Digital Humanities as Empirical Philosophy of Science

XPhiSci 2019, Aarhus, 15.10.2019

Charles H. Pence, Oliver Lean, and Luca Rivelli
@pencelab

UCLouvain
Institut supérieur de philosophie (ISP)

The Pence Lab

Outline

1. The empirical turn in philosophy of science
    a. An analogy with X-Phi?
2. Focusing on the right challenge
3. Examples from textual analysis
    a. Biodiversity and conceptual analysis
    b. Binding specificity and tacit understanding
4. Coda: The role of text in science

The take-home: Philosophers of science looking to use digital methods should follow X-Phi's example, reflect on metaphilosophical commitments!

Empirical Philosophy

The Growth of "Empirical Philosophy"

Combination of a number of trends:

- Interdisciplinary collaboration
- Co-authoring and collaboration in philosophy
- New sources of data, experimental techniques

Overall: **massive growth in use of empirical methods and data for philosophical analysis.**

Empirical Philosophy

Most mature example: **experimental philosophy**

Stable and fruitful relationship with a variety of "nearby" disciplines:

- experimental psychology
- moral psychology
- traditional ethics

Empirical/experimental philosophy of science is a relative newcomer!

Whence our Empirical Data?

One obvious question: What are the "nearby" disciplines for experimental philosophy of science?

- Some methods shared with traditional X-Phi (surveys, analysis of processes of reasoning)
- Normative interventions in science itself
- Scientometrics
- Metascience

Metascience

1930s: CW Morris. Semiotics as Metascience

1938: Charles W. Morris,
*International Encyclopedia of Unified Science*:

- "Science, as a body of signs with certain specific relations to one another, to objects, and to practice, is at once a language, a knowledge of objects, and a type of activity. The interrelated study of syntactics, semantics, and pragmatics of the language of science in turn constitutes **metascience**—the science of science"
- "In this way, and on a comprehensive scale, science is made an object of scientific investigation. Metascience appears both as a tool for, and as an element within unified science (…) it itself is open to the same false leads and growth and revisions as are the other sciences."

8

1930s: CW Morris. Semiotics as Metascience

- Morris' *metascience* is the application of his *semiotics*, comprising the study of syntax, semantics and of *pragmatics,* to the language of science, by virtue of the fact that science makes use of and expresses its results in terms of *signs.*
- Morris' semiotical metascience comprises a syntactic, a semantic, and a *pragmatic* point of view on the language of science, the latter resembling an *empirical* point of view on scientific *practice*: "Here belong the problems as to how the scientist operates, the connection of science as a social institution with other social institutions, and the relation of scientific activity to other activities. (1938b)"

2000-today: metascience reappears

Metascience is today a lively field!
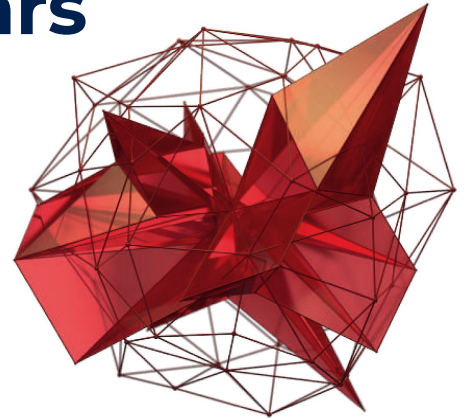
METASCIENCE
2019 The Emerging Field of Research
on the Scientific Process.
SYMPOSIUM

September 5th-8th, 2019
Stanford University | Cubberley Auditorium

www.metascience2019.org

2000-today: metascience reappears

Metaresarch: metascience and research methods.

2018: Enserink, in "Research on Research", in *Science*:

- Metaresearchers aim "to make research more robust and efficient, from preregistering studies and establishing reporting standards to the recent push to make study data freely available for others to explore."
- "whether the reforms actually work has become a study object in itself."

11

2000-today: metascience reappears

2019: Babic et al, "Incidences of Problematic Cell Lines Are Lower in Papers That Use RRIDs to Identify Cell Lines." *ELife 8*:

- They try to check if, as they suppose, the use of RRIDs, by alerting researcher that their cell lines are on a known list of contaminated ones, actually reduces the prevalence of misidentified and contaminated cell lines in the produced literature.
- To test the hypothesis, they text-mined the methods sections of about two million papers in PubMed Central, to identify in them unique cell-line names". Results confirm the hypothesis.

