Networks of concepts and ideas

Maria Petrou

Informatics and Telematics Institute,
CERTH, Thessaloniki, Greece
and
Department of Electrical and Electronic Engineering,
Imperial College London

Networks everywhere...

- The neuronal network: a physical communication network
- The metabolic network: a chemical reaction network
- The network of ideas..

The Behavioursome project

http://www.eubios.info/menmap.htm

D.R.J. Macer, "The next challenge is to map the human mind," *Nature*, vol. 402, pp 121, 2002.

- Count ideas: are they finite, uncountable or infinite?
- Can we map them and their inter-relationships?
- How are they organised?
- Do they form a network? If yes, what type?

The omnipresent of scale-free networks...

L. Barabasi, *Linked: The New Science of Networks*, Perseus Publishing, 2002.
A-L. Barabasi and E. Bonabeau, "Scale-free networks," *Scientific American*, vol. 288, pp. 50–58, 2003.

- The Internet: a scale-free network

- Social interactions: can be modelled by a scale-free network

- The organisation of living organisms: can be modelled by a scale-free network

Inspiration of many projects...

- Opte: the whole Internet mapping project
- The human brain mapping project
- Memetics: ideas=memes$\leftrightarrow$genes

If ideas are created, propagate and die, ie have a biological cycle, is it possible that their interplay is organised in a scale-free structure, like it has been shown for natural organisms?

Is it useful to understand how ideas are related?

- In cognitive systems..

- In presenting information to the user...

- In the artificial intelligence of robots that interact with humans...

- In creating ontologies that capture semiotic content...

- In human machine interfaces

Some peculiarities of the network of ideas

• It is triggered by external stimuli

• Does it matter what is the modality of the stimulus?

• Is the network triggered by a particular type of stimulus the same as the network triggered by another type of stimulus?

• If not, does the network triggered by a particular type of stimulus have the same topology as the network triggered by another type of stimulus?

If you cannot examine the Universe, check a rock...

- The enormity of the task...
- The scaled down version...

Some preliminaries

Ideas: mental conceptualisation of things, including physical objects, actions or sensory experiences, that may or may not be linguistically expressible.

Classes of ideas:

- conceptualisations of physical objects;
- psychological meanings associated with objects;
- memories;
- plans for the future;
- intentions to modify the behaviour of self;
- intentions to modify the behaviour of surrounding beings;
- processing of sensory states (e.g. pleasure, pain);
- inhibition of a response based on immediate evolutionary benefit (e.g. selfish genes or memes);
- interactive conceptualisations of ideas in a community based response;
- creativity ideas (e.g. images, plans, relations, values).

Concepts: the building blocks of knowledge, both in practical and in more abstract disciplines.

Examples:

- a definition complemented by like-this examples;
- a definition complemented by hands-on examples;
- a collection of applications;
- a list of related ideas.

Networks

Small-world network: it may contain billions of nodes, but it takes only a few intermediate nodes to move from one node to any other.

• **shortest path:** the length, in term of edges, of the shortest connection between two nodes

• **mean path:** the average of the shortest paths calculated on all pairs of nodes in the graph between which a path exists.

• **degree of a node:** the number k of incoming and outgoing links from a node

• $P(k)$: probability density function of degree k

Small-world networks: Two types:

- **random:** $P(k)$ has a *Poisson* distribution

- **scale-free:** $P(k) \sim k^{-c}$, where $c > 0$.

Examples:
The Internet: scale-free with $2 \leq c \leq 3 \Longrightarrow$ robustness and resilience to failure

Challenges of designing an experiment

• the restriction of the term "idea", to something manageable in a real-world experiment;
• the restriction of the term "concept";
• a choice of stimuli which could describe a world small enough for making the building of a network through connections a sensible operation, but also as sparse as it is needed, in order to avoid the risk of self-referencing.
• stimuli cardinality, in order to strike a balance between the subjects' freedom in the connection search process and intrinsic limits of visual and verbal memory.

Idea Restriction:

Our ideas will be objects such as tools, animals, vegetables, represented through their images and names.

Concept restriction:

We use concept as the term indicating a relationship between ideas, and specifically a pair of ideas.

Choice:

We sample the ideas from a standard database commonly used in psychological experiments on perception, the elements of which were tested for recognisability, name agreement and familiarity.

Cardinality:

We use 96 ideas, allowing the subject to create chains of connections up to a maximum 48-ideas long.

