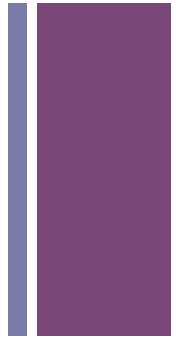


Edwardian Proofs as Futuristic Programs

Valeria de Paiva
Rearden Commerce,
Foster City, CA, USA



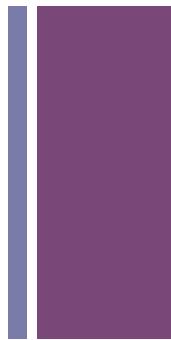
outline



- Personal history
- Unity behind disparity:
- Categorical logic
- Curry-Howard Correspondence
- Learning to enjoy change

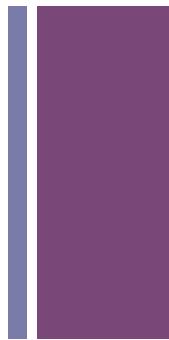
+

Personal and Inspiring?...



+

PUC—Rio de Janeiro
Dept Maths, BA and MA (Algebra)



+ University of Cambridge DPMMS 1984-88, Lab 1989-1995





University of Birmingham, UK

Computer Science 1996-2000



+ Xerox PARC NLTT: Natural Language Theory and Technology

October 2000-May 2008



cuil

Search 127 billion pages

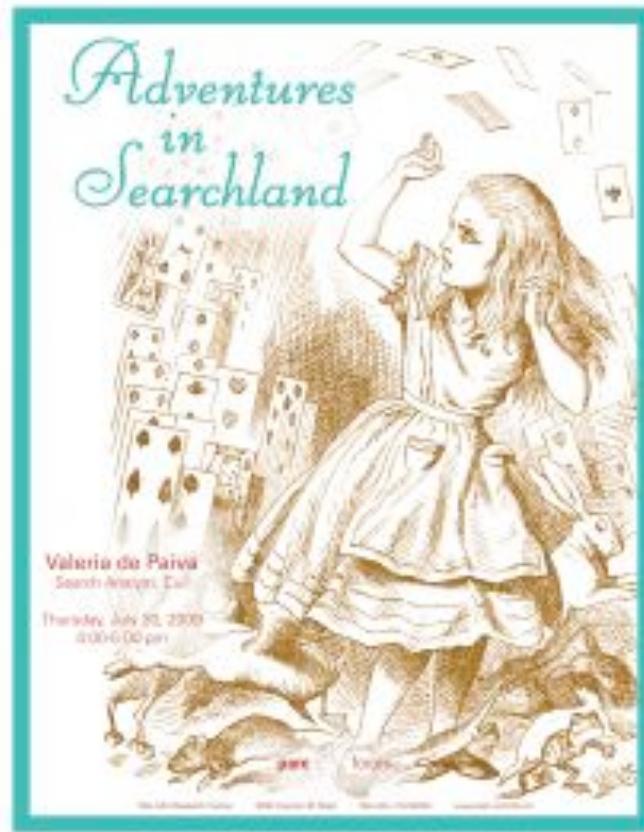
[About Cuil](#) | [Preferences](#) | [Add Cuil to Firefox](#)

[Privacy Policy](#) | © 2010 Cuil, Inc.



Adventures in Searchland

PARC Forum



- **Valeria de Paiva, Cuil,
Search Analyst**
- 30 July 2009
4:00-5:00pm
George E. Pake Auditorium,
PARC, Palo Alto, CA
[map/ directions](#)
- [http://www.parc.com/event/
934/adventures-in-
searchland.html](http://www.parc.com/event/934/adventures-in-searchland.html)
- Video and audio

+

Launching a hot product...





After the hype, the blogsphere....

The image shows the TechCrunch website homepage. At the top left is the 'TC' logo in a green pixelated font. To the right is a navigation bar with links: HOME, STARTUPS, MOBILE, GADGETS, EUROPE, and V. Below the navigation is an advertisement for 'Family Messaging Unlimited' featuring a smartphone and the text 'only \$30 /month'. At the bottom of the header is a sub-navigation bar with links: HOT TOPICS, APPLE, FACEBOOK, GOOGLE, ANDROID, DISRUPT NYC, and HACKTHON.

NEWS How To Lose Your Cuil 20 Seconds After Launch

Comment 0

 ERICK SCHONFELD

Tweet 0

Share

+1 0



Tuesday, July 29th, 2008

Comments

The hype cycle now lasts less than a day. Take yesterday's **over-hyped launch** of stealth search startup **Cuil**, which was quickly followed by a backlash when everyone realized that it was selling a bill of goods. This was entirely the company's own fault. It pre-briefed every blogger and tech journalist on the planet, but didn't allow anyone to actually test the search engine before the launch.

The company's founders have a **good pedigree**, and have developed a unique way to index the Web cheaply and at massive scale. But creating a big index is only half the battle.