Metascience and Philosophy of science

- Metascience is based on empirical observation of scientific practice

- It purports to have a normative (methodological) role toward scientific practice.

- It makes use of quantitative scientific methods. It is producing its specialized methods, "metaresearch", **including *massive automated literature-mining* techniques** to draw conclusions about the state of scientific research and on the efficacy of some of its normative proposals.

Metascience and Philosophy of science

As philosophers of science, we should keep a well-focused eye on metascience, for upon its results we could build important philosophical reflections.

Metascience can become for experimental philosophy of science something akin to what psychological research is for X-Phi.

# The Challenge: Using Data Analysis for Empirical Philosophy of Science

The Challenge for Digital Philosophy of Science

Question: *How can we use digital analysis of scientific literature to inform discussion in the philosophy of science?*

- X-Phi has developed a range of methodologies for using empirical work to inform and constrain philosophical discourse
- So X-Phi's development potentially holds lessons for how to develop an analogous methodology for digital Phil Sci

X-Phi and metaphilosophy: The role of intuitions

- Psychological experiment can yield descriptive results about people's intuitions about knowledge, moral responsibility, etc.
  - and how these intuitions vary with context, culture, gender, etc.
- But (some) X-Philosophy makes **further inferences** from these experimental results to philosophical conclusions
  - Negative: diversity and sensitivity of intuitions casts doubt on reliability
  - But also *positive* conclusions: facts about intuitions as relevant for philosophy
- Justifying these inferences requires **metaphilosophical** arguments about the relation between intuitions and philosophical "truths":
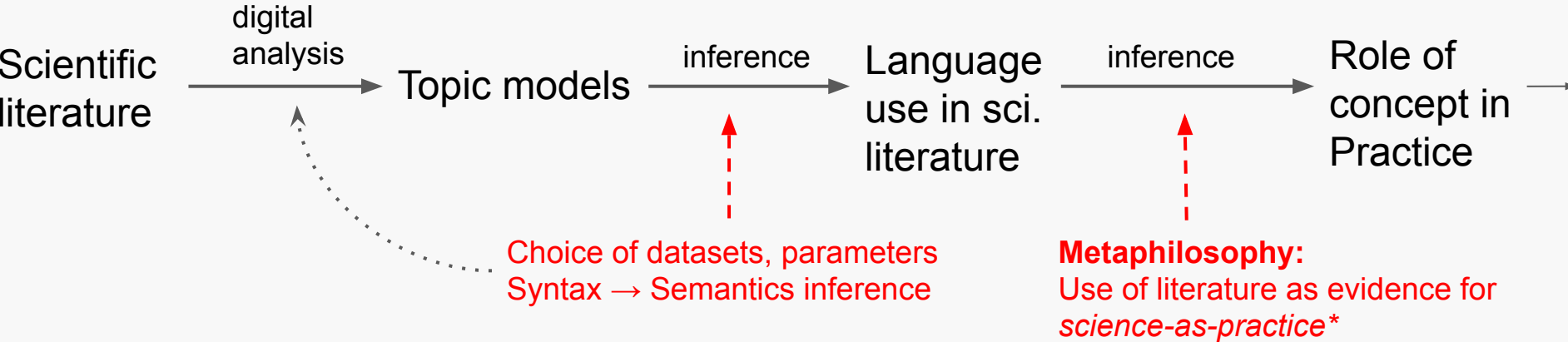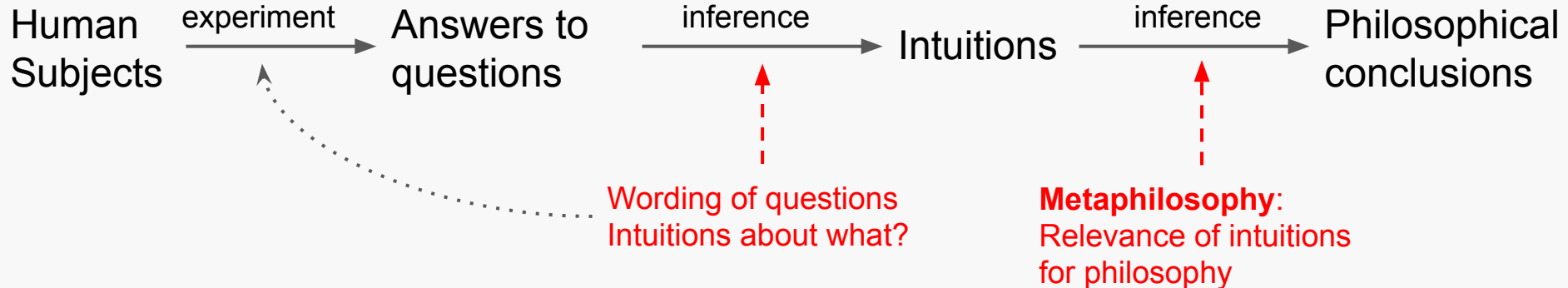
Case Study: Biodiversity
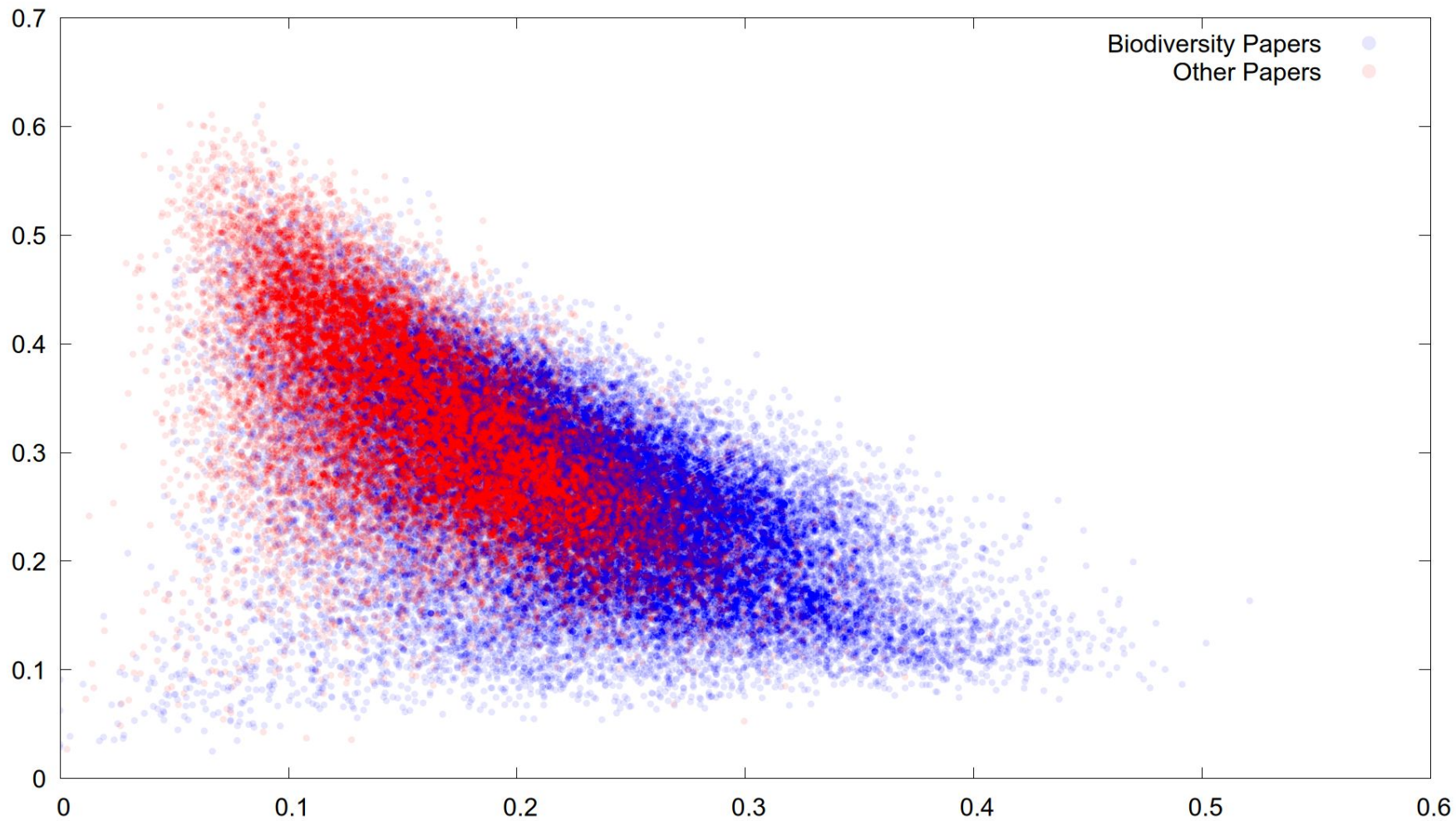
Digital Conceptual Analysis

How could we approach the use of a concept across an important corpus?