Subjects

90 subjects, in two different Universities:
Imperial College in London, UK, and
Universita degli Studi di Palermo, Italy.

gender		*mother tongue*					*age group*		
M	F	ital.	engl.	greek	chin.	other	18-24	25-35	>35
60%	40%	51%	9%	12%	12%	16%	58%	33%	9%

Ideas selected for the experiment

001	Anchor	Ancora	049	Igloo	Igloo
002	Apple	Mela	050	Kangaroo	Canguro
003	Arrow	Freccia	051	Key	Chiave
004	Backpack	Zaino	052	Funnel	Imbuto
005	Banana	Banana	053	Knife	Coltello
006	Bottle	Bottiglia	054	Leaf	Foglia
007	Bed	Letto	055	Lion	Leone
008	Bone	Osso	056	Lipstick	Rossetto
009	Book	Libro	057	Mask	Maschera
010	Broom	Scopa	058	Mirror	Specchio
011	Butterfly	Farfalla	059	Moon	Luna
012	Button	Bottone	060	Mushroom	Fungo
013	Cake	Torta	061	Pear	Pera
014	Candle	Candela	062	Pen	Penna
015	Cigarette	Sigaretta	063	Pencil	Matita
016	Carrot	Carota	064	Penguin	Pinguino
017	Crown	Corona	065	Piano	Pianoforte
018	Chair	Sedia	066	Ladder	Scala
019	Cheese	Formaggio	067	Pineapple	Ananasso
020	Church	Chiesa	068	Light-bulb	Lampadina
021	Comb	Pettine	069	Plane	Aeroplano
022	Cross	Croce	070	Puzzle	Puzzle
023	Dog	Cane	071	Pyramid	Piramide
024	Dolphin	Delfino	072	Rake	Rastrello
025	Door	Porta	073	Ring	Anello
026	Dragon	Drago	074	Robot	Robot
027	Drawer	Cassetto	075	Saddle	Sella
028	Ear	Orecchio	076	Saw	Sega
029	Elephant	Elefante	077	Scarf	Sciarpa
030	Fire	Fuoco	078	Shoe	Scarpa

031	Fish	Pesce	079	Skeleton	Scheletro
032	Flag	Bandiera	080	Slider	Scivolo
033	Flower	Fiore	081	Spider	Ragno
034	Fork	Forchetta	082	Strawberry	Fragola
035	Ghost	Fantasma	083	Sun	Sole
036	Giraffe	Giraffa	084	Table	Tavolo
037	Glasses	Occhiali	085	Tent	Tenda
038	Globe	Mappamondo	086	Tie	Cravatta
039	Frog	Rana	087	Train	Treno
040	Guitar	Chitarra	088	Tree	Albero
041	Hammer	Martello	089	Turtle	Tartaruga
042	Hand	Mano	090	Umbrella	Ombrello
043	Hat	Cappello	091	Watch	Orologio
044	Heart	Cuore	092	Well	Pozzo
045	Helicopter	Elicottero	093	Wheel	Ruota
046	Helmet	Casco	094	Whistle	Fischietto
047	Horse	Cavallo	095	Window	Finestra
048	House	Casa	096	Zebra	Zebra

Stimuli

The stimuli used in the experiment were a subset of the 520 stimuli used by the Internet Picture Naming Project http://crl.ucsd.edu/ aszekely/ipnp/

520 300 $\times$ 300 pixels black and white images depicting common objects

B. Rossion, G. Pourtois, "Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition", Perception, 33, 217-236.

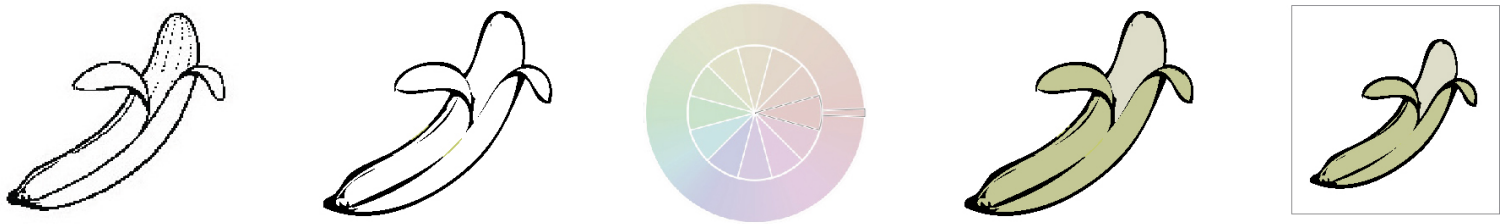
Stimuli choice criteria

- The ideas were first ranked by familiarity and name agreement, so that the ideas on top of the list were the most familiar, and the ones that people tended to define using one specific name.

- Ideas with ambiguous names were discarded from the list.

- The first 96 images left, and their corresponding words, were selected for the experiment.