A good search engine has to bring back the best results from that haystack as well. Here Cuil falls short, as **we pointed out** an hour after the site launched and we could actually check it out.

The story quickly turned from Cuile to Cuile's launch (make that an amuse bouche). The

+The reasons for Cuil

- There is (too much) information on the web.
- Cuil 'organizes' the web so that you can find information that you didn't know you wanted..

The screenshot shows the Cuil search interface with the query 'melanoma' in the search bar. The results are organized into tabs: All Results, Malignant Melanoma, Ocular Melanoma, Nodular Melanoma, and more... The 'Malignant Melanoma' tab is selected. The first result is a link to Wikipedia's page on Melanoma, which includes a snippet about amelanotic melanomas and a histological image. The second result is a link to 'Melanoma – skin cancer reviewed' from melanoma.blogsome.com, featuring a snippet about nodular melanoma and two small images. The third result is a link to the 'MRF Home' page, which has a snippet about supporting melanoma research and education, along with a photo of a person running on a beach.

melanoma

Search Preferences English ▾

All Results Malignant Melanoma Ocular Melanoma Nodular Melanoma more... Melanoma Vaccine

Melanoma

Wikipedia: Amelanotic (colorless or flesh-colored) **melanomas** do not have pigment and may not even be visible. Lentigo maligna, a superficial **melanoma** confined to the topmost layers of the skin (found primarily in older patients) is often described as a "stain" on the skin. Some patients with metastatic **melanoma** en.wikipedia.org/wiki/Melanoma

Melanoma – skin cancer reviewed

Nodular **melanoma** (about 15% of cases) – This type of **melanoma** rises very rapidly and is the most aggressive type of skin cancer. From the beginning this type of **melanoma** grows vertically up and down. So there is a danger from the beginning that **melanoma** will spread in to inner tissues. This type of **melanoma** melanoma.blogsome.com/

MRF Home

To support medical RESEARCH for finding effective treatments and eventually a cure for **melanoma**. To EDUCATE patients and physicians about the prevention, diagnosis and treatment of **melanoma**. To act as an ADVOCATE for the **melanoma** community to raise the awareness of this disease and the need for a cure. www.melanoma.org/

+The reasons are still there...

- Reports estimate we can see only 15% of the existing web.
- Probing the web is mostly **popularity based**. You're likely to see what others have seen before. But your seeing increases the popularity of what you saw, thereby reducing the pool of available stuff.
- Vicious or virtuous circle? How to measure?
- Eli Pariser, The Filter Bubble, book and TED talk,

www.ted.com/talks/eli_pariser_beware_online_filter_bubbles.html

+

Rearden Commerce? What's that? ...

The screenshot shows the homepage of the Rearden Commerce website. At the top left is the Rearden Commerce logo, featuring a stylized red 'R' icon followed by the text 'REARDEN COMMERCE'. The top navigation bar includes links for 'Business Solutions', 'Consumer Solutions', 'Merchant Services', 'Mobile', 'News', and 'Company'. To the right of the navigation is a dark blue header area containing a large red rectangular button with white text that reads 'WELCOME TO SMART COMMERCE.'. Below this button is a black callout box with white text that says 'Seamless. Integrated. Relevant. Empowering.' and 'Say "hello" to Deem. Watch the video. >'. The background of the page features a dark, abstract graphic of glowing orange lines forming a network or globe against a dark background.

REARDEN
COMMERCE

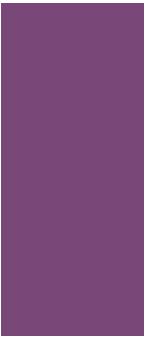
Business Solutions Consumer Solutions Merchant Services Mobile News Company

WELCOME TO
SMART COMMERCE.

Seamless. Integrated. Relevant. Empowering.
Say "hello" to Deem. Watch the video. >

+ Women in Engineering and diversity...

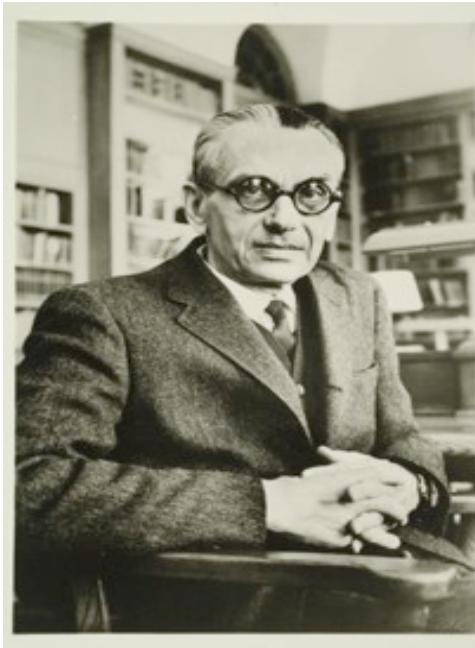
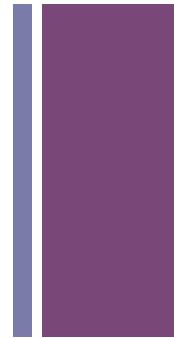
Diwali celebration 2011





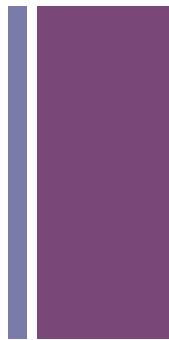
Inspiration?

