Rethinking Language

May 5, 2022

John Terilla

"Language is the core property that defines human beings"

Noam Chomsky

“Modern human beings process information symbolically, rearranging mental symbols to envision multiple potential realities. They also express the ideas they form using structured articulate language. No other living creature does either of these things.”

Ian Tattersall

“Modern human beings process information symbolically, rearranging mental symbols to envision multiple potential realities. They also express the ideas they form using structured articulate language. No other living creature does either of these things.”

Ian Tattersall

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

1 Introduction

Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Language Models are Few-Shot Learners

Tom B. Brown[*] **Benjamin Mann**[*] **Nick Ryder**[*] **Melanie Subbiah**[*]

Jared Kaplan[†] **Prafulla Dhariwal** **Arvind Neelakantan** **Pranav Shyam** **Girish Sastry**

Amanda Askell **Sandhini Agarwal** **Ariel Herbert-Voss** **Gretchen Krueger** **Tom Henighan**

Rewon Child **Aditya Ramesh** **Daniel M. Ziegler** **Jeffrey Wu** **Clemens Winter**

Christopher Hesse **Mark Chen** **Eric Sigler** **Mateusz Litwin** **Scott Gray**

Benjamin Chess **Jack Clark** **Christopher Berner**

Sam McCandlish **Alec Radford** **Ilya Sutskever** **Dario Amodei**

OpenAI

Abstract

Recent work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. While typically task-agnostic in architecture, this method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples. By contrast, humans can generally perform a new language task from only a few examples or from simple instructions – something which current NLP systems still largely struggle to do. Here we show that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine-tuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and test its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic. At the same time, we also identify some datasets where GPT-3's few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora. Finally, we find that GPT-3 can generate samples of news articles which human evaluators have difficulty distinguishing from articles written by humans. We discuss broader societal impacts of this finding and of GPT-3 in general.

[*]Equal contribution
[†]Johns Hopkins University, OpenAI

Author contributions listed at end of paper.

Training

The quick brown fox jumps over the lazy do…

The quick brown fox jumps over the lazy doV

The quick brown fox jumps over the lazy dog!

After training…

I went to the grocery store and bought …

I went to the grocery store and bought … a can of chickpeas.

I went to the grocery store and bought … some cucumbers.

I went to the grocery store and bought … a new yacht.