The original image (a) is cleaned and vectorised (b), using colours from the subdued equi-illuminant palette (c) is coloured (d), and then resized (e).



a *b* *c* *d* *e*

Apparatus

The tests were administered in a quiet room, using a 21″ Trinitron monitor, connected to a personal computer. The software driving the tests had been programmed in-house using commonplace development tools and an object oriented language. The subject was given a mouse connected to the system, in order to execute the required task.

Procedure

The subject was introduced to the test room, and invited to sit in front of the screen. A brief data collection session, useful for statistical purposes, followed.

Procedure - Visual

• All 96 images were displayed on the monitor, arranged in a 12×8 grid.

• Their position was randomised for each subject, and was kept constant during each trial.

• Explicit remarks were made to make it clear that the test was not against the clock, and that there were no right answers per se, but the interest of the experimenters was in the subjective response.

• The subject was left alone for 5 minutes, in order to avoid perceived pressure from the experimenter.

• The experimenter returned to the room and asked the subject if any of the images represented on the screen was ambiguous, or meaningless. In case of an affirmative answer, the experimenter gave the subject clarifications about the ambiguous image, avoiding explicitly naming the idea.

• The computer chose a random object from which to start the connection building process, and the experimenter invited the subject to pinpoint the object that they felt being more similar to the one selected, and click on it using the mouse.

• The previously selected object disappeared, and the newly clicked object was selected. The experimenter asked the subject to repeat the task, now starting from the newly selected object, again and again.

• When 48 of the 96 objects had been selected, a greeting message was displayed, and the test ended. The software automatically recorded all the links created by the subject, and the total elapsed time for the linking process, for statistical purposes.

• The subject was reminded of the fact that a second session of the experiment would be held in the near future.

Two months later: Procedure - Verbal

The procedure was almost the same, except:

- The grid was now 6×16, and the names representing the ideas were presented black on white, in a sans serif font.

- After the familiarisation phase, the experimenter asked the subject to scan the grid for unknown names, and in case some was found, explained which idea the name represented, avoiding using circular definitions and other names already present in the grid.

For each subject, the starting ideas used in the verbal experiment were kept the same as the ones used in the visual experiment.

visual grid

Elephant	Mask	Banana	Cake	Ear	Leaf
Fork	Book	Pen	Fish	Puzzle	Key
Knife	Flower	Dolphin	Fire	Tree	Comb
Whistle	Apple	Bed	Door	Plane	Strawberry
Moon	Glasses	Helicopter	House	Candle	Tie
Butterfly	Hand	Igloo	Ring	Lion	Penguin
Mushroom	Watch	Chair	Flag	Pencil	Table
Hat	Heart	Hammer	Button	Slider	Robot
Crown	Window	Tent	Giraffe	Zebra	Anchor
Drawer	Piano	Helmet	Wheel	Dragon	Church
Rake	Cigarette	Scarf	Ghost	Skeleton	Spider
Well	Saddle	Bone	Carrot	Pinapple	Backpack
Umbrella	Lightbulb	Turtle	Frog	Pyramid	Horse
Train	Broom	Pear	Shoe	Saw	Cross
Arrow	Mirror	Globe	Lipstick	Sun	Dog
Cheese	Kangaroo	Guitar	Ladder	Funnel	Bottle

verbal grid

Networks of ideas

• Each idea is a node.

• The links between nodes are representations of the connections created by the experimental subjects during the trials.

• The network is built by adding to the graph all of the connections made by all subjects.

• The first object of any trial, being randomly chosen, is not included in the network.

Network analysis: a small-world

The mean shortest path m for both the visual and verbal networks of ideas, the standard deviation of their distribution, and for comparison, $\log(n)$ (where n is the number of nodes of the network)

	visual network	verbal network
m	1.79	1.78
σ	0.45	0.44
$\log(n)$	4.56	4.56

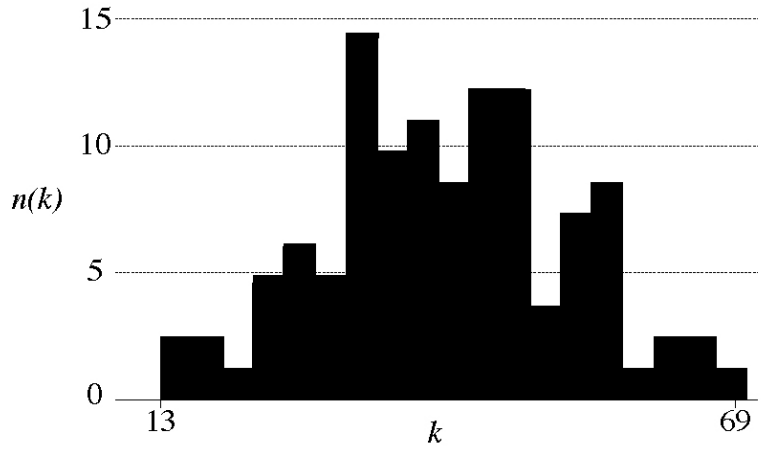
Random or scale-free?