Universidade de Cambridge DPMMS





Unity Behind Disparity: Proofs are Programs?...



- It's no secret that the bulk of mathematics, as we know it today got crystallized in the last years of the 19th century, first years of the 20th century.
- The shock is still being felt. Notices of the AMS Jan 2012:
[A Revolution in Mathematics? What Really Happened a Century Ago and Why It Matters Today](#) Frank Quinn
- **Very controversial...**
- Interested today in two aspects: notions of **Proofs** and the birth of **Algebra**...

⁺ The easy one: Edwardian Algebra?

- (Abstract) Algebra: area of mathematics that studies algebraic structures, such as **groups**, **rings**, **fields**, **modules**, **vector spaces**, and **algebras** generally...
- Abstract (instead of elementary) algebra: turn of the 20th century, van der Waerden first textbook 'Modern Algebra', 1930.
- Sources: linear algebra, permutation groups, diophantine equations,...

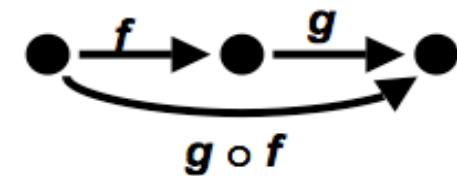
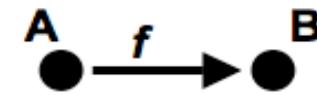
+ Category Theory: 1945-now

- Basic idea: there's an underlying unity of mathematical concepts/theories. More important than the mathematical concepts themselves is how they relate to each other.
- Topological spaces come with **continuous maps**, while vector spaces come with **linear transformations**, for example.
- **Morphisms** are how structures transform into others in the (reasonable) way to organize the mathematical edifice.
- Detractors call CT abstract Nonsense
- The language of CT is well-accepted in all branches of Math, the praxis and the philosophy less so.

+ Categories: a picture

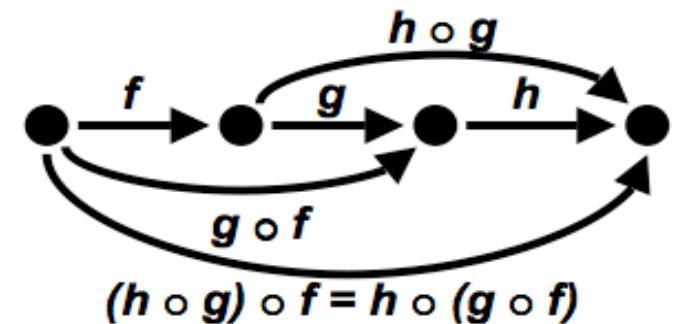
A category consists of:

- a class of *objects*
- a class of *morphisms* (“arrows”)
- for each morphism, f , one object as the *domain* of f and one object as the *codomain* of f .
- for each object, A , an *identity morphism* which has domain A and codomain A . (“ ID_A ”)
- for each pair of morphisms $f:A \rightarrow B$ and $g:B \rightarrow C$, (i.e. $\text{cod}(f) = \text{dom}(g)$), a *composite morphism*, $g \circ f: A \rightarrow C$



With these rules:

- *Identity composition*: For each morphism $f:A \rightarrow B$,
 $f \circ ID_A = f$ and $ID_B \circ f = f$
- *Associativity*: For each set of morphisms $f:A \rightarrow B$,
 $g:B \rightarrow C$, $h:C \rightarrow D$,
 $(h \circ g) \circ f = h \circ (g \circ f)$



+ Category Theory

- Category: a collection of objects and of morphisms, satisfying obvious laws
- Functors: the **natural** notion of morphism between categories
- Natural transformations: the natural notion of morphisms between functors
- Constructors: products, sums, limits, duals....
- Adjunctions: an abstract version of **equality**
- Does this relate to logic?



Back to Proofs?

