Digging by Debating (DbyD):
From Big Data Text Repositories to Argument Analysis

➢ **The InPhO Group**, Indiana University**:** Colin Allen with Robert Rose, Jaimie Murdock, Jun Otsuka, Doori Lee, Indraneel Dey

➢ **The Cyberinfrastructure Network Science Center**, Indiana University: Katy Börner with Robert Light

➢ **The International Centre for Public Pedagogy (ICPuP)**, University of East London (UEL): Andrew Ravenscroft with Simon McAlister

➢ **ARG:dundee**, University of Dundee: Chris Reed with John Lawrence

➢ **Centre for Digital Philosophy**, University of Western Ontario**:** David Bourget

From Big Data to Argument Analysis

Linking massive datasets to specific arguments, where 'text is data'

Project Goals

- Uncover and represent the key argumentative structures of digitized documents from a large philosophy/science corpus;
- Allow users to find and interpret detailed arguments in the broad semantic landscape of books and articles, and to support innovative interdisciplinary research and better-informed critical debates

Data Sources

Stanford Encyclopedia of Philosophy, HathiTrust Collection, PhilPapers

4 Levels of Analysis: Macro (Sci/Phil Maps) to Micro (detailed arguments)

1. *Visualizing* points of contact between philosophy and the sciences
2. *Topic modeling* to identify the volumes/pages 'rich' in a chosen topic;
3. Identify and map *key arguments*; apply a novel analysis framework for propositions and arguments;
4. *Sentence modeling* to get back to HathiTrust materials

Mapping Philosophy in the Sciences
(analysis level 1)
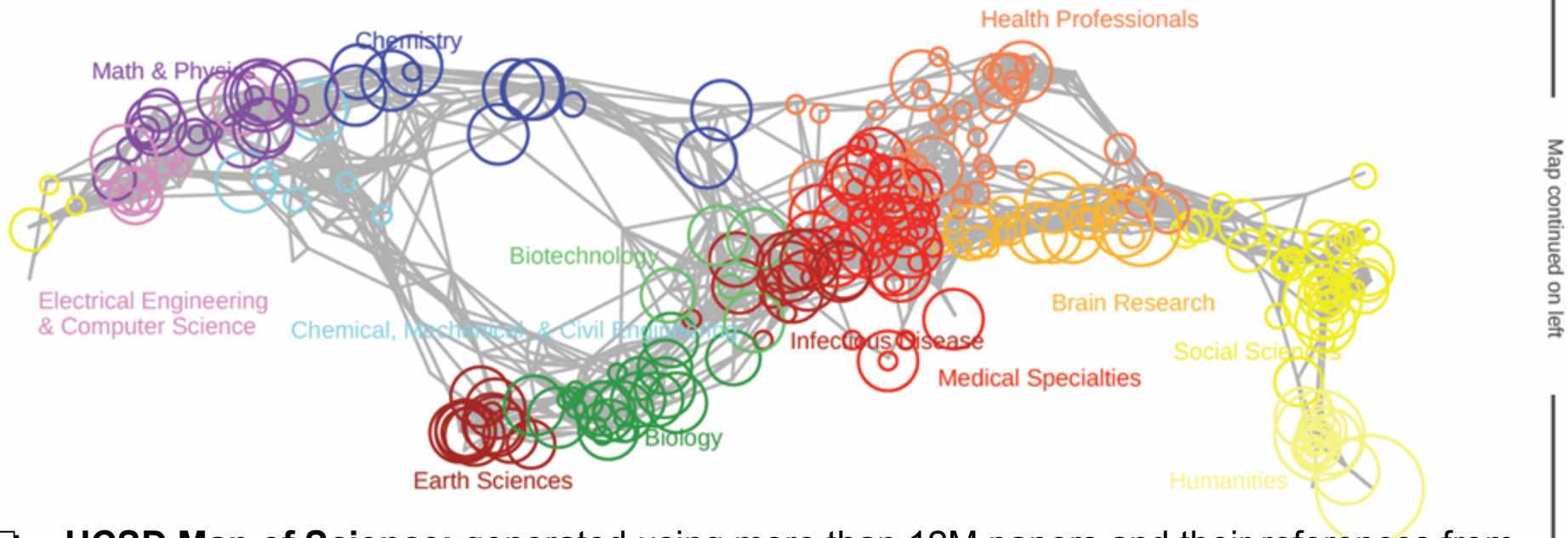
- ❏ **UCSD Map of Science:** generated using more than 12M papers and their references from Elsevier's Scopus and Thomson Reuters' Web of Science (25,000 journals), see http://sci.cns.iu.edu/ucsdmap
- ❏ Shows **554 subdiscipline nodes** aggregated into **13 color-coded disciplines**.
- ❏ Overlaid are citations made by the **Stanford Encyclopedia of Philosophy to** visualize the impact of the sciences on philosophy.
- ❏ **Node sizes** scale from 0 (no circle) to 43. Highest numbers are in **Humanities, Earth Sciences, and Math & Physics.**

