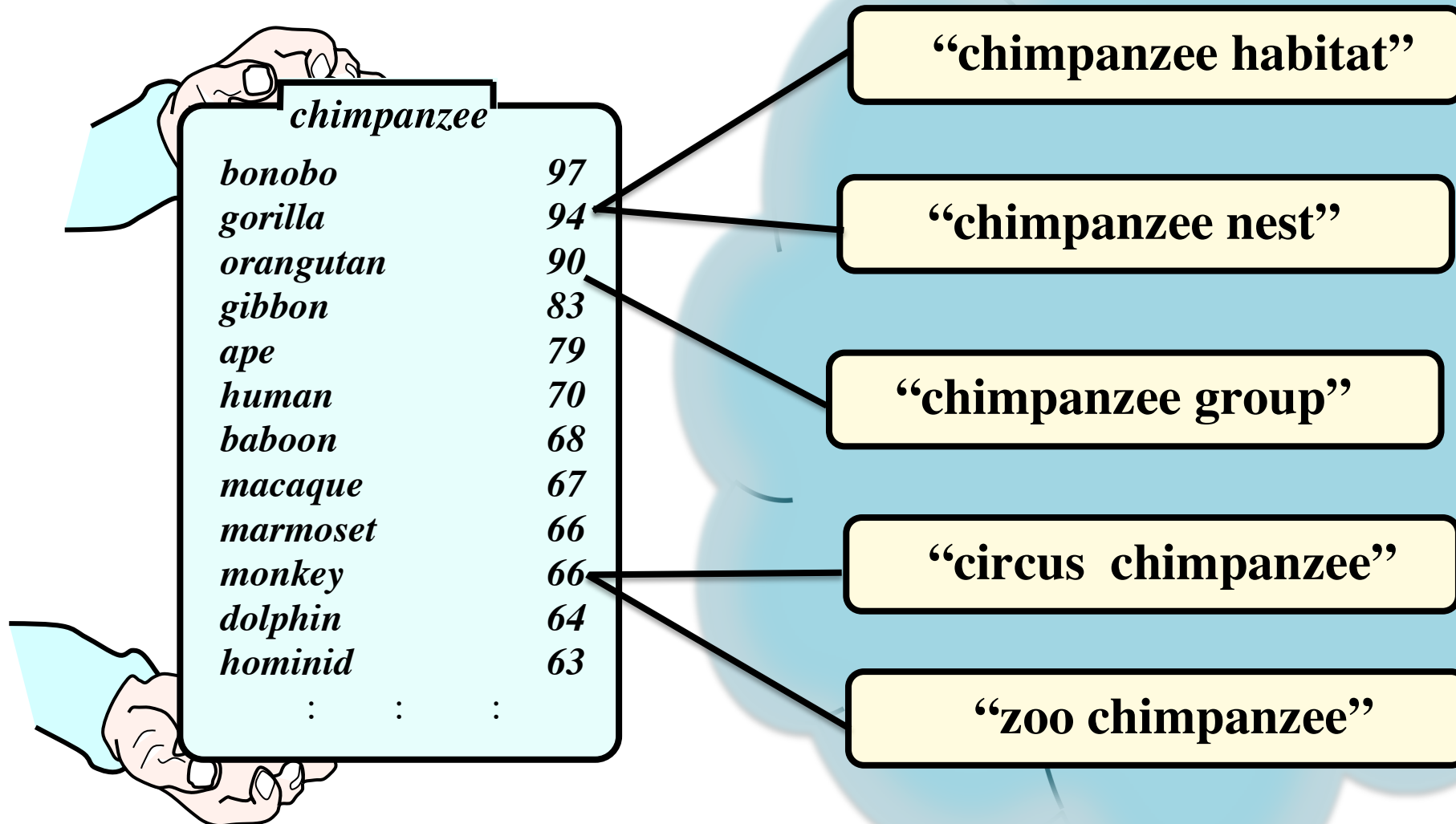
pourquoi les chats **ronronnent**  
pourquoi les chats **n'aiment pas l'eau**  
pourquoi les chats **mordent**  
pourquoi les chats **miaulent**  
pourquoi les chats **remuent la queue**

[En savoir plus](#)

Environ 144 000 000 résultats (0,05 secondes)