Mathematical Logic divided in: Model theory, Proof theory, Recursion theory and Set theory

Category theory used for Model theory and Proof theory

Proof theory is Hilbert's brainchild:
'proofs' as objects of mathematical study

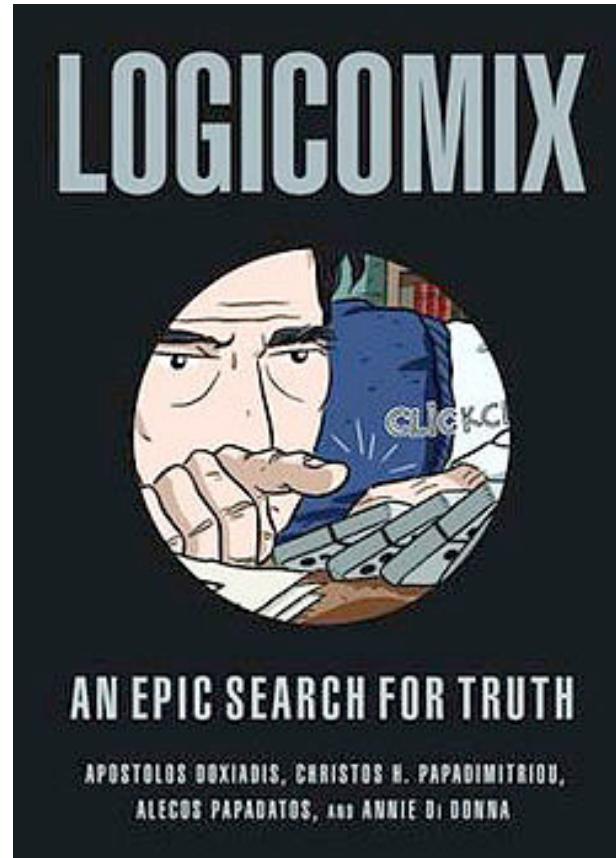
Famous address in 1900 to ICM, 23 problems that shaped Mathematics
“...in mathematics there is no 'ignorabimus',”





Which Proofs?

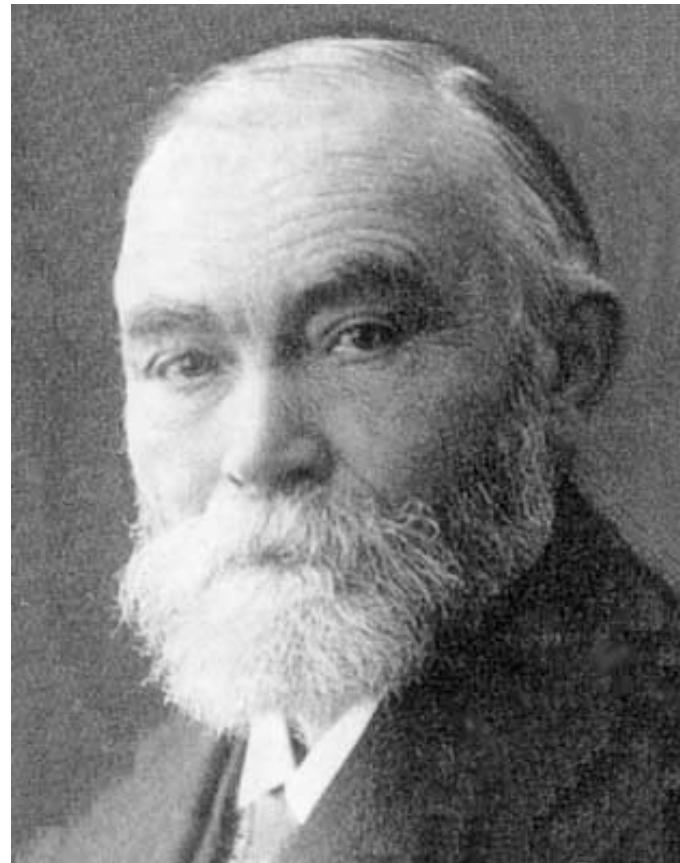
- Mathematics in turmoil in the turn of the century because of paradoxes e.g. Russell's
- Hilbert's Program: Base all of mathematics in finitistic methods
- Proving the consistency of Arithmetic the big quest
- Read the graphic novel!!





Edwardian Proofs? Frege

- One of the founders of modern symbolic logic put forward the view that mathematics is reducible to logic.
- *Begriffsschrift, 1879*
- Was the first to write proofs using a collection of abstract symbols





Proofs of the 20th century? Gentzen Proofs

- To prove the consistency of Arithmetic Gentzen invented his systems of
- NATURAL DEDUCTION

how mathematicians think...

- SEQUENT CALCULUS

how he could formalize the thinking to obtain the main result he needed, his *Hauptsatz*. (1934)



Illustration: Curry-Howard for Implication

Natural deduction rules for implication (without λ -terms):

$$\frac{A \rightarrow B \quad A}{B} \rightarrow_E \frac{[A] \quad \vdots \quad \vdots \quad B}{A \rightarrow B} \rightarrow_I$$

Natural deduction rules for implication (with λ -terms):

$$\frac{M : A \rightarrow B \quad N : A}{M(N) : B} \rightarrow^E \quad \frac{[x : A] \quad \begin{matrix} \\ \vdots \\ M : B \end{matrix}}{\lambda x. M(x) : A \rightarrow B} \rightarrow_I$$

Function application *λ -abstraction*



Church and lambda-calculus

- Alonzo Church introduced the lambda calculus in 1932.
- By 1936 Church had realized that lambda terms could be used to express every function that could ever be computed by a machine.
- Instead of saying ‘the function f where $f(x) = t$ ’, he simply wrote $\lambda x. t$.
- The lambda calculus is an universal programming language.
- The Curry-Howard correspondence led logicians and computer scientists to develop a cornucopia of new logics based on the correspondence between proofs and programs.