Topic Modeling in HT Books
(analysis level 2)

- LDA Topic Modeling: Bayesian method generates set of **"topics"** – **probability distributions** over terms in the corpus
  - **Every topic contains every term** – different probabilities in the different topics
  - The **number of topics** is a user-selected parameter
- Finds the set of topics best able to reproduce term distributions in the **documents**
- Documents may be whole **volumes**, **chapters**, **articles**, single **pages**, even individual **sentences** – modelers' choice

```
In [36]:  # finding related topics using multiple terms
          v1.sim_word_top(['anthropomorphism', 'animal', 'psychology'])
Out[36]:
```

	Sorted by Word Similarity
Topic	**Words**
26	consciousness, experience, p, psychology, process, individual, object, activity, relation, feeling
16	animals, evolution, life, animal, development, man, species, cells, living, theory
10	animals, water, animal, food, birds, one, leaves, insects, species, many
1	pp, der, vol, die, de, des, und, ibid, university, la
58	man, among, tribes, primitive, men, people, also, races, women, race
47	college, university, professor, school, law, work, students, degree, education, new
25	nature, ii, us, mr, without, life, human, natural, language, every
29	fig, two, body, form, cells, animals, first, ii, side, organs
12	child, children, first, development, movements, play, life, little, mental, mother

```
In [38]:  # showing documents by combined topics
          v1.sim_top_doc([10,16,26], print_len=20, label_fn=htrc_label_link_fn_1315)
Out[38]:
```
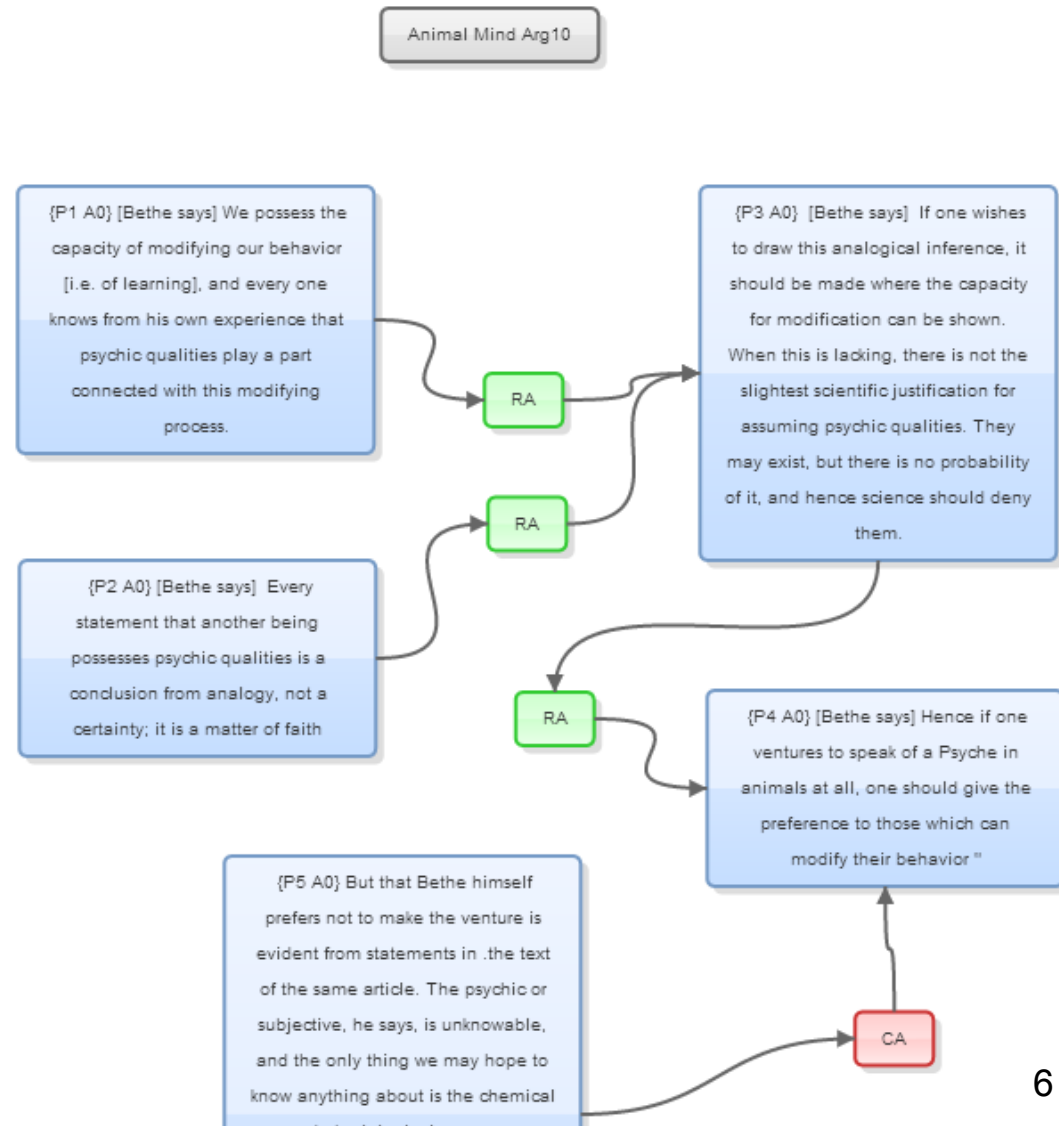
Topics: 10, 16, 26	
Document	**Prob**
Secrets of animal life,, http://hdl.handle.net/2027/uc2.ark:/13960/t7wm15g73	0.63954
Comparative studies in the psychology of ants and of higher , http://hdl.handle.net/2027/uc2.ark:/13960/t6057f659	0.63085
The colours of animals, their meaning and use, especially co, http://hdl.handle.net/2027/uc2.ark:/13960/t9t14w82w	0.55333
The foundations of normal and abnormal psychology., http://hdl.handle.net/2027/loc.ark:/13960/t9m33nm99	0.54171
The bird rookeries of the Tortugas., http://hdl.handle.net/2027/uc2.ark:/13960/t3pv6cc9j	0.53789