I went to the grocery store and bought … aKTyy8@vQ$$q

image from nations online project

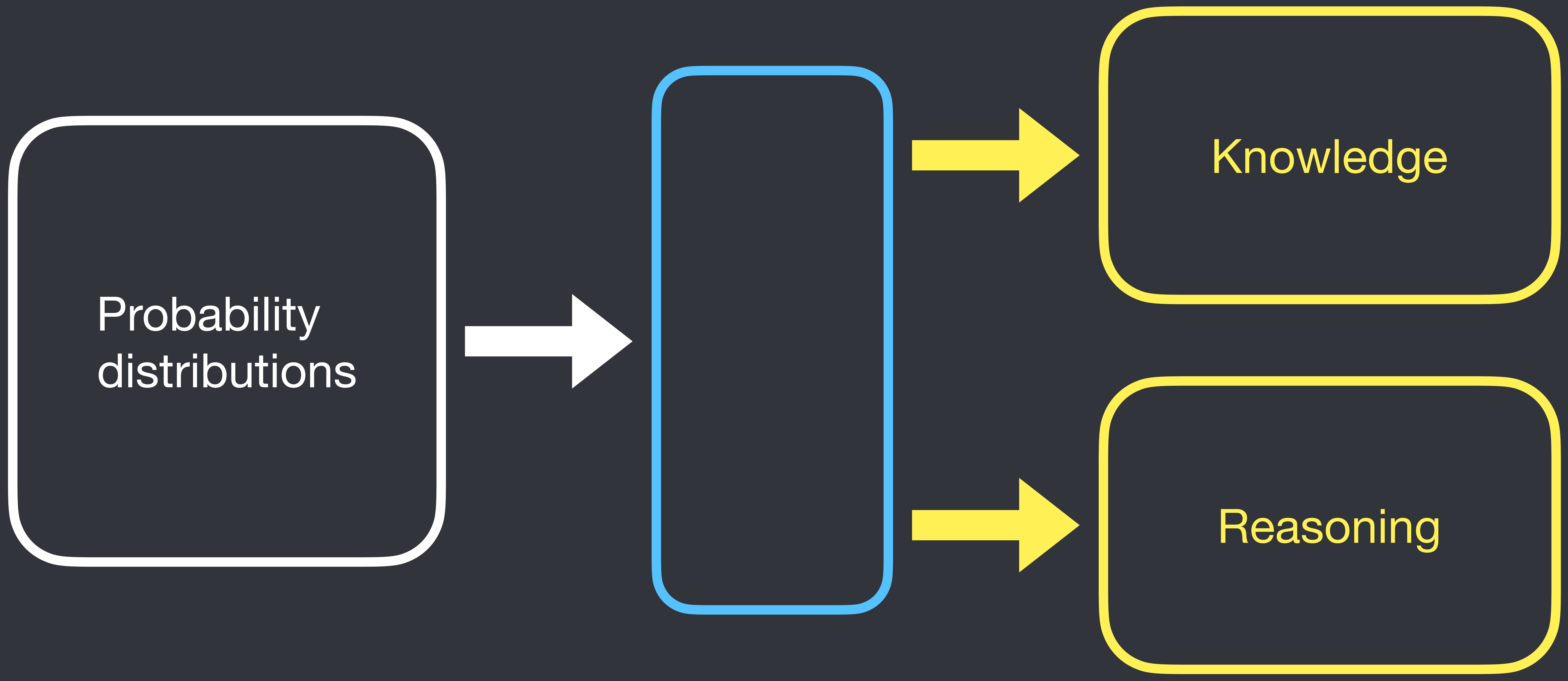
image from nations online project

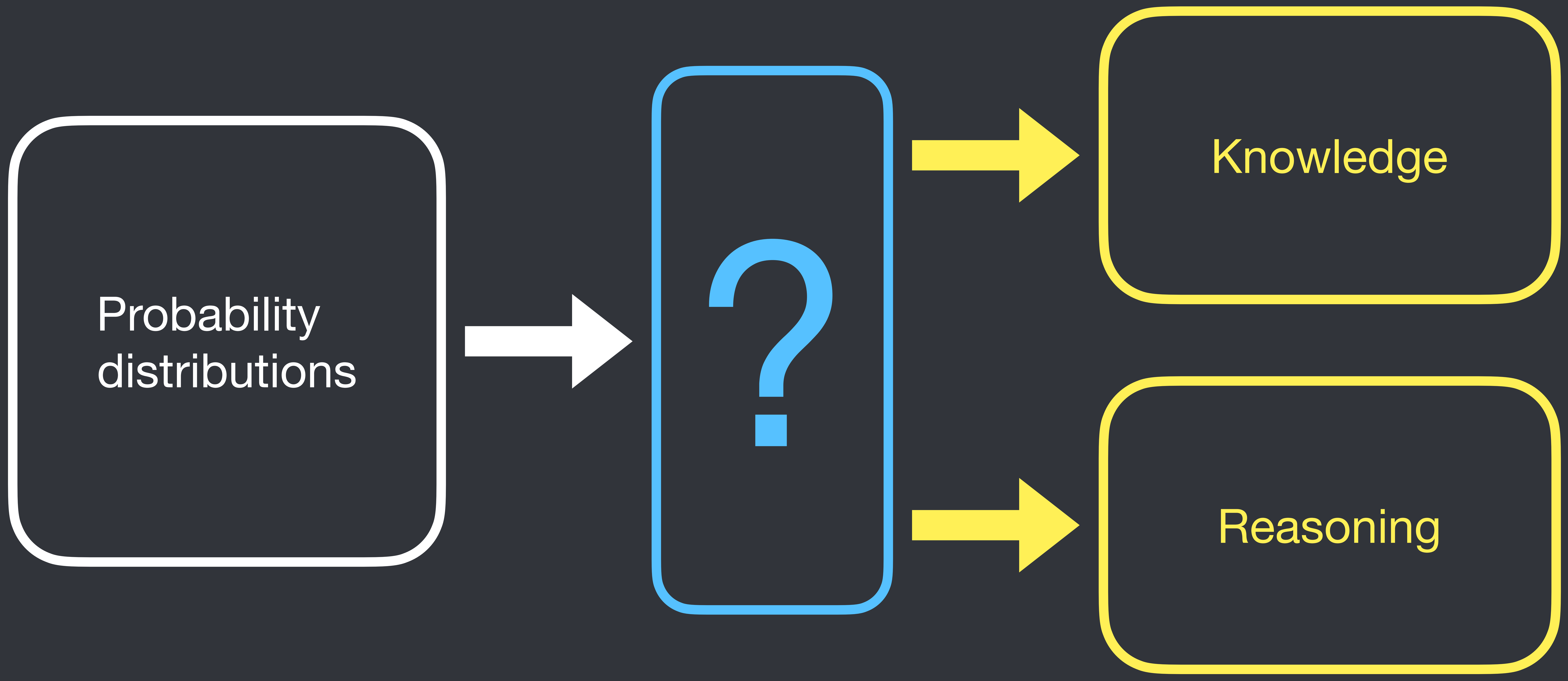
Tuesday

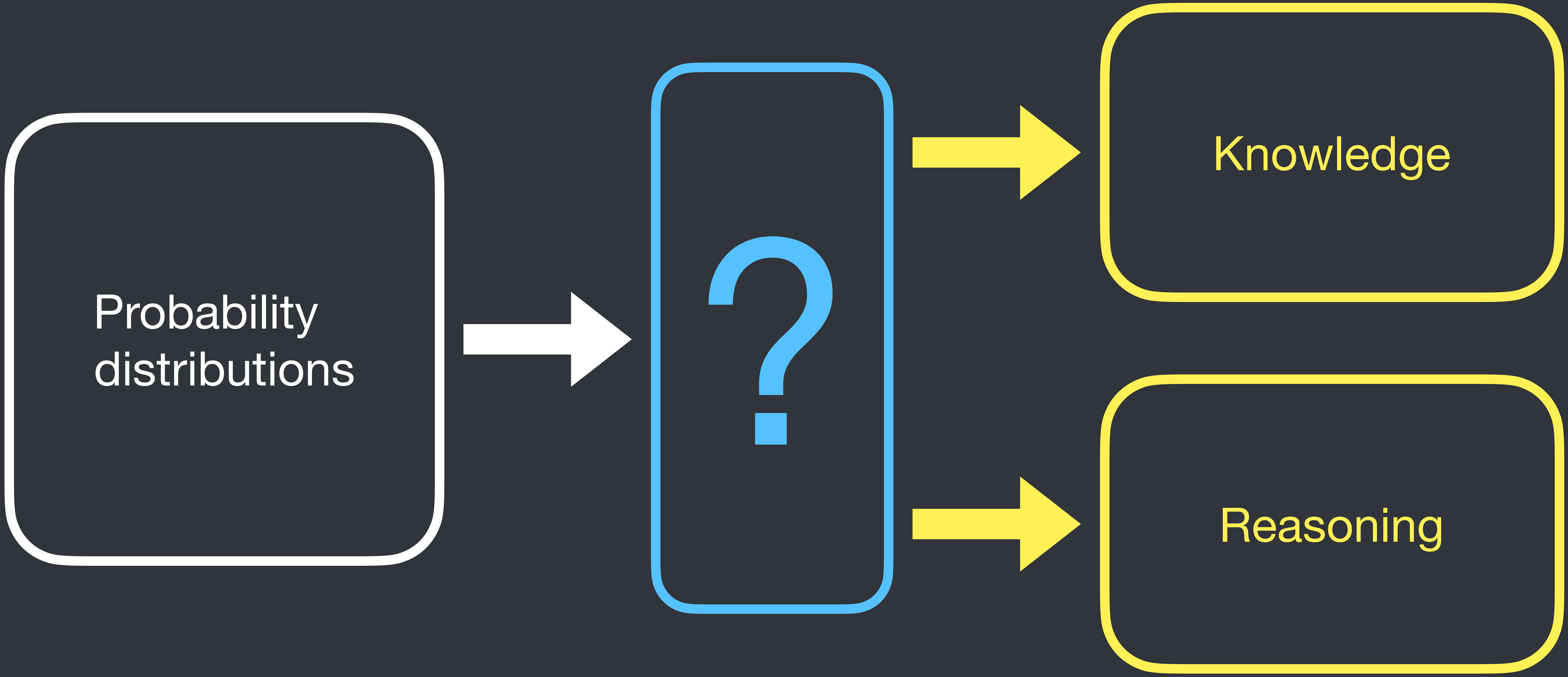
Monday —> Tuesday

Main Question

Main Question: *How are knowledge and rules for reasoning about that knowledge encoded in probability distributions of next character continuations?*