Test case: Looking at the use of *biodiversity* across the journal *Conservation Biology*.

Digital Conceptual Analysis

First major question: Are there any systematic differences between papers that use "biodiversity" and those that don't?

Answer: A few.

Biodiversity

Words which distinguish biodiversity articles:

- species
- areas
- forest
- land
- diversity
- richness
- protected
- *development*
- *policy*
- *international*
- Years from 1997–2006

Words which significantly occur within 500 words of uses of 'biodiversity':

- biointegrity
- macroclimatic
- distributive
- imperils
- *bureaucratically*
- countdown
- *neoliberalization*
- *postmodernism*
- *manifesto*
- hotspots
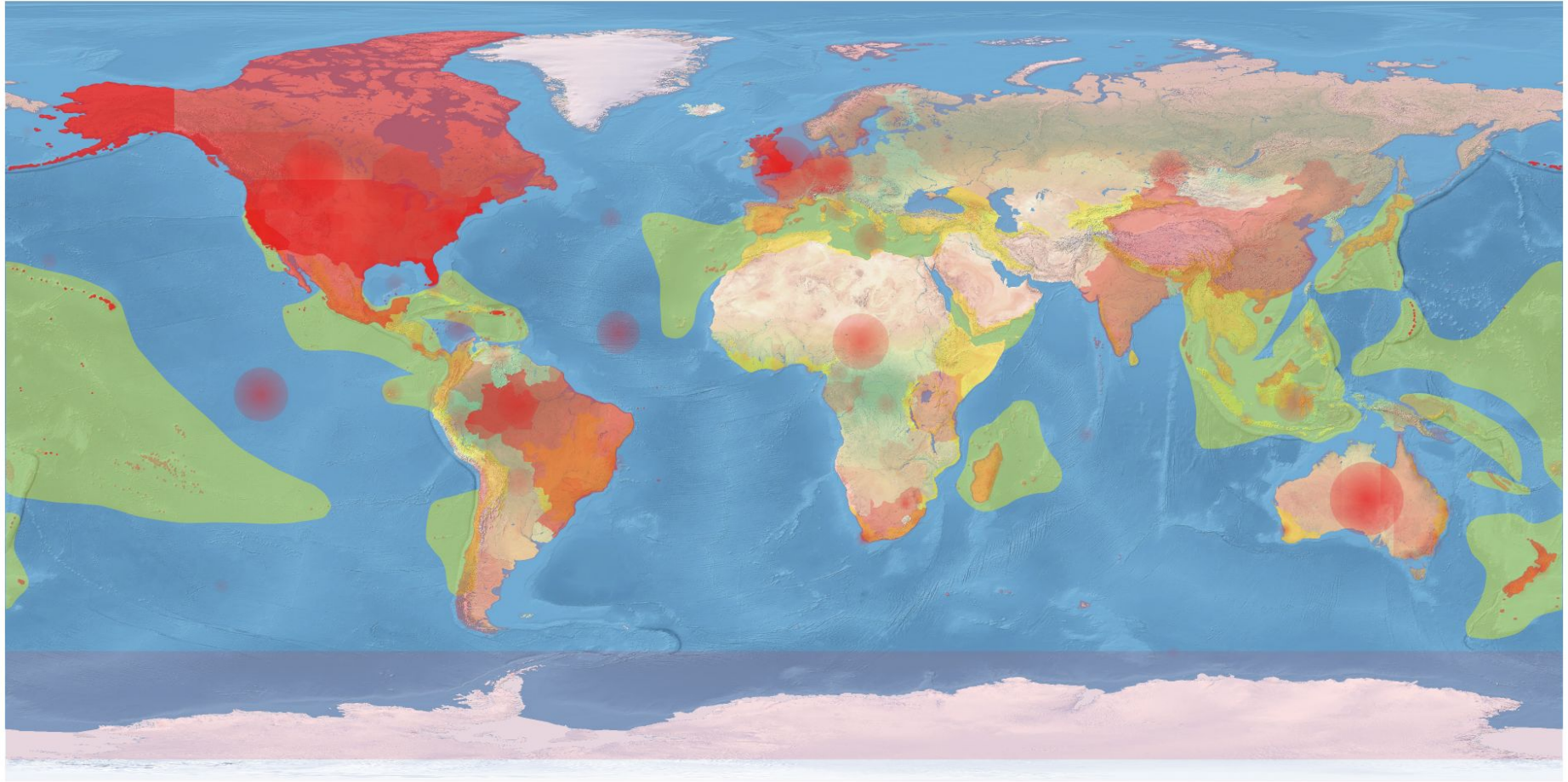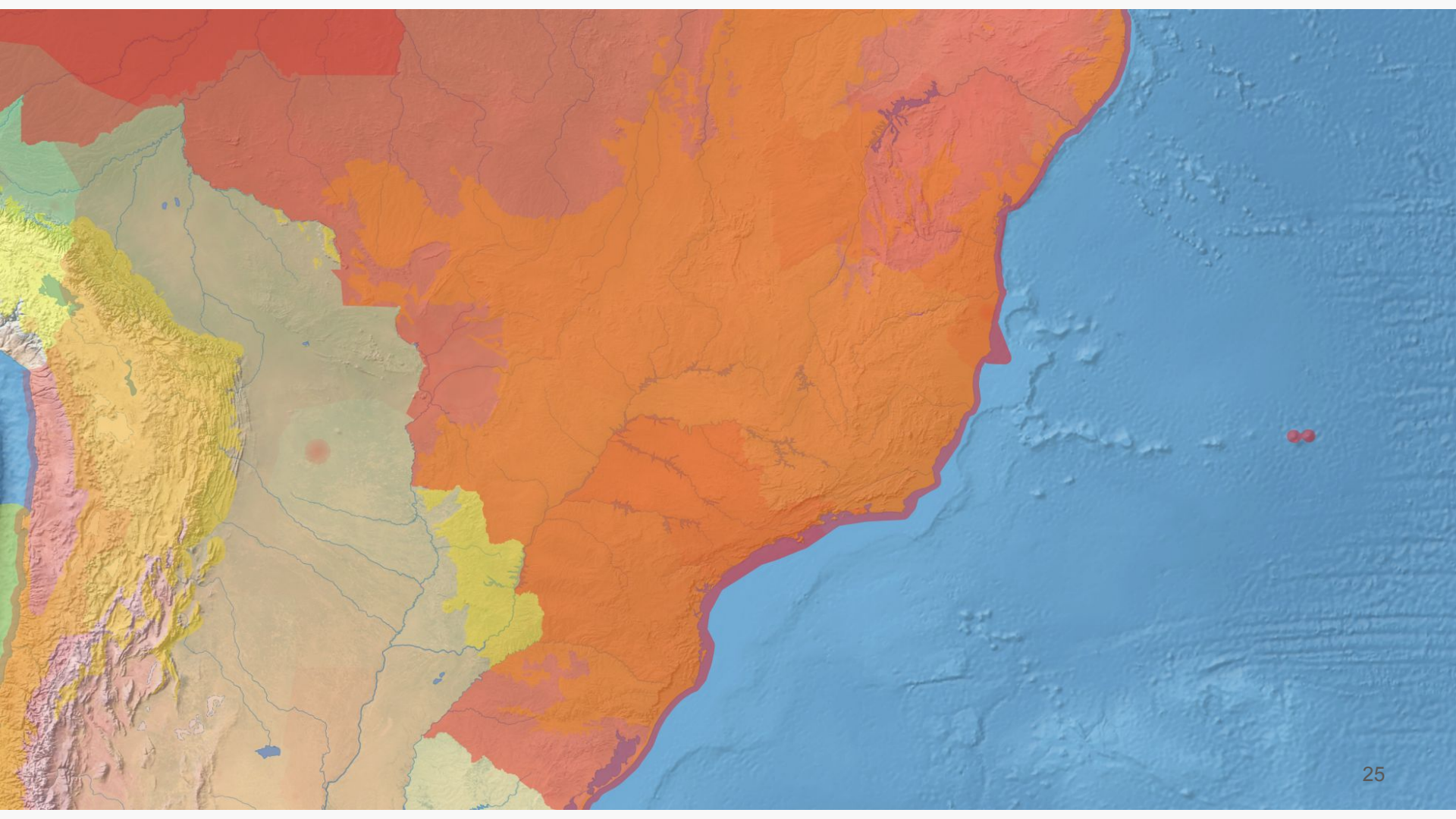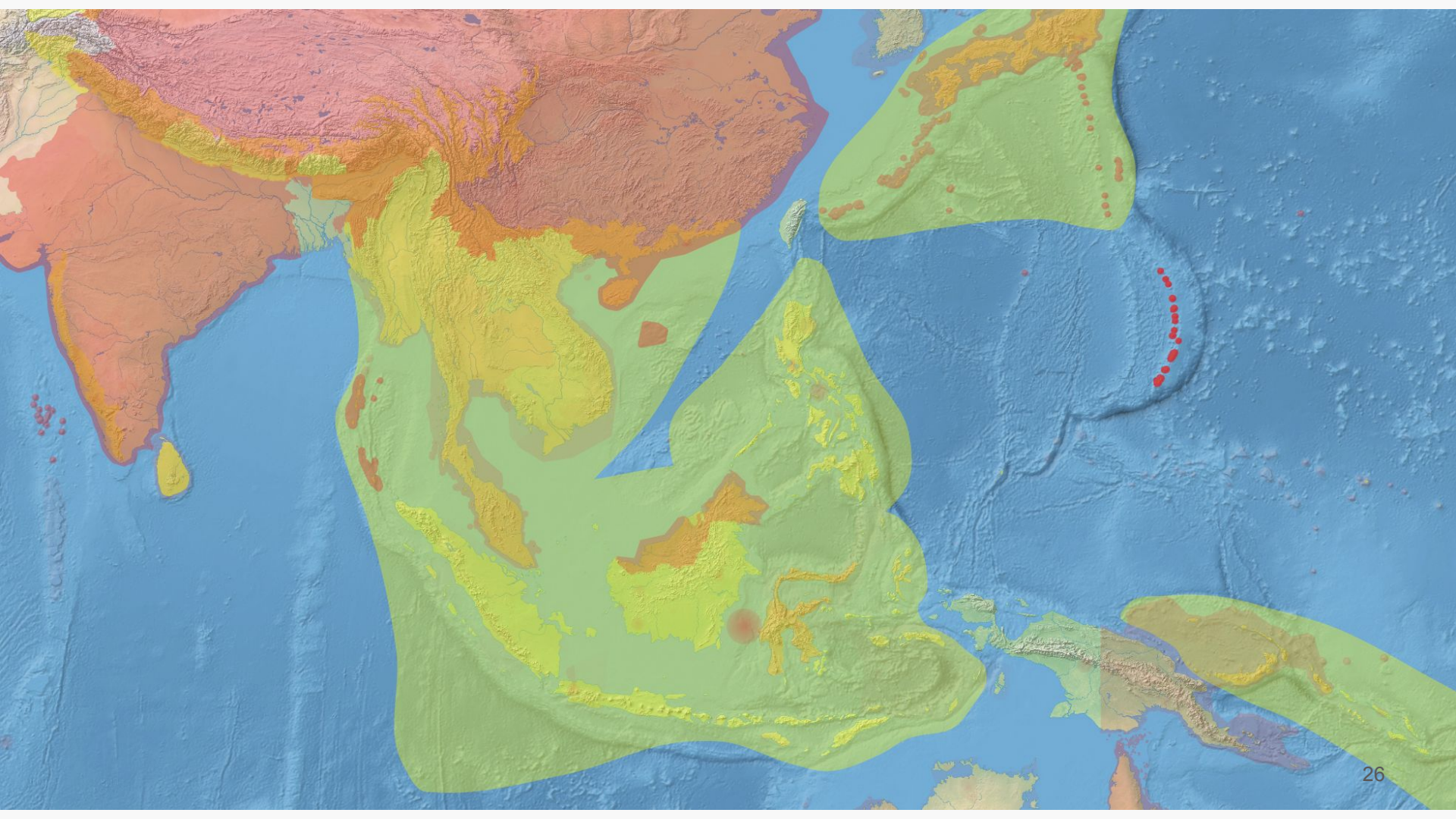
Digital Conceptual Analysis: Biodiversity