Mind in animals., http://hdl.handle.net/2027/mdp.39015005169357	0.53783
Ants and some other insects; an inquiry into the psychic pow, http://hdl.handle.net/2027/wu.89095158218	0.53606
Systematic science teaching ; a manual of inductive elementa, http://hdl.handle.net/2027/uc2.ark:/13960/t11n8195t	0.53152
The riddle of the universe at the close of the nineteenth ce, http://hdl.handle.net/2027/uc2.ark:/13960/t5v69b880	0.52804
The riddle of the universe at the close of the nineteenth ce, http://hdl.handle.net/2027/loc.ark:/13960/t2s47h57b	0.52763
The riddle of the universe at the close of the nineteenth ce, http://hdl.handle.net/2027/loc.ark:/13960/t2s47h57b	0.52035
The colour-sense: its origin and development. An essay in co, http://hdl.handle.net/2027/mdp.39015009245195	0.51658
Advanced elementary science;, http://hdl.handle.net/2027/uc2.ark:/13960/t8z895h4k	0.50012
Report of the Brown-Harvard Expedition to Nachvak, Labrador,, http://hdl.handle.net/2027/uc2.ark:/13960/t41r70n1f	0.49748
The germ-plasm : a theory of heredity /, http://hdl.handle.net/2027/uc2.ark:/13960/t0qr4pc9z	0.49713
Biology and its makers, with portraits and other illustratio, http://hdl.handle.net/2027/uc2.ark:/13960/t1td9pw8v	0.49495
Report of the Brown-Harvard expedition to Nachvak, Labrador,, http://hdl.handle.net/2027/loc.ark:/13960/t4vh6bb8k	0.49243
Last words on evolution; a popular retrospect and summary,, http://hdl.handle.net/2027/nyp.33433081629184	0.48496
Thought and things; a study of the development and meaning o, http://hdl.handle.net/2027/nc01.ark:/13960/t9x06sq8q	0.48331
The psychic life of micro-organisms. A study in experimental, http://hdl.handle.net/2027/uc2.ark:/13960/t73t9h556	0.47866

4

Topic Modeling in HT Books
(analysis level 2)

- LDA Topic Modeling: Bayesian method generates set of **"topics"** – **probability distributions** over terms in the corpus
  - **Every topic contains every term** – different probabilities in the different topics
  - The **number of topics** is a user-selected parameter
- Finds the set of topics best able to reproduce term distributions in the **documents**
- Documents may be whole **volumes**, **chapters**, **articles**, single **pages**, even individual **sentences** – modelers' choice