[Google.com in English](#) [Recherche avancée](#)

*We “milk” question completions from Google, and parse them into axioms*



## Mashups for chimpanzee, top 200

-  **why are chimpanzees supported by habitats** like gorilla chimpanzee habitat  score: 192, support: 64, similarity: 94  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj habitat) (supported\_by ?Subj ?Obj)))))
-  **why do chimpanzees live in habitats** like gorilla chimpanzee habitat  score: 190, support: 32, similarity: 94  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj habitat) (live\_in ?Subj ?Obj)))))
-  **why do chimpanzees build nests** like gorilla chimpanzee nest  score: 183, support: 8, similarity: 94  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj nest) (build ?Subj ?Obj)))))
-  **why don't chimpanzees live in groups** like orangutan chimpanzee group  score: 181, support: 10, similarity: 90  
(=> (instance ?Subj chimpanzee) (not (exists (?Obj) (and (instance ?Obj group) (live\_in ?Subj ?Obj)))))
-  **why do chimpanzees regurgitate their food** like gorilla chimpanzee food  score: 162, support: 2, similarity: 94  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj food) (regurgitate ?Subj ?Obj)))))
-  **why do chimpanzees live in circuses** like monkey circus chimpanzee  score: 162, support: 18, similarity: 66  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj circus) (live\_in ?Subj ?Obj)))))
-  **why are chimpanzees found in zoos** like monkey zoo chimpanzee  score: 162, support: 18, similarity: 66  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj zoo) (found\_in ?Subj ?Obj)))))
-  **why do chimpanzees live in zoos** like monkey zoo chimpanzee  score: 162, support: 18, similarity: 66  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj zoo) (live\_in ?Subj ?Obj)))))



# Filling the Knowledge Gap: Mash-Ups and Blends

**Mashes**

- Bonobo (2#)
- Gorilla (74#)
- Orangutan (14#)
- Gibbon (4#)
- Ape (25#)
- Human (671#)
- Baboon (21#)
- Monkey (228#)
- Dolphin (219#)
- Dog (1192#)
- Lion (351#)
- Tiger (384#)
- Camel (101#)
- Whale (249#)
- Wolf (343#)
- Cat (746#)
- Elephant (348#)
- Mouse

chimpanzees

E.g., Scientists as Artists, or, just Scientists

[Metaphorize This!](#)

Mashups for chimpanzee,

why are chimpanzees supported by habitats like gorilla chimpanzee habitat score: 192, support: 64, similarity: 94  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj habitat) (supported\_by ?Subj ?Obj)))))

why do chimpanzees live in habitats like gorilla chimpanzee habitat score: 190, support: 32, similarity: 94  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj habitat) (live\_in ?Subj ?Obj)))))

why do chimpanzees build nests like gorilla chimpanzee nest score: 183, support: 8, similarity: 94  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj nest) (build ?Subj ?Obj)))))

why don't chimpanzees live in groups like orangutan chimpanzee group score: 181, support: 10, similarity: 90  
(=> (instance ?Subj chimpanzee) (not (exists (?Obj) (and (instance ?Obj group) (live\_in ?Subj ?Obj)))))

why do chimpanzees regurgitate their food like gorilla chimpanzee food score: 162, support: 2, similarity: 94  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj food) (regurgitate ?Subj ?Obj)))))

why do chimpanzees live in circuses like monkey circus chimpanzee score: 162, support: 18, similarity: 66  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj circus) (live\_in ?Subj ?Obj)))))

why are chimpanzees found in zoos like monkey zoo chimpanzee score: 162, support: 18, similarity: 66  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj zoo) (found\_in ?Subj ?Obj)))))

why do chimpanzees live in zoos like monkey zoo chimpanzee score: 162, support: 18, similarity: 66  
(=> (instance ?Subj chimpanzee) (exists (?Obj) (and (instance ?Obj zoo) (live\_in ?Subj ?Obj)))))

[ngrams.UCD.ie/metaphor-eye/](http://ngrams.UCD.ie/metaphor-eye/)



Google is a cult

E.g., Scientists as Artists, or, just Scientists



Metaphorize This!