Probability distributions
Knowledge
Reasoning


Probability distributions
?
Knowledge
Reasoning

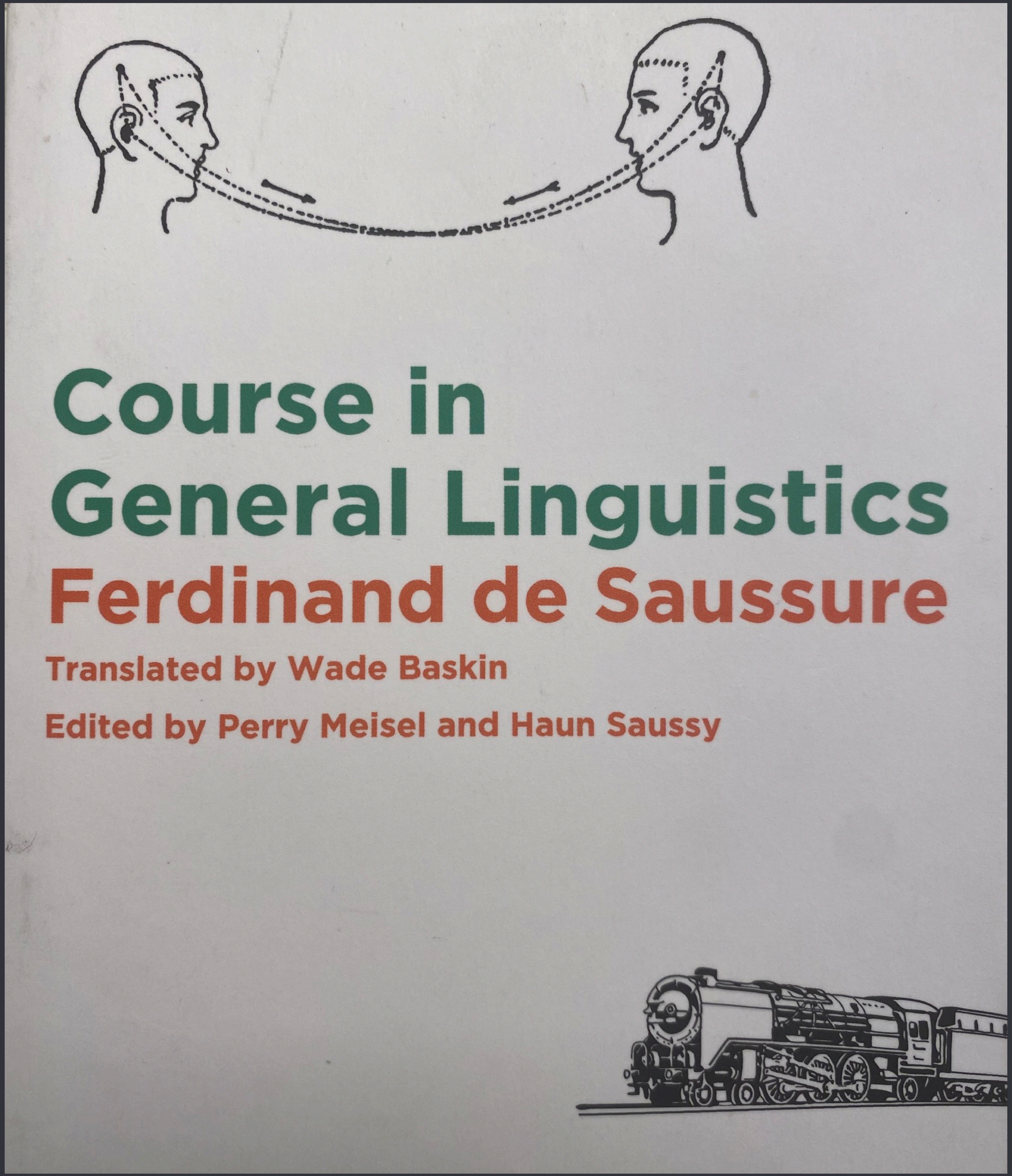

Probability distributions
?
Knowledge
Reasoning

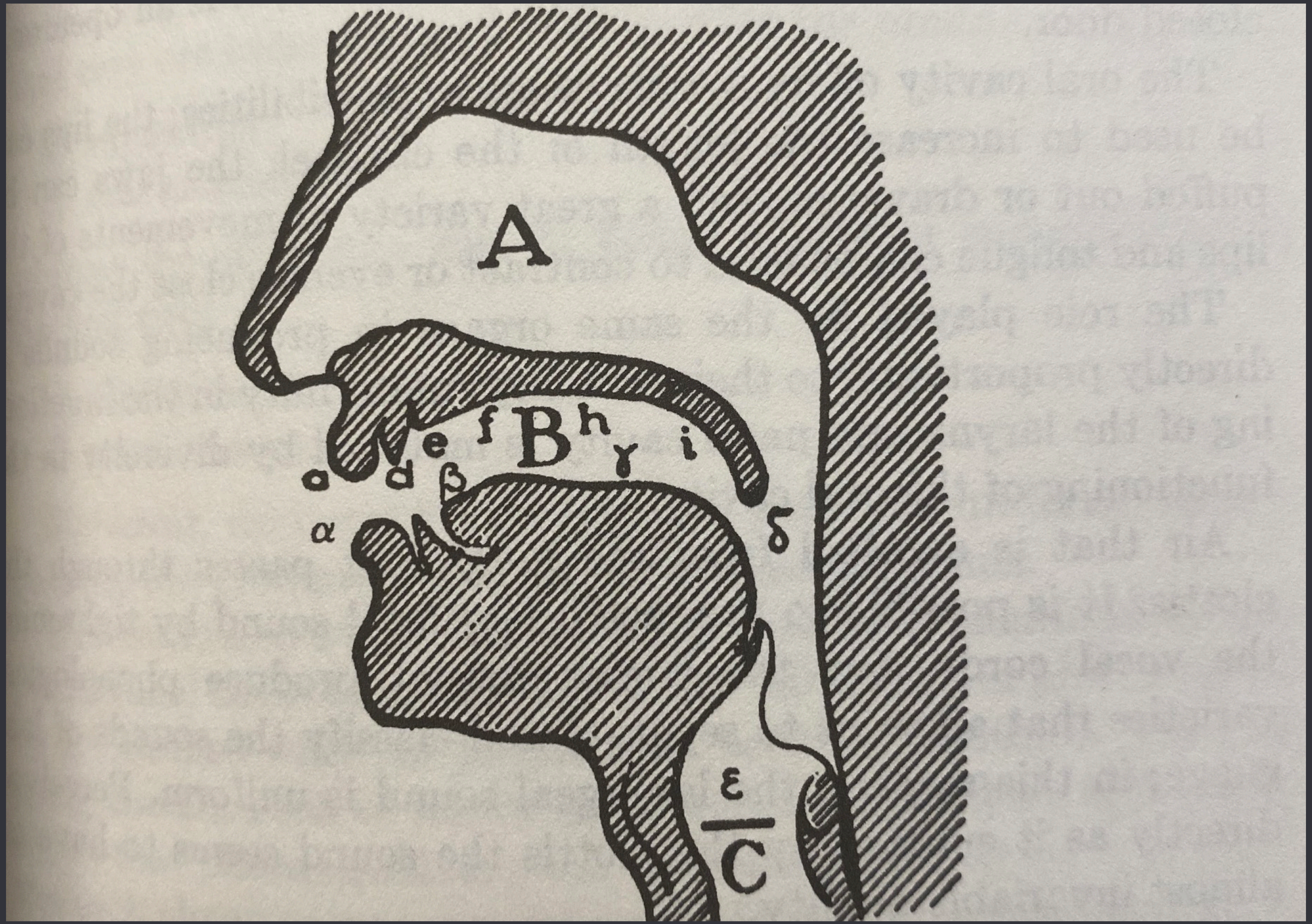
One more thing…

Surprising Fact: ***The ability to continue stories can be learned by simple trial and error!***

You can’t learn how to give birth to kittens by a trial and error.