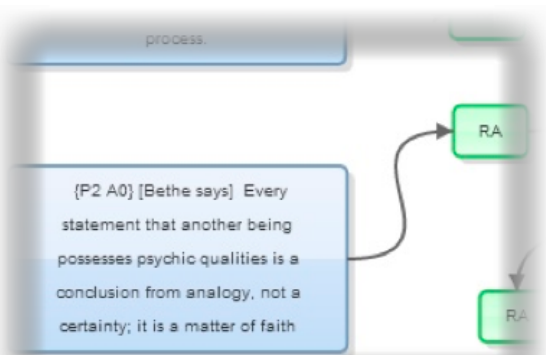# Argument Selection, Mapping and Analysis (analysis level 3)

1. Identifying arguments from rated pages (currently human/ manual, but developing algs for automation)

2. Mapping of key arguments with OVA mapping tool:

- Provides a formal framework, or 'lens', for investigation and comparative analysis, e.g. role, structure, status

- More indirectly: meaning, importance and context

Animal Mind Arg10

{P1 A0} [Bethe says] We possess the capacity of modifying our behavior [i.e. of learning], and every one knows from his own experience that psychic qualities play a part connected with this modifying process.

{P3 A0} [Bethe says] If one wishes to draw this analogical inference, it should be made where the capacity for modification can be shown. When this is lacking, there is not the slightest scientific justification for assuming psychic qualities. They may exist, but there is no probability of it, and hence science should deny them.

{P2 A0} [Bethe says] Every statement that another being possesses psychic qualities is a conclusion from analogy, not a certainty; it is a matter of faith

RA

RA

RA

{P4 A0} [Bethe says] Hence if one ventures to speak of a Psyche in animals at all, one should give the preference to those which can modify their behavior "

{P5 A0} But that Bethe himself prefers not to make the venture is evident from statements in .the text of the same article. The psychic or subjective, he says, is unknowable, and the only thing we may hope to know anything about is the chemical

CA

Back to the Sentences
(analysis level 4)

['every', 'statement', 'another',
'possesses', 'psychic','qualities',
'conclusion', 'analogy',
'certainty', 'matter', 'faith']

Topic	Words
	Topics Sorted by Index
0	j, l, f, e, cole, w, bentley, uexkull, loeb, akad
1	sense, smell, hearing, taste, sight, organ, insects, fishes, sensory, study
2	psychology, comparative, animal, experimental, mind, human, study, writers, psychologist, upon
3	process, may, become, food, many, long, reaction, used, reduced, state
4	h, c, g, f, e, m, b, j, parker, o
5	body, surface, animal, food, contact, occur, solid, end, water, two
6	experience, learning, animal, memory, form, individual, ideas, rapid, idea, learn
7	vol, physiol, ibid, psych, nat, sci, studies, n.s, science, arts
8	cilia, current, paramecium, water, swim, bodies, reaction, fact, jennings, animals
9	animal, method, behavior, interpretation, methods, error, facts, anecdote, trial, experiment
10	p, 1, 15, 8, 10, 4, ibid, 5, 13, 23
11	sensation, different, sensations, may, consciousness, qualities, experience, quality, animal, human
12	p, 17, 14, 16, 21, 22, ibid, 20, 12, 18
13	pp, two, seen, already, observation, pages, ways, may, dark, rapidly
14	water, sounds, case, produced, response, current, vibrations, respond, electric, react
15	side, direction, toward, would, one, opposite, move, object, moved, fish
16	effect, upon, body, movement, like, light, position, sides, influence, motor
17	make, conditions, tell, find, behavior, study, observation, ordinary, good, careful
18	placed, food, one, containing, side, fish, water, either, tank, forceps
19	jour, neur, trans, leipzig, abth, und, wiss, wien, abstract, st
20	bees, hive, flowers, visual, memory, bethe, back, insects, explain, smell
21	reaction, negative, stimulus, mechanical, contact, may, stimulation, positive, form, response
22	human, experience, animals, power, would, ideas, know, beings, mind, difference
23	reaction, substances, stimuli, also, edible, gonionemus, chemical, reactions, hand, metridium
24	evidence, must, distance, us, use, give, could, learning, experience, think
25	animals, processes, mental, methods, visual, life, form, test, discrimination, phenomena
26	r, c, acad, w, n, m, soc, v, e, yerkes
27	number, found, place, romanes, years, series, mollusk, takes, cause, compartment
28	wasp, nest, wasps, one, solitary, new, least, locality, flight, prey
29	sensation, sense, may, specific, quality, special, reactions, accompanied, organ, organs

Back to the Sentences
(analysis level 4)