Curry-Howard correspondence?

Basic idea:

types are formulae/objects in **appropriate** category,

terms are proofs/morphisms in same category,
logical constructors are appropriate
categorical constructions.

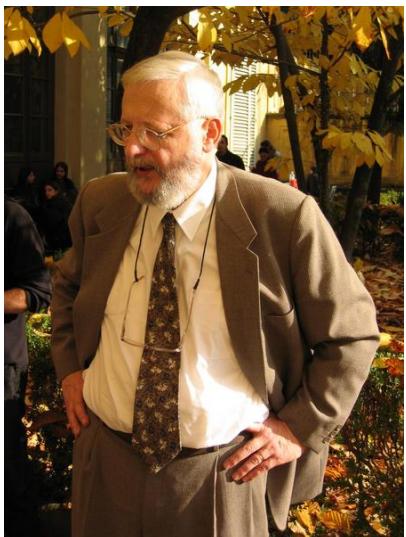
Most important: Reduction is proof
normalization

Outcome: Transfer results/tools from logic to
CT to CS

.

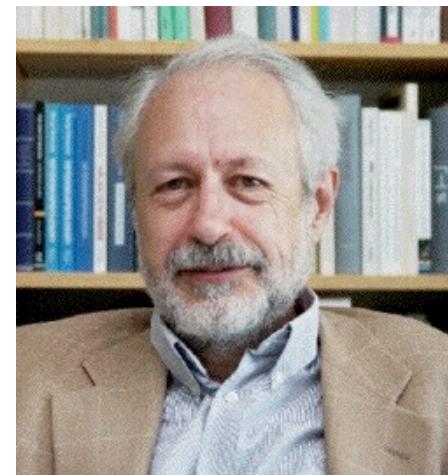


Curry-Howard Correspondence Triangles



1963

Lambda-
calculus



1965

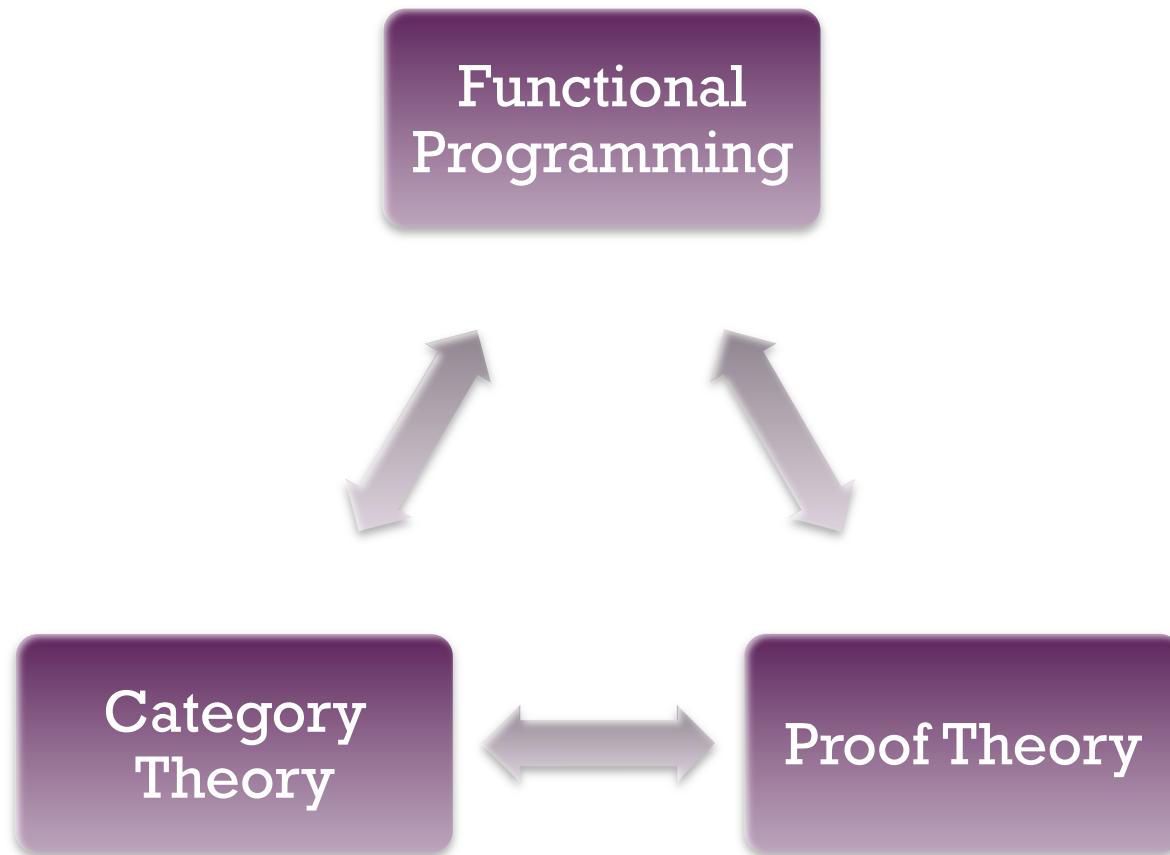
Cartesian
Closed
Categories

Intuitionistic
Propositional
Logic



+

Curry-Howard Correspondence Triangles



A long time in the making: Curry analogy, Howard (1969), published 1981, book 1986 ..