43167
Winesap
Silas Wilson
Wenatchee Co.
Wash
D.G. Passmore
12.22.08
4.18.09


Course in
General Linguistics
Ferdinand de Saussure
Translated by Wade Baskin
Edited by Perry Meisel and Haun Saussy


A
B
h
f
e
d
o
α
β
γ
i
δ
ε
C

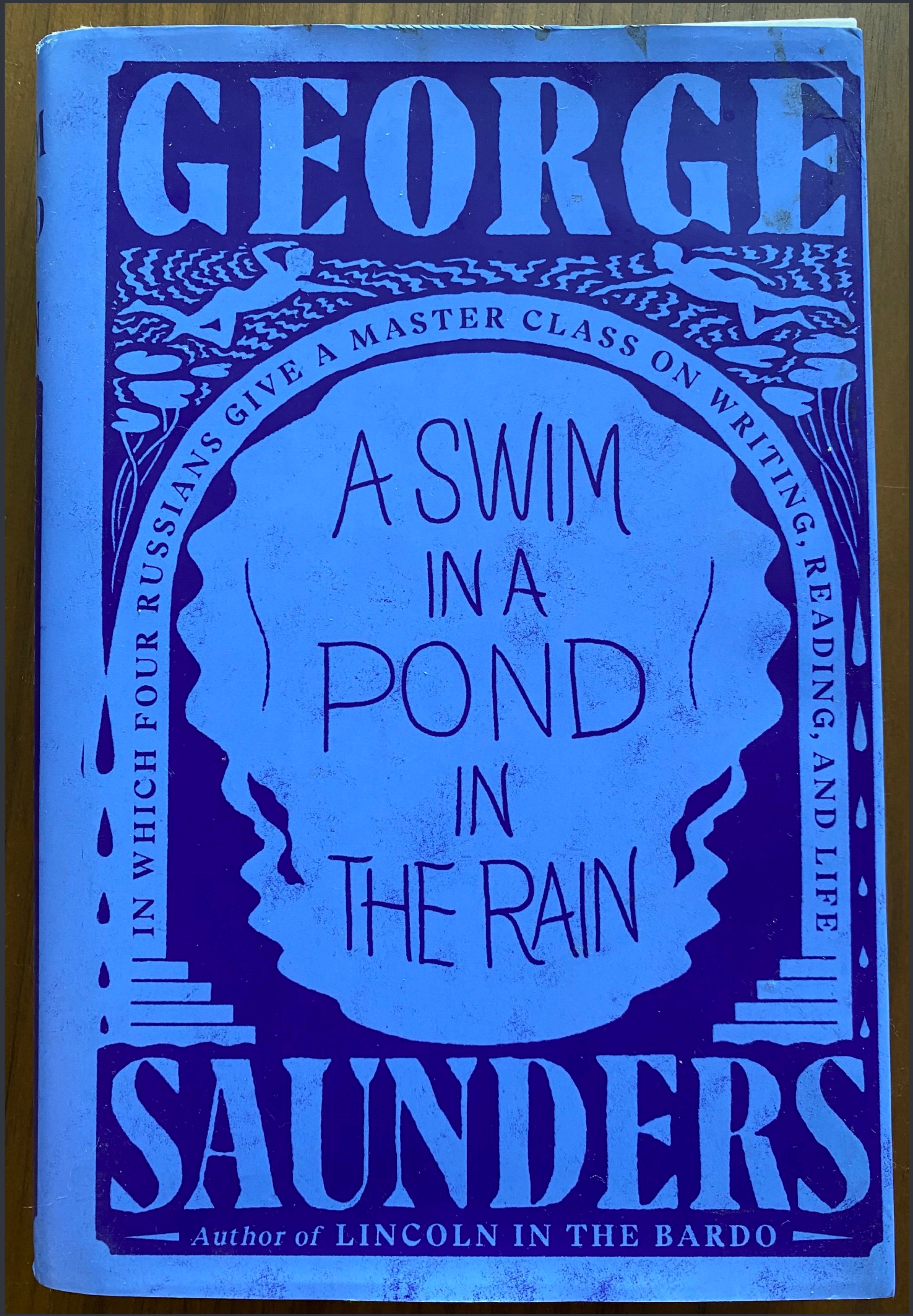
2015

2022

<u>**Main Question**</u>**:** ***How are knowledge and rules for reasoning about that knowledge encoded in probability distributions of next character continuations?***

Dynamic Semantics:

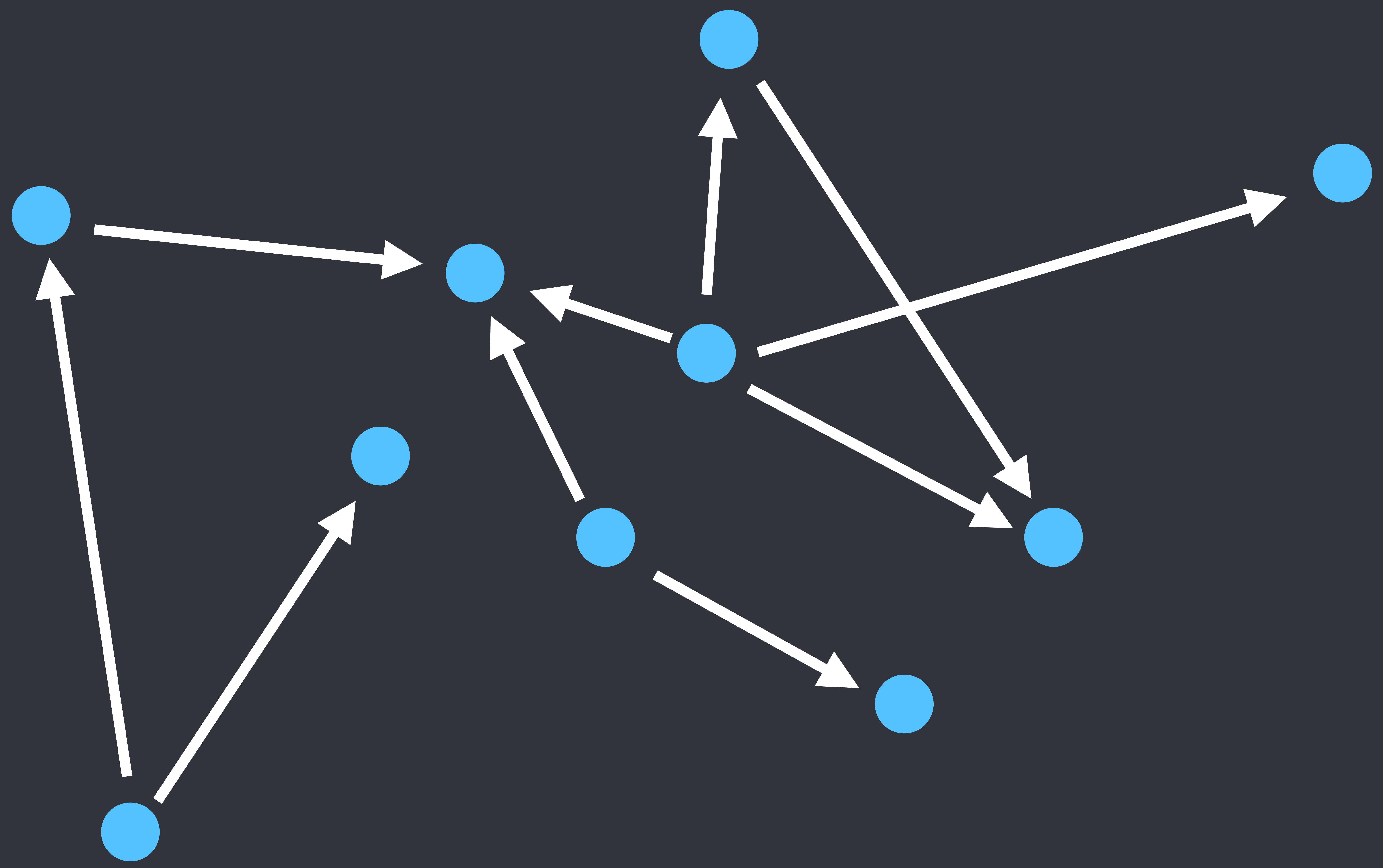
Meaning is context change potential.