For each network, we counted the number of edges connected to each node, and then analysed the resulting frequency distribution using the standard *chi-square goodness-of-fit test* with parameter estimation.

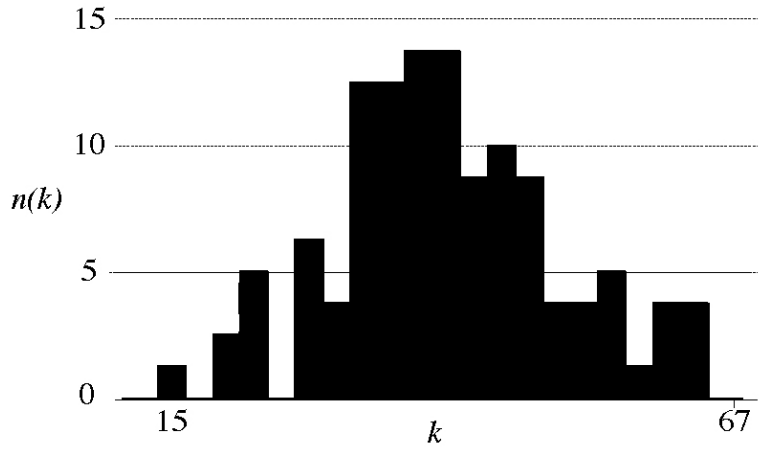
•: **Hypothesis** $H1$**:** the observed distribution is a power law: $P_2(k) = \frac{k^{-c}}{c-1}$, with $c > 1$, typical of scale-Free networks,

•: **Hypothesis** $H2$**:** the observed distribution is a Poisson distribution, i.e. $P_1(k) = \frac{e^{-\lambda}\lambda^k}{k!}$, where λ is the mean value of k, characteristic of Random networks.

Histograms of the frequency distribution of k for both the visual and the verbal networks of ideas.



visual network



verbal network

Results of the chi-square goodness-of-fit test on the networks of ideas

visual network of ideas			
Hypothesis	Score	Threshold	Hypothesis rejected
H1(Scale-free)	73.4	12.83	Yes
H2(Random)	18.54	19.02	No
verbal network of ideas			
Hypothesis	Score	Threshold	Hypothesis rejected
H1(Scale-free)	102.5	12.83	Yes
H2(Random)	18.89	19.02	No

At the 90% confidence level the networks of ideas we have built are characterised by Poissonian distributions of the nodes, typical of **random networks**.

Most connected nodes of the networks of ideas

visual network	k	**verbal network**	k
Tree	69	**House**	67
Sun	67	**Sun**	63
House	63	**Table**	60
Fire	62	**Fire**	60
Hand	61	**Hand**	58
Window	60	Glasses	58
Door	59	**Tree**	57
Skeleton	58	**Window**	56
Horse	57	**Horse**	56
Table	56	Chair	56
Apple	56		

Node measurements for the subnetworks of ideas

	UK visual subnetwork	**UK verbal subnetwork**
m	2.07	1.99
σ	0.58	0.54
$\log(n)$	4.56	4.56
	Italian visual subnetwork	**Italian verbal subnetwork**
m	2.05	1.98
σ	0.57	0.53
$\log(n)$	4.56	4.56