Categorical Proof Theory

models derivations/proofs, not whether theorems are true or not

Proofs definitely first-class citizens

How? Uses extended Curry-Howard isomorphism

Why is it good? Modeling derivations useful in linguistics, functional programming, compilers..

Why is it important? Widespread use of logic/algebra in CS means new important problems to solve with our favorite tools, as well as jobs for mathematicians & logicians.

Why so little impact on math itself?



Hey, where are my futuristic programs?

“Time and again, a logician motivated by logical concerns and a computer scientist motivated by practical concerns have discovered exactly the same type system, usually with the logician getting there first. Most modern functional languages use the Hindley-Milner type system, discovered by the logician Hindley in 1969 and re-discovered by the computer scientist Milner in 1978.”

- Prof P. Wadler, University of Edinburg, “Proofs are Programs: 19th Century Logic and 21st Century Computing”, Avaya 2000





Successes/Opportunities of Categorical Semantics

- +Models for the untyped lambda-calculus
- +Typed programming language & polymorphism
- +Dependent Type Theory
- +Operational Semantics & Full abstraction results
- +Game Semantics
- Proof theory of Classical Logic/Modal logics
- Effect full computation, mobile/grid computing, etc
- Practical impact
- Acceptance of methodology in math

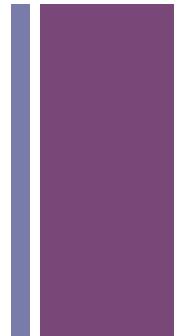


Languages & provers

- StandardML Cornell
- Haskell
- O'Caml
- Scala
- F#
- HOL
- Isabelle
- Twelf, Elf,...
- Coq
- TAL, TIL, PCC

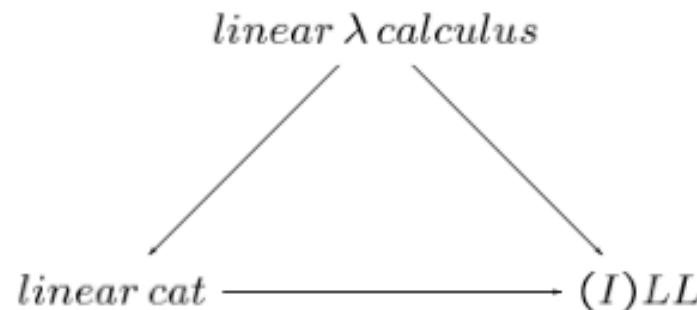
Warning: FP In the Real world...

+ How do I get into this? The same picture...



Categorical Semantics of (Intuitionistic) Linear Logic

Want linear version of Extended Curry-Howard Isomorphism



- Logic intuitionistic
- Already have Intuitionistic Linear Logic (ILL),
 - must find other two sides of triangle...

+ ...different logic

Linear Logic, a proof theoretic logic described by Jean-Yves Girard in 1986.

Basic idea: assumptions cannot be discarded or duplicated. They must be used exactly once—just like dollar bills...

Other approaches to accounting for logical resources. Great win of Linear Logic:

Account for resources when you want to, otherwise fall back on traditional logic,

$$A \Rightarrow B \text{ iff } !A \multimap B$$



+ My example: Dialectica categories

Nice version of Curry Howard correspondence for Intuitionistic LINEAR LOGIC in 1987

Why is it nice?

Godel's Dialectica interpretation (1958)

Gödel's **result**: an interpretation of intuitionistic arithmetic HA in a quantifier-free theory of functionals of finite type T

idea: translate every formula A of HA to $A^D = \exists u \forall x.A_D$ where A_D is quantifier-free.

use: If HA proves A then T proves $A_D(t, y)$ where y is a string of variables for functionals of finite type, t a suitable sequence of terms not containing y .

+ Dialectica categories

- Goal: to be as constructive as possible while being able to interpret all of classical arithmetic
- De Paiva 1987 „Dialectica Categories in AMS Boulder Meeting, vol 92: An internal categorical version of the Dialectica, modelling the main connective, implication

If $A^D = \exists u \forall x. A_D$ and $B^D = \exists v \forall y. B_D$

$$(A \Rightarrow B)^D = \exists V(u) \exists X(u, y) \forall u \forall y [A_D(u, X(u, y)) \Rightarrow B_D(V(u), y)]$$

+ What is the point?