GEORGE
IN WHICH FOUR RUSSIANS GIVE A MASTER CLASS ON WRITING, READING, AND LIFE
A SWIM IN A POND IN THE RAIN
SAUNDERS
Author of LINCOLN IN THE BARDO

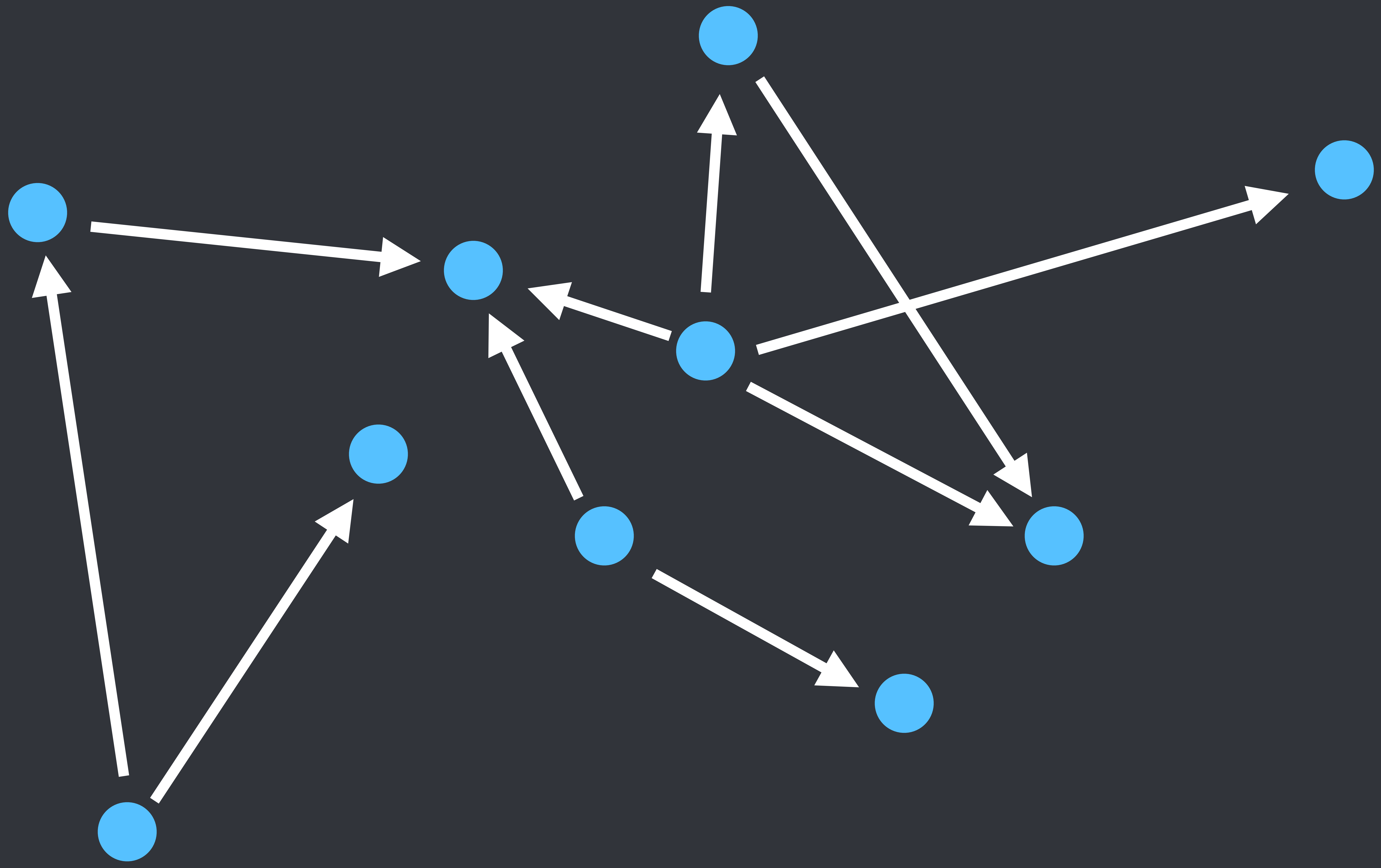
Category Theory

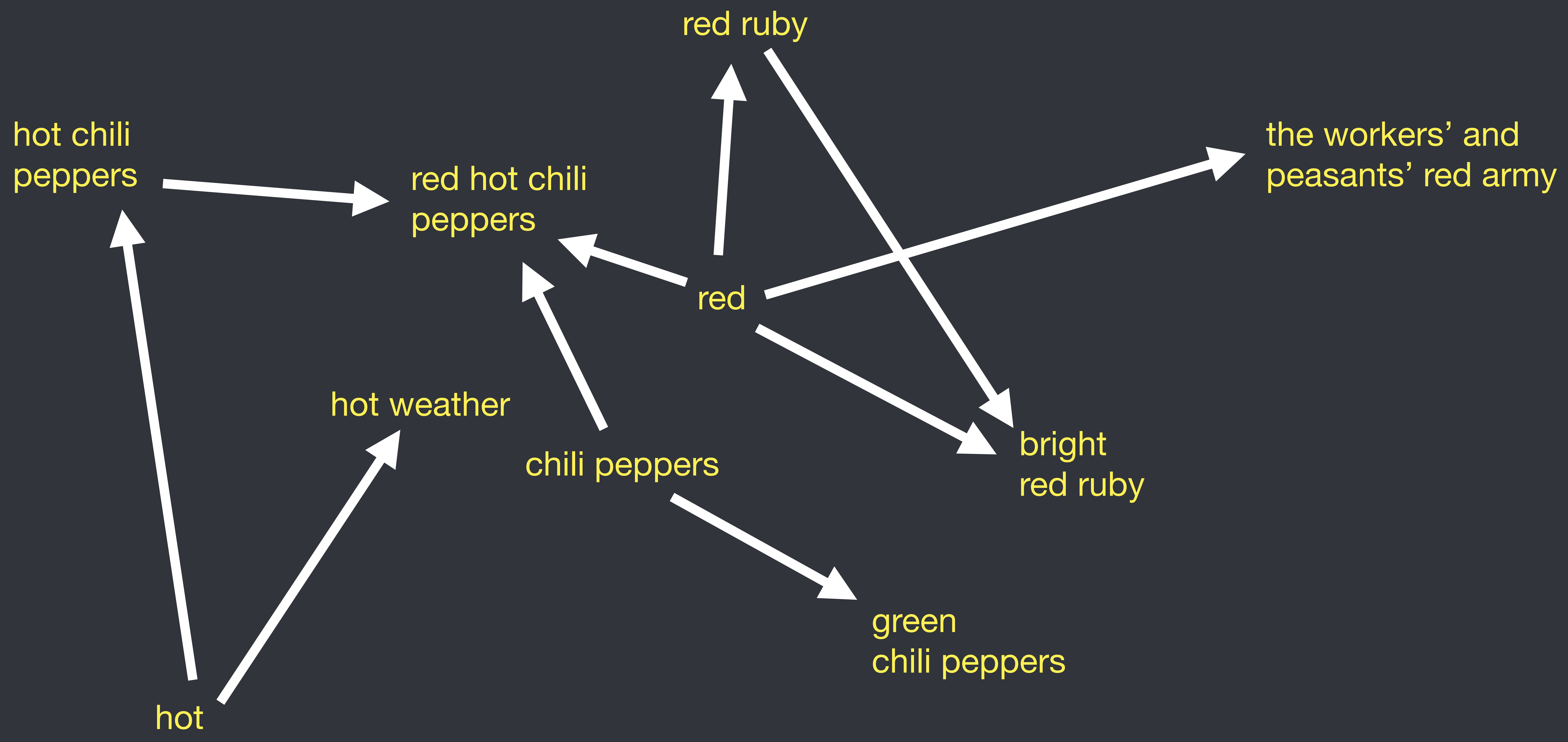
“The mathematical language developed by the end of the 20th century by far exceeds in its expressive power anything, even imaginable, say, before 1960.”