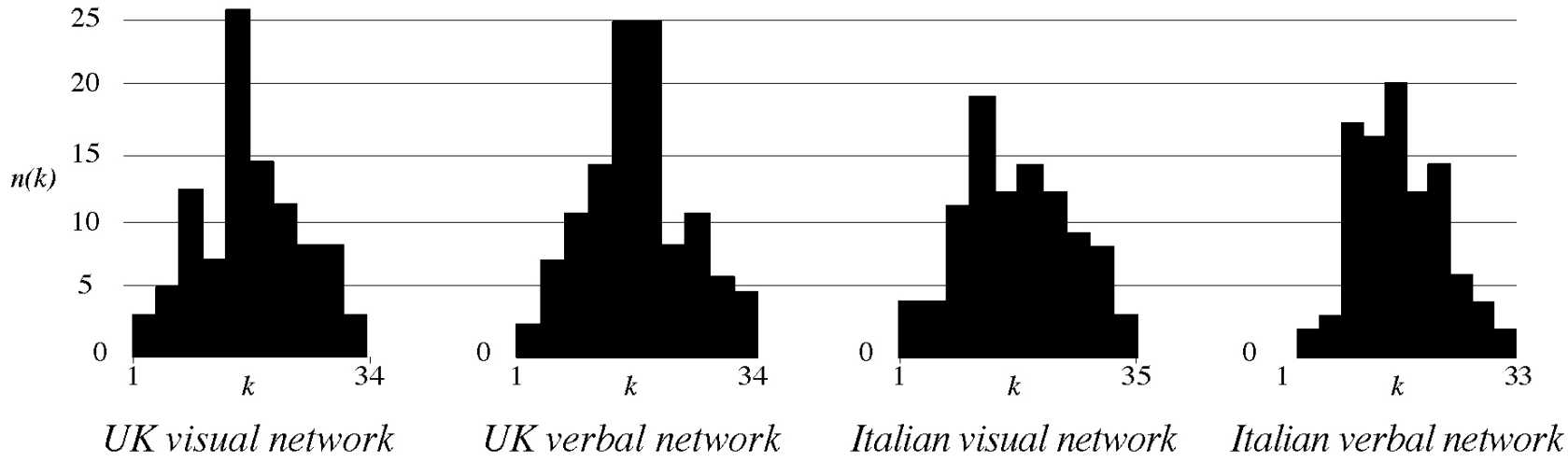
Results of the chi-square test on the subnetworks of ideas

UK visual subnetwork			
Hypothesis	Score	Confidence	Hypothesis rejected
H1(Scale-free)	80	95%	Yes
H2(Random)	17.0	90%	No
UK verbal subnetwork			
Hypothesis	Score	Confidence	Hypothesis rejected
H1(Scale-free)	98	95%	Yes
H2(Random)	14.2	90%	No
Italian visual subnetwork			
Hypothesis	Score	Confidence	Hypothesis rejected
H1(Scale-free)	107	95%	Yes
H2(Random)	21.4	80%	No
Italian verbal subnetwork			
Hypothesis	Score	Confidence	Hypothesis rejected
H1(Scale-free)	102	95%	Yes
H2(Random)	24.3	80%	No

Most connected nodes of the subnetworks of ideas

UK vis. sub.		**UK verb. sub.**		**Ital. vis. sub.**		**Ital. verb. sub.**	
34	Tree	34	House	35	Tree	33	House
32	Sun	33	Sun	35	Sun	32	Fire
32	House	33	Hand	33	Fire	30	Sun
31	Hand	32	Window	32	Lion	30	Glasses
30	Window	31	Table	32	Horse	29	Table
30	Table	31	Moon	32	Door	29	Horse
30	Leaf	31	Door	31	House	28	Chair
30	Apple	30	Tree	30	Window	28	Book
29	Skeleton	29	Candle	30	Hand	27	Tree
29	Fork	28	Glasses	30	Drawer	26	Zebra
29	Fire	28	Fire	30	Cross	26	Globe
		28	Chair			26	Bed

Frequency distributions of k for the subnetworks



Discussion concerning the networks of ideas

• There is about 80% overlap in the hubs of the two networks.

• The topologies of the subnetworks of UK and Italian subjects are pretty consistent with each other and with the topologies of the full networks.

• The overlapping of the hubs between the two visual subnetworks and between the two verbals subnetworks is about 50%, in all pair-wise comparisons we can make.
$\Rightarrow$ The difference between the hubs in the visual and the verbal networks is not an indication of two different networks. In the full networks the overlap is about 80%.

Conclusion concerning the networks of ideas

Our results indicate the presence of a single network of ideas, of random network topology, with hubs that may be invoked either verbally or visually.

Networks of concepts

• nodes: representations of concepts, i.e. relationships between ideas

• edges: consequential connections between concepts.

• An oriented graphs. Each node of the graph is a specific concept relating an idea with another, represented by a pair (a, b), in which a is the connecting idea and b is the connected idea. Each edge in the graph connects two concepts if those two concepts have been selected in a temporal succession, and has the direction of the temporal flux.

• The networks are built by adding to the graph each connection made by each subject.

The visual network of concepts

Network analysis

They are both small-world networks with similar topologies.

The mean shortest path m for both the visual and verbal networks of concepts, the standard deviation of their distribution, and for comparison, $\log(n)$ (where n is the number of nodes of the network)

	visual network	verbal network
m	10.15	10.49
σ	3.43	3.59
$\log(n)$	6.62	6.63

Random or scale-free?

At the 95% confidence level, both our connectivity distributions are compatible with a power law distribution with the c value between 2.5 and 2.8 for the visual network of concepts and between 2.6 and 2.7 for the verbal network of concepts. Such a c value is in the range of a proper scale-free network.

A result of chance?