[ngrams.UCD.ie/metaphor-eye/](http://ngrams.UCD.ie/metaphor-eye/)

### Mashups for Google as cult

-  why does Google have apologists like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj apologist) (have ?Subj ?Obj)))))
-  why does Google enforce beliefs like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj belief) (enforce ?Subj ?Obj)))))
-  why does Google promote beliefs like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj belief) (promote ?Subj ?Obj)))))
-  why does Google worship celebrities like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj celebrity) (worship ?Subj ?Obj)))))
-  why does Google worship gods like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj god) (worship ?Subj ?Obj)))))
-  why does Google worship gurus like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj guru) (worship ?Subj ?Obj)))))
-  why is Google led by gurus like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj guru) (led\_by ?Subj ?Obj)))))
-  why does Google follow gurus like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj guru) (follow ?Subj ?Obj)))))
-  why is Google established by gurus like cult  score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj guru) (established ?Subj ?Obj)))))

## Mashups for Google as cult

-  why does Google have apologists like cult google apologist score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj apologist) (have ?Subj ?Obj)))))
-  why does Google enforce beliefs like cult google belief score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj belief) (enforce ?Subj ?Obj)))))
-  why does Google promote beliefs like cult google belief score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj belief) (promote ?Subj ?Obj)))))
-  why does Google worship celebrities like cult google celebrity score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj celebrity) (worship ?Subj ?Obj)))))
-  why does Google worship gods like cult google god score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj god) (worship ?Subj ?Obj)))))
-  why does Google worship gurus like cult google guru score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj guru) (worship ?Subj ?Obj)))))
-  why is Google led by gurus like cult google guru score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj guru) (led\_by ?Subj ?Obj)))))
-  why does Google follow gurus like cult google guru score: 100, support: 0, similarity: 0  
(=> (instance ?Subj google) (exists (?Obj) (and (instance ?Obj guru) (follow ?Subj ?Obj)))))
-  why is Google established by gurus like cult google guru score: 100, support: 0, similarity: 0

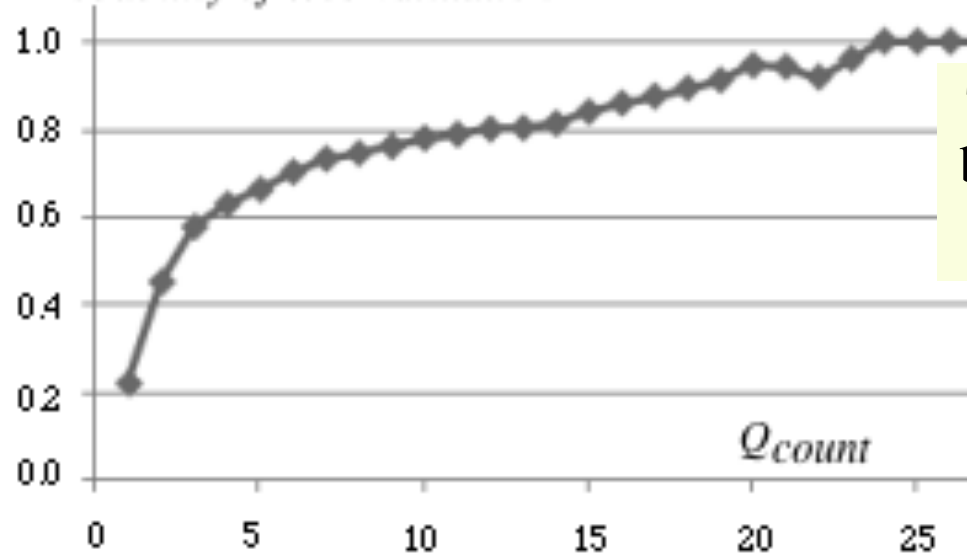
## Mashes

Passion (12#)  
 Hate (4#)  
 Loyalty (35#)  
 Worship (40#)  
 Lover (459#)  
 Emotion (27#)  
 Sex (1#)  
 Lust (30#)  
 Fear (26#)  
 Joy (78#)  
 Hatred (29#)  
 Reverence (1#)  
 Pet (148#)  
 Game (245#)  
 Pleasure (61#)  
 Desire (6#)  
 Feeling (17#)  
 Want (28#)  
 Belief (18#)  
 Friend (183#)  
 Dream (229#)  
 Memory (64#)  
 Study (178#)  
 Longing (56#)  
 Tease (8#)  
 Thought (111#)  
 Family (286#)  
 Experience (2#)  
 Client (46#)