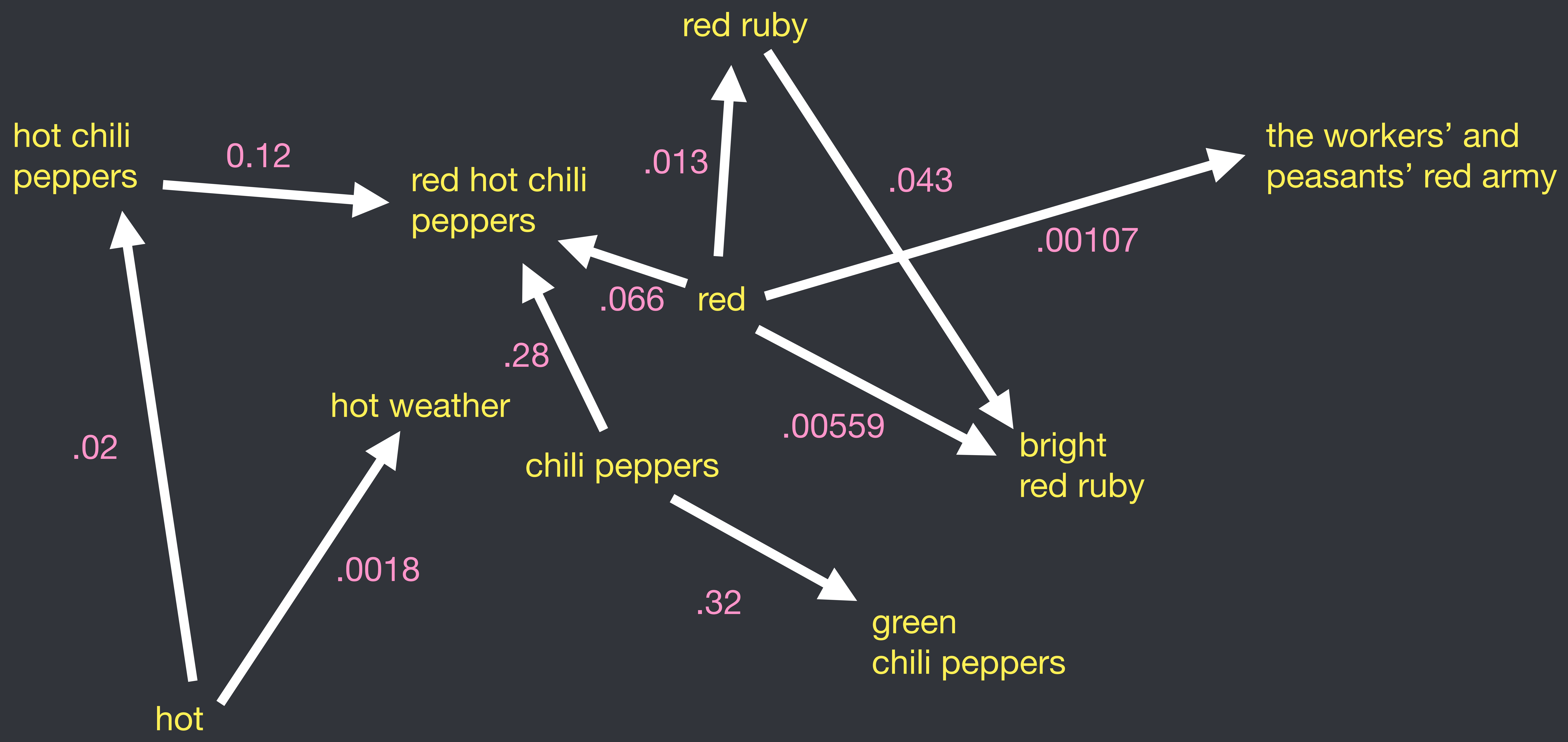
Misha Gromov

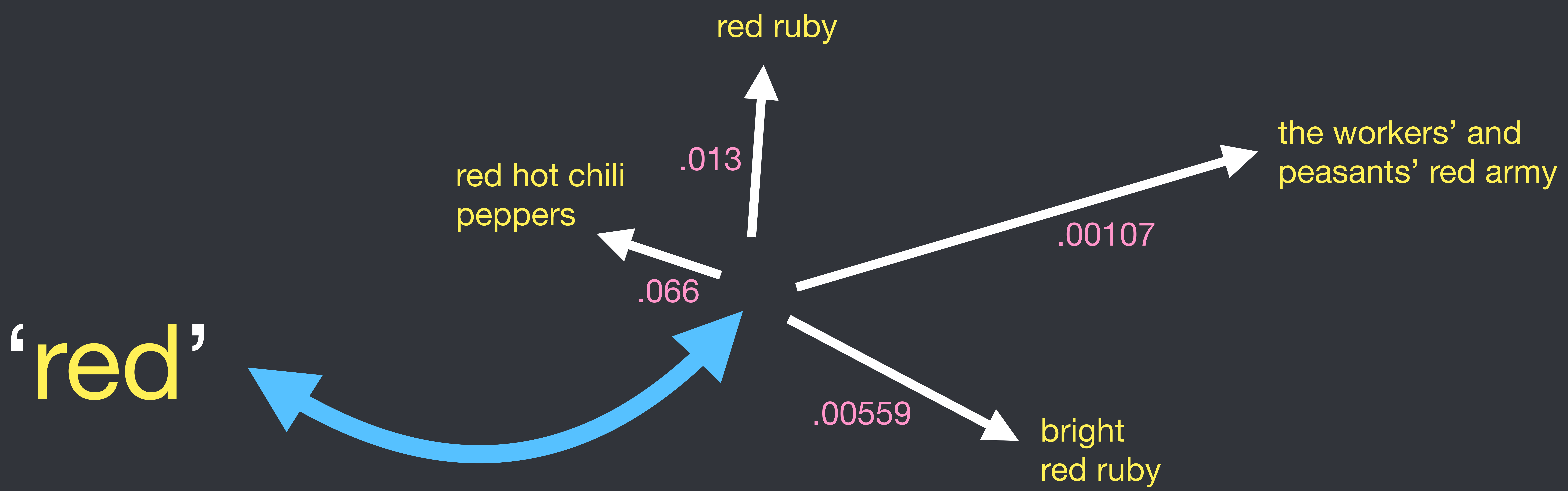


Tai-Danae Bradley
Yiannis Vlassopoulos




red ruby
hot chili peppers
red hot chili peppers
the workers' and peasants' red army
red
hot weather
chili peppers
bright red ruby
green chili peppers
hot


red ruby
hot chili peppers
0.12
red hot chili peppers
.013
.043
the workers' and peasants' red army
.00107
.066
red
.28
hot weather
.00559
bright red ruby
.02
chili peppers
.0018
.32
green chili peppers
hot



Yoneda Lemma

Semantic Category

Language Category

Yoneda Embedding

Semantic Category:

Functions (sheaves) on the language category.