Stanford Encyclopedia of P × IPy IPython Dashboard × IPy InPhO–HTRC–Sept 2013 ×

https://inphodata.cogs.indiana.edu:8883/209646c0-3b60-470d-87d7-b7113532774d#

Bochum   BF   LAZ   Home Stuff   SEP/Coding   Apple   COGS   Animal Minds   Funding   News   Bikes/Races   »   Other Bookmarks

IP[y]: Notebook   InPhO-HTRC-Sept 2013   Last saved: Sep 12 11:11 PM   Logout

File   Edit   View   Insert   Cell   Kernel   Help

Code
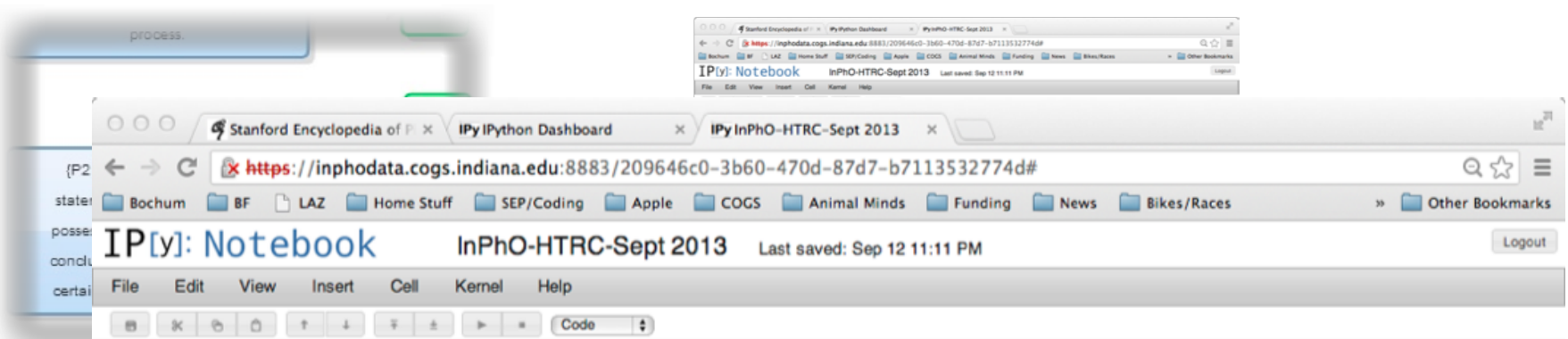
What did we get? -- tokenized sentences (word lists) followed by original text...

In [22]: #tok_sents

In [23]: orig_sents

Out[23]: ['Every statement that another being possesses \n psychic qualities is a conclusion from analogy, not a certainty; it is a matter of faith.',
          "If any consciousness \n accompanies it, then the nearest human analogy to such \n consciousness is to be found in organic sensations, and these, \n as has just been said, must necessarily be in the human mind \n wholly different in quality from anything to be found in an \n animal whose structure is as simple as the Amoeba's.",
          '; learning, 208, \n \n 214.',
          'On the other \n \n1).',
          'Dytiscus, 86.',
          'Burnett, 126, 170.',
          'Willem, 130, 192.',
          'Caterpillars, 192, 196.',
          'Murbach, 107, 158.',
          'Fancy, for example, one of us entering a \n room in the dark and groping about among the furniture.',
          'This, of \n course, does not refer to the power to judge distance.',
          'Again, a bodily structure entirely unlike our own \n must create a background of organic sensation which renders \n the whole mental life of an animal foreign and unfamiliar to \n us.',
          'She disposes of the psychic learning by experience theory of Nagel \n by saying that the only experience upon which the animal \n could reject the filter paper must be experience that it is not \n good for food.',
          'We speak, for example, of an "angry" wasp.',
          'All traces of the "learning" \n \ne "learning" \n \n \n give evidence of the power to judge distance.']

Not great! Not bad?!

Digging by Debating

DbyD: Conclusions so far...

Project Achievements - *proof-of-concept and loose integration of key components:*

1. Method for visualizing points of contact between philosophy and sciences

2. Method for Text selection from Big Data using multi-scale modeling techniques

3. Identified, represented and mapped key arguments about topics (OVA) and devised a novel framework for investigation and comparative analysis

4. Used sentence-level Topic Modeling to 'go back' to the texts to find similar propositions to those mapped (investigating context)

Future work

- From loose integration to usable tools, e.g. linked to Philpapers
- To understand and incrementally construct argument maps (mark-up interface) AND automated extraction and mapping
- Funding from future DiD or similar Big Data initiatives ('open data'?)

diggingbydebating.org

9