	<b>why does <i>love</i> treat patients</b> like drug	love patient , patient's love	score: 100, support: 8, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj patient) (treat ?Subj ?Obj)))))</pre>		
	<b>why does <i>love</i> have interactions</b> like drug	love interaction	score: 100, support: 2, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj interaction) (have ?Subj ?Obj)))))</pre>		
	<b>how does <i>love</i> numb minds</b> like drug	love mind , love's mind	score: 100, support: 1, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj mind) (numb ?Subj ?Obj)))))</pre>		
	<b>why does <i>love</i> have dangers</b> like drug	love danger	score: 100, support: 0, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj danger) (have ?Subj ?Obj)))))</pre>		
	<b>why does <i>love</i> have doses</b> like drug	love dose	score: 100, support: 0, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj DOS) (have ?Subj ?Obj)))))</pre>		
	<b>why does <i>love</i> have generics</b> like drug	love generic	score: 100, support: 0, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj generic) (have ?Subj ?Obj)))))</pre>		
	<b>why does <i>love</i> have reactions</b> like drug	love reaction , love's reaction	score: 100, support: 0, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj reaction) (have ?Subj ?Obj)))))</pre>		
	<b>why does <i>love</i> have risks</b> like drug	love risk	score: 100, support: 0, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj risk) (have ?Subj ?Obj)))))</pre>		
	<b>why does <i>love</i> affect cells</b> like drug	love cell	score: 100, support: 0, similarity: 0
	<pre>(=&gt; (instance ?Subj love) (exists (?Obj) (and (instance ?Obj cell) (affect ?Subj ?Obj)))))</pre>		

## Mash-Ups as Sources of **Introspective Questions**

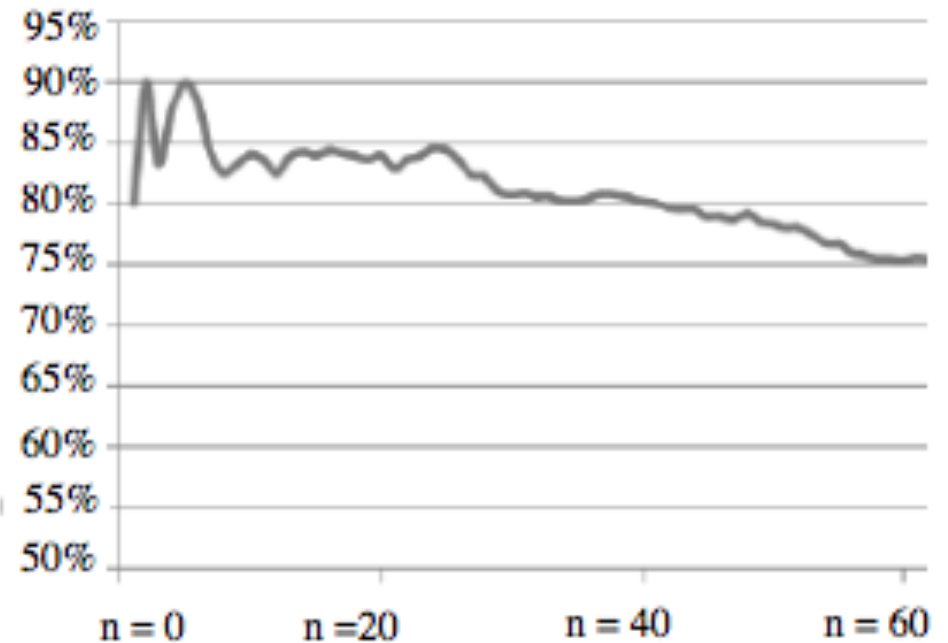
Probability of Web validation



The *more times* a hypothesis is suggested by *pragmatic neighbours*, the more likely it is to be validated on the Web