What do the data show? A preliminary hypothesis:

- Even among practicing scientists in a professional journal, biodiversity is a deeply politicized concept.
- Analyses that ignore social and political context are likely to be *scientifically* inadequate.
- And there's evidence of different concepts deployed in different scientific/social/political domains.

Supporting the Hypothesis

To return to the way of framing the question above: what are the critical inferential leaps here?

- Inferring from differences in articles to differences in the concept
- Inferring a univocal use of "biodiversity"
- Inferring that the articles accurately reflect scientists' ideal concepts
- Inferring that the algorithms actually capture intended reference (e.g., mapping inaccuracy)

(Future!) Case Study: Binding Specificity in Molecular Biology

3.2 Case Study: Binding Specificity

- Specificity (of binding interactions) is centrally important in molecular biology, but almost completely ignored by philosophers
- Why? Possibly because it's not a "hot topic" in most molecular research, but essential *background* — part of the conceptual furniture

Can we use digital analysis to demonstrate this?

3.2 Case Study: Binding Specificity

Proposal: Use topic modeling of literature to detect **tacit understanding** — concepts that are 1) not discussed, but 2) necessary for understanding what *is* being discussed

How can we detect implicit ideas using purely syntactic analysis?

- Step 1: Detect "specificity" topic(s), i.e. cluster(s) of words in which "specificity" is prominent
- Step 2: Identify documents that feature these topics but *without* "specificity"

3.2 Case Study: Binding Specificity

Following above, this requires justification at different steps of the process:

- Inferring from collocation (syntax) to semantic link
- Inferring from semantic link to *tacit understanding*: Semantic relationship between A and B doesn't necessarily imply A *necessary for understanding* B
  - Can we distinguish this digitally?
- Complementary evidence to experimental questioning of scientist's understanding, or just inferior?

Coda: a possible skeptical objection

Coda: a possible skeptical objection

Objection:

- Publications are some sort of embellished advertising devices that belie the actual scientific practice that's behind them and precedes them.
    - In other words, they bear no useful traces of the practice and artifacts actually produced by scientists while carrying out their research.
- Thus, analysis of the final produced papers, **even** on a massive scale, will not detect any interesting signals reflecting scientific practice.
- So, such an analysis will not produce any interesting result for an empirically-informed philosophy of scientific practice.

Coda: a possible skeptical objection

A possible counter-argument:

- not only philosophers of science, but **scientists themselves**—barring vis-à-vis communication via informal talks between *friend* scientists—in most cases *do* come to form an idea of the scientific landscape in which they are working by getting information from published (or pre-published) papers, or from talks at conferences, which may be just as much, if not more, a distortion of scientific practice as journal articles.

Coda: a possible skeptical objection

- That would mean that not even scientists have a good idea of what in general other scientists practically do.
- So, such an objection would hit not only our own work as empirically-informed philosophers of science, but **the whole scientific enterprise**.
- Do we actually want to bite that bullet? Probably *not*!

  This counter-argument can be thus seen as a sort of (non-logical) reductio ad absurdum of the objection.

Coda: a possible skeptical objection

- Granted, somebody could reply that the fact that papers even sometimes (or perhaps most of the time, in certain disciplines?) plainly *lie* is actually what gives rise to the replicability crisis. Isn't this a well-known problem?
- But, if we want to *generalize* that concern from, say, psychology, to other sciences, then we'll have to dismiss most or all science!
- That would border on *scientific denialism*, plain and simple.
  - In fact, the typical accusations denialists move against science is that most scientists *lie* (if for self-promotion or following a hidden agenda does not matter in the present context).
- Do we want that? Probably *not*!

Coda: a possible skeptical objection

That said, we think that *even* if papers do not completely reflect the actual science behind them, the whole literature can actually carry a *latent signal* of the scientific practice behind the production of journal articles or *of their very being deceptive.*

It is then precisely *our task,* as empirically-informed philosophers of science, that of devising ways to detect that signal.

Coda: a possible skeptical objection

This suggests a further counter-argument:

- Let's suppose that big-data analysis of the scientific literature, after much effort, fails to detect any signal of the influence of scientific practices, or of changes in these practices, in the actual final produced papers.
- This would be *in itself* a very interesting *result* for philosophy of science and of scientific practice, giving rise to a host of new and *interesting* problems!
- So, it seems that in any case, a large scale mining of the scientific literature should produce interesting results for philosophy of science, and could be well worth the effort.