• A Monte Carlo simulation of our experiment using the same parameters as in the original experiment, namely a repetition of 90 trials using 96 ideas and linking 48 ideas in each test, this time choosing the links at random using a pseudo-random number generator.

• We repeated this simulation 90 times, always obtaining distributions of the k values that did not resemble at all the power law distribution.

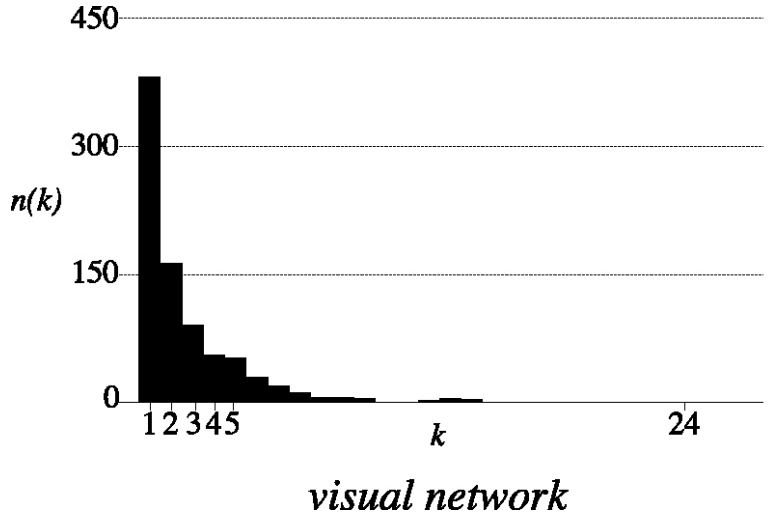
Tests for robustness

• A *leave-one-out* test on the set, repeating our analysis 90 times, each time excluding the set of links corresponding to one subject from the test. In each instance we obtained a scale-free network, with the same relationships as the ones in the original experiment acting as hubs.

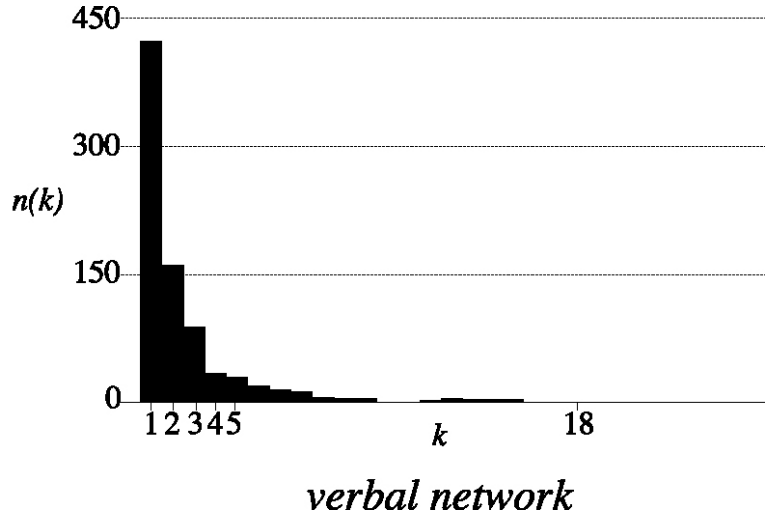
• We performed a Monte Carlo variation of the test, leaving out 5 randomly chosen subjects' sets of choices each time, and repeating this test 90 times.

All of the networks built in such a way were characterised by power law distributions of the node connectivities, and, compared with the original experiment, there were only marginal differences at the lower end of the hub list.

Histograms of the values of k for both the visual and the verbal networks of concepts.



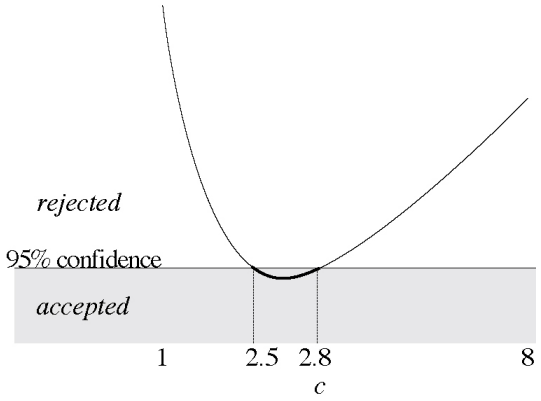
visual network



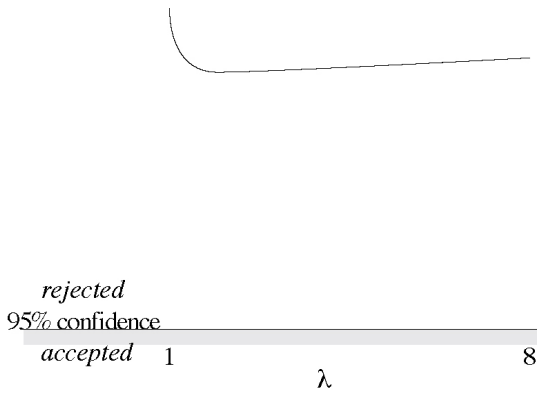
verbal network

The results of the goodness-of-fit tests for both the visual and verbal degree distributions.