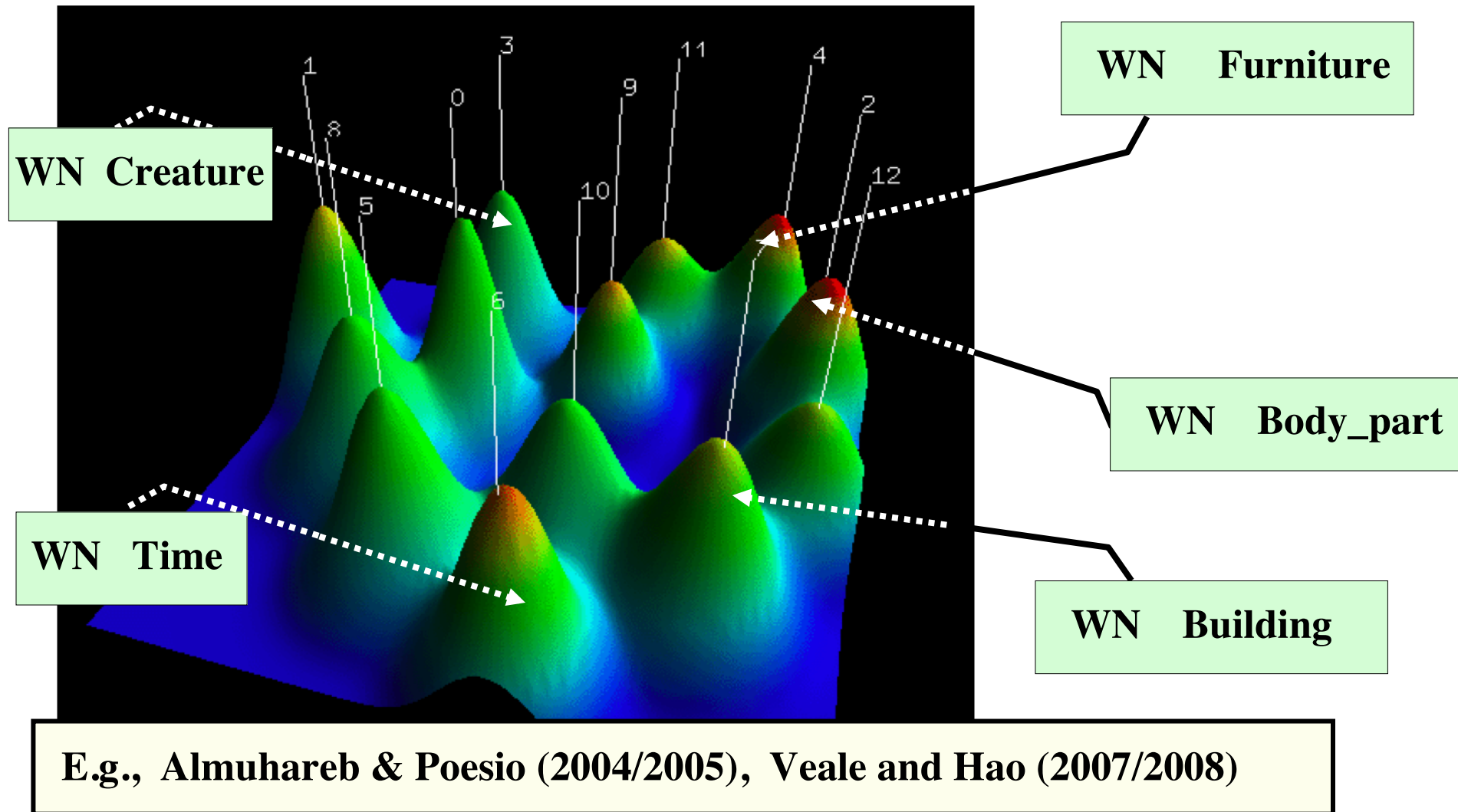
The most highly ranked hypotheses (by *similarity and #times suggested*) are very likely to be validated on the Web

Validation rate (top n)