Semantic Category: ***Build concepts and do some reasoning.***

La Matematica _####################_
https://doi.org/10.1007/s44007-022-00021-2

ORIGINAL RESEARCH ARTICLE

An Enriched Category Theory of Language: From Syntax to Semantics

Tai-Danae Bradley[1] · John Terilla[2] · Yiannis Vlassopoulos[3]

Received: 24 June 2021 / Revised: 7 February 2022 / Accepted: 7 February 2022

Abstract

State of the art language models return a natural language text continuation from any piece of input text. This ability to generate coherent text extensions implies significant sophistication, including a knowledge of grammar and semantics. In this paper, we propose a mathematical framework for passing from probability distributions on extensions of given texts, such as the ones learned by today's large language models, to an enriched category containing semantic information. Roughly speaking, we model probability distributions on texts as a category enriched over the unit interval. Objects of this category are expressions in language, and hom objects are conditional probabilities that one expression is an extension of another. This category is syntactical—it describes what goes with what. Then, via the Yoneda embedding, we pass to the enriched category of unit interval-valued copresheaves on this syntactical category. This category of enriched copresheaves is semantic—it is where we find meaning, logical operations such as entailment, and the building blocks for more elaborate semantic concepts.

Keywords Category theory · Yoneda embedding · Compositionality · Natural language · Probability · Logic

Mathematics Subject Classification 18D20 · 18A25 · 18A30 · 18A35 · 18B25 · 18B35

✉ Tai-Danae Bradley
tai.danae@math3ma.com

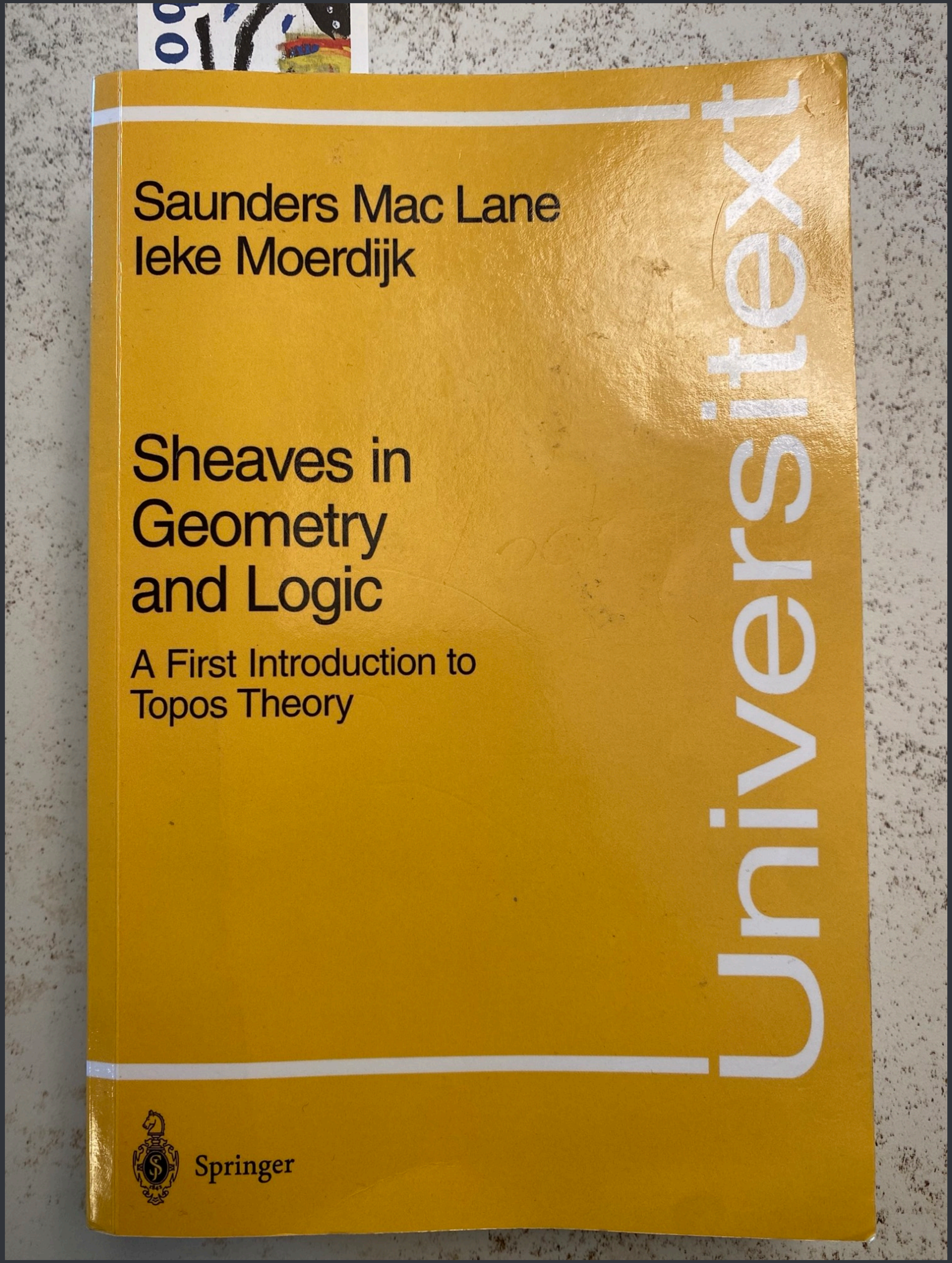
John Terilla
jterilla@gc.cuny.edu

Yiannis Vlassopoulos
yiannis@tunnel.tech

1 X, the Moonshot Factory and Sandbox@Alphabet, Mountain View, CA, USA

2 The City University of New York and Tunnel, New York, NY, USA

3 Tunnel, New York, NY, USA


Saunders Mac Lane
Ieke Moerdijk
Sheaves in Geometry and Logic
A First Introduction to Topos Theory
Universitext
Springer

The kind of reasoning one can do internally from text is something like:

categorical logic + probabilities.

Reprints in Theory and Applications of Categories, No. 1, 2002, pp. 1–37.