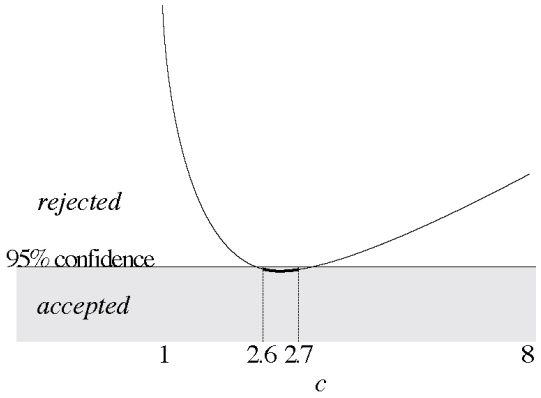
Along the horizontal axis we measure the parameter of the distribution for which the test is done, while along the vertical axis we measure the confidence level of the test.



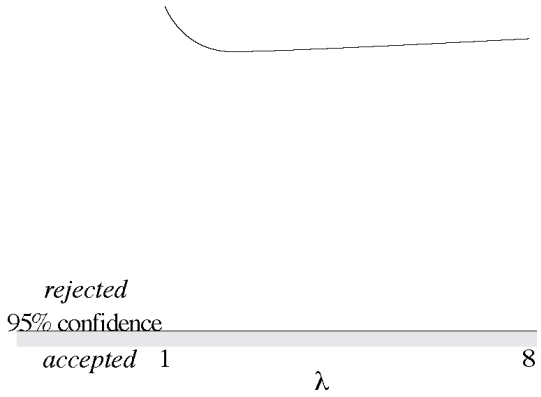
H1 - visual network



H2 - visual network



H1 - verbal network



H2 - verbal network

Hubs and the role of visual and verbal cues

Nodes represent relationships $\Rightarrow$ the hubs are special relationships which are connected to a variety of other relationships.

From the point of view of mind processes: the relationships that play the role of hubs may be the focal points from which many diverse scenarios emanate.

The logical path in the mind is not a sequence of individual ideas, but a sequence of relationships!

Hubs of the networks of concepts

visual network	*k*	*verbal network*	*k*
(knife,fork)	24	**(horse,saddle)**	18
(tree,leaf)	20	(sun,moon)	17
(horse,saddle)	18	**(tree,leaf)**	15
(door,window)	17	(pencil,pen)	14
(church,cross)	16	(plane,helicopter)	14
(table,chair)	16	**(knife,fork)**	13
(leaf,tree)	15	**(church,cross)**	13
(igloo,penguin)	15	**(door,window)**	11
(cross,church)	15	(glasses,mirror)	11
(pen,pencil)	14	(fork,knife)	11
(fork,knife)	14	**(table,chair)**	11
(window,door)	14		

Looking closer...

The difference between the hubs in the two halves of the same experiment is roughly the same as the difference between the hubs of the verbal and visual experiments.

This does not allow us to decide whether the 50% overlap is an indicator of two separate networks, or it is simply due to the differences arising from the randomness of the processes involved.