Questions?

@pencelab · @pencechp · @LeanPhilSci
https://pencelab.be
charles@charlespence.net
omlean@gmail.com
luca.rivelli@gmail.com

[THIS SPACE IS INTENTIONALLY LEFT BLANK]

1930s: CW Morris. Semiotics as Metascience

Morris' pragmatical point of view on science is something akin to an *empirical* point of view on scientific *practice*:

" The study of the relation of signs to scientists may be called, in the spirit of pragmatism, the pragmatical investigation of the language of science. Here belong the problems as to how the scientist operates, the connection of science as a social institution with other social institutions, and the relation of scientific activity to other activities. (1938b)

2000-today: metascience reappears

2014: Johnathan Schooler, in "Metascience could rescue the 'replication crisis' ", *Nature*:

● "Metascience, the science of science, uses rigorous methods to examine how scientific practices influence the validity of scientific conclusions. It has its roots in the philosophy of science and the study of scientific methods, but is distinguished from the former by a reliance on quantitative analysis, and from the latter by a broad focus on the general factors that contribute to the limitations and successes of research."

[THIS SPACE IS INTENTIONALLY LEFT BLANK]

1930s: CW Morris. Semiotics as Metascience

1938: Charles W. Morris, *International Encyclopedia of Unified Science*:

- "Science, as a body of signs with certain specific relations to one another, to objects, and to practice, is at once a language, a knowledge of objects, and a type of activity. The interrelated study of syntactics, semantics, and pragmatics of the language of science in turn constitutes **metascience**—the science of science"

- "In this way, and on a comprehensive scale, science is made an object of scientific investigation. Metascience appears both as a tool for, and as an element within unified science."

1930s: CW Morris. Semiotics as Metascience

Morris, "The Significance of the Unity of Science Movement" (1946)

- *Metascience*: "a science of science is concerned with the methods of the science, their history and their sociology, as well as their language." (1946)
- "a science of science imposes no limitations on science; it will follow its subject-matter as does any other science. And it itself is open to the same false leads and growth and revisions as are the other sciences." (*ibid*)

1930s: CW Morris. Semiotics as Metascience

- Morris' idea of metascience stems from the application to science of his *semiotics*, comprising the study of syntax, of semantics and of pragmatics.

- For Morris, *metascience* is realized by studying through a metalanguage the language of science, by virtue of the fact that science makes use of and expresses its results in terms of signs.

- Being based on semiotics, Morris' metascience comprises also a semantic and, more interestingly, *pragmatic* point of view on the language of science.

1930s: CW Morris. Semiotics as Metascience

Morris' pragmatical point of view on science is something akin to an *empirical* point of view on scientific *practice*:

> " The study of the relation of signs to scientists may be called, in the spirit of pragmatism, the pragmatical investigation of the language of science. Here belong the problems as to how the scientist operates, the connection of science as a social institution with other social institutions, and the relation of scientific activity to other activities. (1938b)

[THIS SPACE IS INTENTIONALLY LEFT BLANK]

2000-today: metascience reappears

2002: Faust & Meehl, in *Using Meta-Scientific Studies to Clarify or Resolve Questions in the Philosophy and History of Science*: "As data are the subject matter for theories, theories and other scientific products are the subject matter for meta-theory and meta-science, organized and directed by methods that, in large part, remain to be developed. However, we believe that the era of meta-science is not far off and that it will make significant, if not revolutionary, contributions to the history and philosophy of science, and to the work of the practicing scientist."

2000-today: metascience reappears

2017: Zeng et al, "The Science of Science: From the Perspective of Complex Systems." in *Physics Reports*

- *SOS* is a rapidly developing field aiming to understand, quantify and predict scientific research and its outcomes
- Topics: measuring the influence of scientific publications, researchers, journals, and universities; modeling scientific collaboration and citation patterns; understanding innovation processes; classifying different scientific domains; predicting the future evolution of science.

[THIS SPACE IS INTENTIONALLY LEFT BLANK]

2000-today: metascience reappears

Metascience can recur to *massive automated literature-mining* techniques to draw conclusions about the state of scientific research.

This is significant for us, because upon results like these we could build our philosophical reflections.

[THIS SPACE IS INTENTIONALLY LEFT BLANK]