METRIC SPACES, GENERALIZED LOGIC, AND CLOSED CATEGORIES

F. WILLIAM LAWVERE

Author Commentary:
Enriched Categories in the Logic of Geometry and Analysis

Because parts of the following 1973 article have been suggestive to workers in several areas, the editors of TAC have kindly proposed to make it available in the present form. The idea on which it is based can be developed considerably further, as initiated in the 1986 article [1]. In the second part of this brief introduction I will summarize, for those familiar with the theory of enriched categories, some of the more promising of these further developments and possibilities, including suggestions coming from the modern theory of metric spaces which have not yet been elaborated categorically. (The 1973 and 1986 articles had also a didactic purpose, and so include a detailed introduction to the theory of enriched categories itself.)

While listening to a 1967 lecture of Richard Swan, which included a discussion of the relative codimension of pairs of subvarieties, I noticed the analogy between the triangle inequality and a categorical composition law. Later I saw that Hausdorff had mentioned the analogy between metric spaces and posets. The poset analogy is by itself perhaps not sufficient to suggest a whole system of constructions and theorems appropriate for metric spaces, but the categorical connection is! This connection is more fruitful than a mere analogy, because it provides a sequence of mathematical theorems, so that enriched category theory can suggest new directions of research in metric space theory and conversely, unusual for two subjects so old (1966 and 1906 respectively).

The closed interval $[0, \infty]$ of real numbers as objects, $\geq$ as maps, $+$ as "tensor" and truncated subtraction as adjoint "hom", constitute a bona fide example of a complete, symmetric, monoidal closed category V. For any such V there is the rich system of constructions and theorems (worked out by Eilenberg and Kelly, Day, and others) involving

- V-valued categories;
- V-strong functors;

Originally published as: Metric spaces, generalized logic, and closed categories, *Rendiconti del seminario matématico e fisico di Milano, XLIII (1973), 135-166*
Received by the editors 2002-04-01 and, in revised form, 2002-06-24.
Transmitted by Michael Barr. Reprint published on 2002-09-1.
2000 Mathematics Subject Classification: 18D20.
Key words and phrases: Metric spaces, enriched categories, logic.

Theory and Applications of Categories, Vol. 28, No. 22, 2013, pp. 696–732.

TIGHT SPANS, ISBELL COMPLETIONS AND SEMI-TROPICAL MODULES

SIMON WILLERTON

Abstract. In this paper we consider generalized metric spaces in the sense of Lawvere and the categorical Isbell completion construction. We show that this is an analogue of the tight span construction of classical metric spaces, and that the Isbell completion coincides with the directed tight span of Hirai and Koichi. The notions of categorical completion and cocompletion are related to the existence of semi-tropical module structure, and it is shown that the Isbell completion (hence the directed tight span) has two different semi-tropical module structures.

Introduction.

This paper grew out of a desire to understand whether the tight span of a metric space could be understood in terms of the enriched category theory approach to metric spaces. This led to understanding a link between two apparently unrelated constructions of Isbell, namely the tight span of metric spaces and the Isbell completion of categories; this is turn led, via categorical completeness, to connections with tropical algebra. It seems interesting that these two constructions of Isbell remained unconnected for nearly fifty years.

In this introduction the main ideas of Isbell completion, semi-tropical algebra and tight spans will be given. The intention is that this paper should be readable by mathematicians interested in metric spaces or tropical algebra, without much category theory background, and to allow them to see how category theoretic methods give interesting insight in this case. This means that some bits of enriched category theory for metric spaces will be spelt out in some detail.

The Isbell completion of a generalized metric space. Lawvere [18] observed that a metric space can be viewed as something similar to a category and that from that perspective there is a natural generalization — generalized metric space — which means a set X with a 'distance' function $\mathrm{d}\colon X \times X \to [0, \infty]$ such that $\mathrm{d}(x, x) = 0$ and $\mathrm{d}(x, y) + \mathrm{d}(y, z) \geq \mathrm{d}(x, z)$ for all $x, y, z \in X$, with no further conditions like symmetry imposed. Generalized metric spaces can be thought of as directed metric spaces. From a category theoretic point of view, generalized metric spaces are precisely $[0, \infty]$-enriched categories and so much of the machinery of category theory can be utilized to study them. In this paper we will look at the 'Isbell completion' for generalized metric spaces.

Received by the editors 2013-03-26 and, in revised form, 2013-08-15.
Transmitted by Anders Kock. Published on 2013-08-22.
2010 Mathematics Subject Classification: Primary: 54E35 Secondary: 18D20, 16Y60.
Key words and phrases: key words: Metric spaces, tropical algebra, injective hull.

The University Of Sheffield.

ON THE FUZZY CONCEPT COMPLEX

Jonathan Arthur Elliott

A thesis submitted for the degree of
Doctor of Philosophy

University of Sheffield
Faculty of Science
School of Mathematics and Statistics

September 2017

Categorical semantics of metric spaces and continuous logic

Simon Cho

Journal of Symbolic Logic 85 (3):1044-1078 (2020)

Abstract

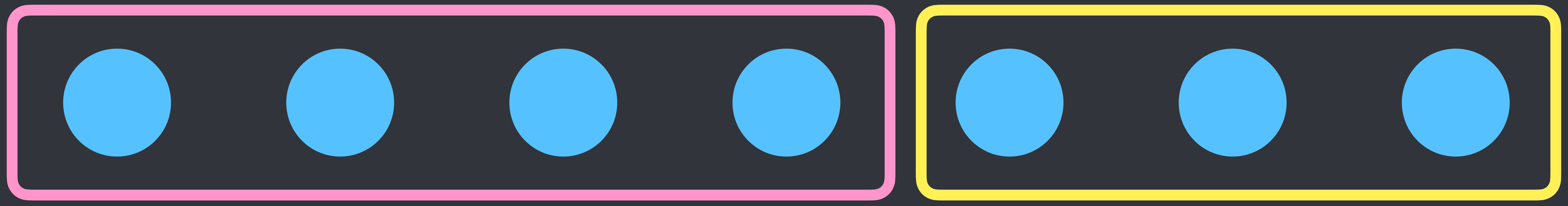
Using the category of metric spaces as a template, we develop a metric analogue of the categorical semantics of classical/intuitionistic logic, and show that the natural notion of predicate in this "continuous semantics" is equivalent to the a priori separate notion of predicate in continuous logic, a logic which is independently well-studied by model theorists and which finds various applications. We show this equivalence by exhibiting the real interval $[0,1]$ in the category of metric spaces as a "continuous subobject classifier" giving a correspondence not only between the two notions of predicate, but also between the natural notion of quantification in the continuous semantics and the existing notion of quantification in continuous logic.Along the way, we formulate what it means for a given category to behave like the category of metric spaces, and afterwards show that any such category supports the aforementioned continuous semantics. As an application, we show that categories of presheaves of metric spaces are examples of such, and in fact even possess continuous subobject classifiers.

Why should anyone care?

Quantum Physics

Language is like a 1d system of interacting quantum particles

2



A quantum physics model for language is compatible with a categorical perspective for language.

Practical Benefits!

Tensor networks

Tensor network language models have been built.

The New York Times Magazine

A.I. Is Mastering Language. Should We Trust What It Says?

OpenAI's GPT-3 and other neural nets can now write original prose with mind-boggling fluency — a development that could have profound implications for the future.

If transformers prove to have fundamental limitations, ideas from quantum physics might help get over them.