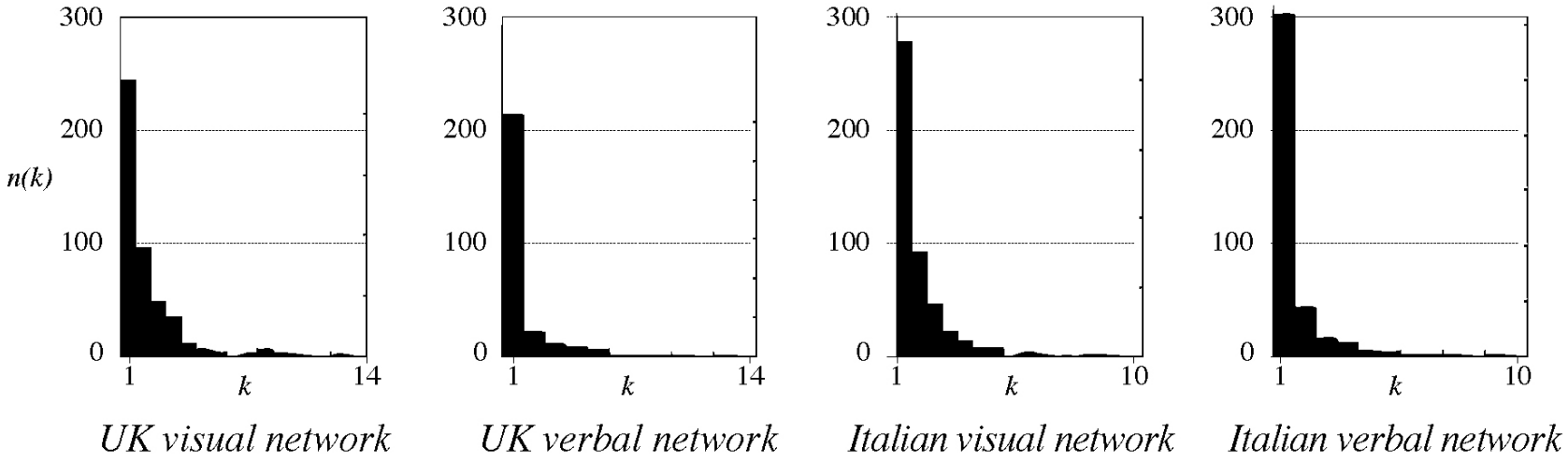
Node measurements for the subnetworks of concepts

	UK visual subnetwork	**UK verbal subnetwork**
m	10.59	4.88
σ	6.60	3.03
$\log(n)$	6.05	5.89
	Italian visual subnetwork	**Italian verbal subnetwork**
m	6.08	5.56
σ	3.61	3.67
$\log(n)$	6.00	6.00

Results of the chi-square test on the subnetworks of concepts

UK visual subnetwork					
Hyp.	Parameter	Score	Threshold	Confidence	Hyp. reje
H1(Scale-free)	2.1	14.03	14.45	90%	No
H2(Random)	4	80.00	12.80	90%	Yes
UK verbal subnetwork					
Hyp.	Parameter	Score	Threshold	Confidence	Hyp. reje
H1(Scale-free)	1.2	13.60	12.80	85%	No
H2(Random)	5	51.33	9.34	90%	Yes
Italian visual subnetwork					
Hyp.	Parameter	Score	Threshold	Confidence	Hyp. reje
H1(Scale-free)	1.8	13.19	14.06	95%	No
H2(Random)	2	43.99	11.01	90%	Yes
Italian verbal subnetwork					
Hyp.	Parameter	Score	Threshold	Confidence	Hyp. reje
H1(Scale-free)	1.5	10.42	12.59	95%	No
H2(Random)	3	32.14	9.38	90%	Yes

Frequency distributions of k for the subnetworks



Hubs of the subnetworks of concepts

UK vis. sub.		**Ital. vis. sub.**	
14	(knife,fork)	10	(knife,fork)
11	(tree,leaf)	10	(window,door)
10	(door,window)	9	(horse,saddle)
10	(fork,knife)	9	(pen,pencil)
9	(leaf,tree)	9	(tree,leaf)
9	(horse,saddle)	9	(church,cross)
9	(igloo,penguin)	9	(cross,church)
9	(strawberry,apple)	8	(table,chair)
8	(saw,hammer)	8	(skeleton,bone)
8	(table,chair)	8	(pear,apple)

Hubs of the subnetworks of concepts

UK verb. sub.		**Ital. verb. sub.**	
14	(sun,moon)	10	(horse,saddle)
9	(candle,fire)	9	(church,cross)
8	(plane,helicopter)	8	(tree,leaf)
8	(knife,fork)	8	(fork,knife)
8	(horse,saddle)	7	(pencil,pen)
7	(door,window)	7	(leaf,flower)
7	(pencil,pen)	6	(chair,table)
7	(glasses,mirror)	6	(hand,ring)
7	(tree,leaf)	6	(train,plane)
7	(door,house)	6	(plane,helicopter)
7	(zebra,giraffe)		

Conclusions

We have evidence that:

- while networks of ideas show a *random* topology, networks of concepts show a *scale-free* topology, irrespective of the fact that they are built using visual or verbal cues, as made evident by their distributions of the number of links per node;

- there is a syntactic correspondence in the topology of the visual and verbal networks of concepts, the two networks being statistically equivalent in topology;

- there is a possible semantic difference between the visual and verbal networks of concepts, indicated by the partial difference in the lists of concepts that act as hubs in the two networks.

M Petrou, M E Tabacchi and R Piroddi, 2010. "Networks of Concepts and ideas", The Computer Journal, Vol 53(10), pp 1738–1751
DOI:10.1093/comjnl/bxp113.