## WordNet(s) as a Gold-Standard for our Mash-up Categories

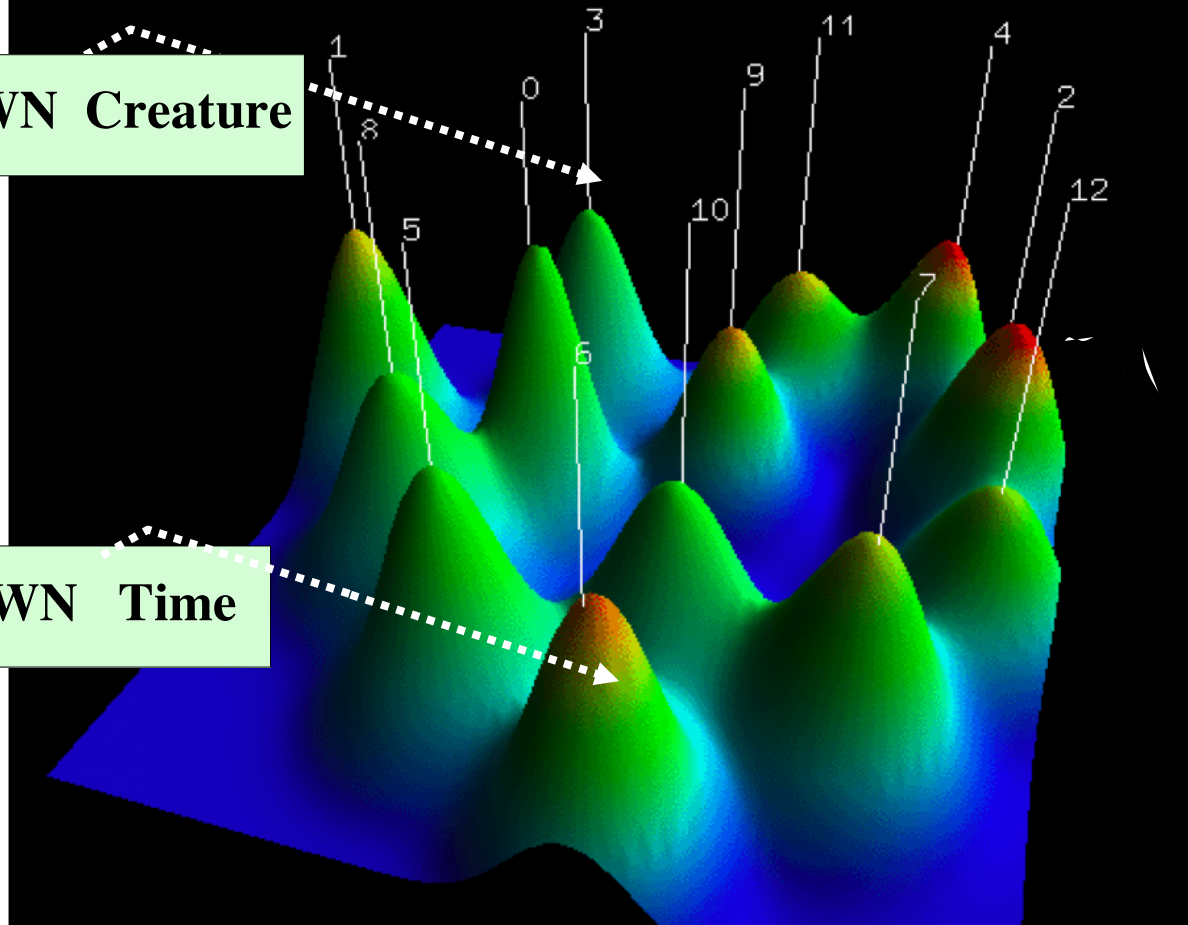


## How Well Do Mash-ups Reflect Real Category Boundaries?

13-way clustering of 214 nouns, compared to WordNet

WN Creature

WN Time



Almuhareb & Poesio (2004)

*Weak text-derived features*

~ 60,000 features for 214 nouns

Result: 0.855 cluster purity

Veale & Hao (2007 / 2008)

*Strong simile-derived features*

~ 7,300 features for 214 nouns

Result: 0.902 cluster purity

Veale & Li (2009)

*Features: Pragmatic Neighbors*

~ 8,300 features for 214 nouns

Result: 0.934 cluster purity

## Introspective Reasoning: The Value of Questions

- To Exhibit Creativity and Introspection, Computers must “ask” questions

By learning from web questions, a computer can learn to ask questions of its own

- The Web contains a great deal of explicit knowledge we don't need/want  
... and a paucity of the explicit stereotypical / normative knowledge we do want

- A computer can learn from similes, from coordinations, and from questions  
We need a way to generalize from small pockets of widely-accepted knowledge

- Creative devices like “blending” hold the key

“Big Blending” still too ambitious; conceptual mash-ups are an effective tool



*Afflatus.UCD.ie*





*Questions?*

*Fin*