- First, the construction ends up as a model of Linear Logic, instead of constructive logic. Then it allows us to see where the assumptions in Gödel's argument are used
- It justifies linear logic in terms of more traditional logic and conversely explains the more traditional work in terms of a 'modern' (linear, resource conscious) decomposition of concerns.
- Theorems(87/89): Dialectica categories provide models of linear logic as well as an internal representation of the dialectica interpretation. Modeling the exponential ! Is hard, first model to do it.

+ Dialectica categories:

Very pretty: models of Intuitionistic and classical linear logic and special connectives that allow us to get back to usual logic

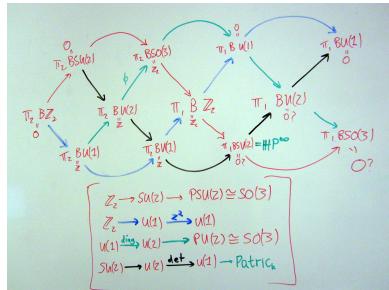
Extended it in a number of directions—a robust proof can be pushed in many ways.

In CS as a model of Petri nets(> 2 phds), it has a non-commutative version for Lambek calculus (linguistics), it has been used as a model of state (Correa et al) and even of quantum groups.

Also used for generic models of Linear Logic (with Schalk04) and for Linguistics Analysis of the syntax-semantics interface, the Glue Approach (Dalrymple, Lamping and Gupta).



Bridges?



culum avencimento in
membrum tuum iustitiae
meum iustitiae
quibus a sumo capite
in his ad summum pede
iustitiae huius et
ab impetuoso enim
scabie laniatus et si
assimile sanguine tuo
rubratur quia man
intudinem dolere in
magnum carne tua
pro nobis pertulisti
Pie ihesu quid ultra de
bunstu facere quod non



Translation





NOT TOY BRIDGES ...

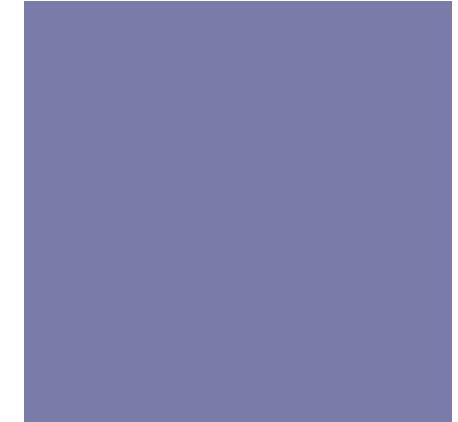
Categorical Logic, Dialectica
Categories and Their Applications

Automated Theorem Proving and
Semantics of Programming Languages

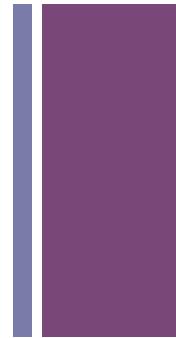
Linear Functional Programming,
Abstract Machines and Modal Type
Theories

Logics for NLP, for Linguistic Inference
and for Contexts in AI

Combining symbolic methods and
analytics to solve big data problems



+20 years later



- Extension to higher-order logic via topos theory, Bodil Biering phd thesis (2008),
- The Copenhagen interpretation
- also connections between different functional interpretations (Oliva, Gernest and Trifonov), via Shirahata's work (TAC vol 17).



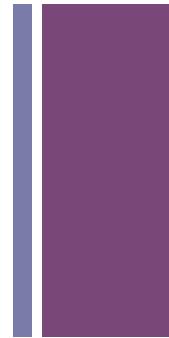
Taking Stock

Working in interdisciplinary areas is hard,
but rewarding.

- The frontier between logic, computing, linguistics and categories is a fun place to be.
- Mathematics teaches you a way of thinking, more than specific theorems.
- Barriers: proprietary software and unwillingness to 'waste time' on formalizations
- Enablers: international scientific communities, open access, NLP and other open software, growing interaction between fields,...
- Handsome payoff expected...
- Fall in love with your ideas and enjoy talking to many about them...



NASSLI 2012: nasslli2012.com/ Come to Texas for the fun...



The Fifth North American
Summer School of Logic,
Language, and Information

at the University of Texas at Austin

Johan van Benthem University of Amsterdam /
Logical Dynamics of Information and Interaction

Craig Roberts Ohio State University
Questions of Discourse

Noah Goodman Stanford University
Stochastic Lambda Calculus
and its Applications in Semantics and Cognitive Science

Mark Steedman University of Edinburgh
Combinatory Categorical Grammar: Theory and Practice

Chris Potts Stanford University
Extracting Social meaning and Sentiment

June 18–22, 2012

nasslli2012.com

Registration: \$175 (academic rate) / \$400 (professional rate)
Student scholarships available, see website for application instructions
Accommodation provided for \$70 / night (single) or \$35 / night (double)

Additional Courses:

- Catherine Legg - University of Victoria
Foundations of Computational Semantics
(for Computer Scientists and Linguists)
- Adam Lopez - Johns Hopkins University
Social Choice Theory for Logicians
- Valerie de Paiva - Radon Center
Urbachova 19, Bratislava, Slovakia
Introduction to Category Theory
- Adam Poerw - Radon Center
Urbachova 19, Bratislava, Slovakia
On-line Semantics and Application with SUMOTM
- Edzard Reckermann - University of Frankfurt
Intensionality
- Thomas Icard - Stanford University
Surface Reasoning
- Nina Gierasimczuk - University of Goettingen
Initial Revision Meets Formal Learning Theory

Special Events:

- June 14–17: *Teaching Semantics*, University of Paris
- June 14–17: *Turing Centennial Conference*, Royal Holloway College, University
Type theory with records for natural language
interpretation
- June 21–24: *Special sessions on American Sign Language,
Semantics, and Pragmatics*, University of Massachusetts
Details regarding Call for Papers at nasslli2012.com
- June 23: *Turing Centennial Symposium*

<http://www.cs.bham.ac.uk/~vdp/>
<http://www.valeriadepaiva.org/>

+

Thanks!



+ Some references

- Godel's Collected Works, eds Feferman and Dawson
- Categories for the Working Mathematician Saunders MacLane
- The Curry-Howard Correspondence ed de Groote, Gallier's paper
- Proofs and Types – Jean-Yves Girard
- <http://www.cs.bham.ac.uk/~vdp/publications/papers.html>



Projects, Forums: a small sample

- CLiCS: Categorical Logic in Computer Science, I and II from Oct 89-Dec95, report:
<http://www.disi.unige.it/person/MoggiE/PROJECTS/clics/>
- APPSEM I and II: Apr 1998-Mar 2002 and Jan03-Dec05
- LINEAR: May 1998-Apr2002
- CMU/Penn Manifest Security
<http://www.cis.upenn.edu/~plclub/ms/>
- Grey project, ZAP(Princeton) 06-10
- LOGOSPHERE: digital libraries/mathematical knowledge
- TRUST – Stanford Security Lab

+ Applications of categories in CS

- semantics/design of programming languages:
 - Lisp, ML, O'CamL, **Haskell**, (Lua?), etc.
- Compilers, proving optimizations in code, ...
- categorical semantics of specification (work on Institutions,)
 - security, (CMU, Microsoft, etc)
 - concurrency...
- Modeling databases (sketches?)
- (at large) functional programming, language design (Type hackery – SPJ)
- Formal methods and high assurance
- interactive theorem proving (HOL, Isabelle, Coq, PVS, Twelf, HOLlight..)

+ Dialectica categories

Assume \mathbf{C} is cartesian closed category

DC objects are relations between U and X
monics $A \xrightarrow{\alpha} U \times X$ written as $(U \xleftarrow{\alpha} X)$.

DC maps are pairs of maps of \mathbf{C}

$f: U \rightarrow V, F: U \times Y \rightarrow X$

making a certain pullback condition hold.

Condition in Sets: $u\alpha F(u, y) \Rightarrow f(u)\beta y$.

+ Prospects?

- Still want to see more work on Hopf algebras and dialectica categories: quantum mechanics anyone?
- Same about state modeling and separation logic: a couple of ideas not written anywhere
- Also about super power games and relationships between functional interpretations



Why do we want this bridge?

- Want our computer to understand us
- Want to ask questions
 - large-scale intelligent information access
- Want to search for content on the web
 - exploiting the large amounts of textual data on the web
- Want automatic, high quality translation into other languages
- Want intelligent, grammatical summarization, etc...



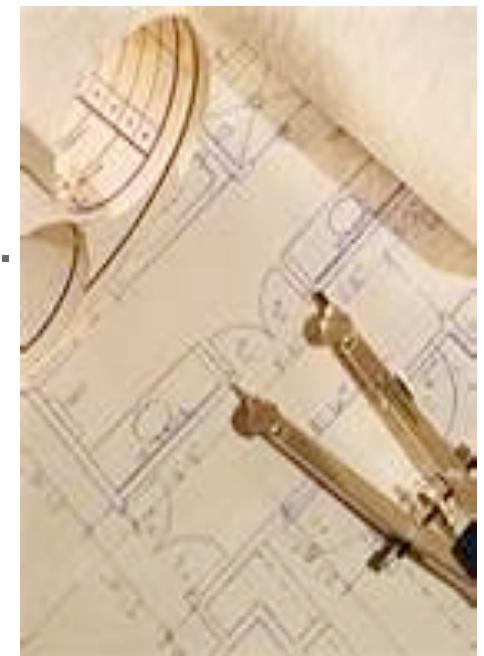


Blueprint for bridge between language and logic

- translation compositional and principled,
- meaning preserving, at least truth value preserving...
- a reasonable fragment of all language
- generic and specific texts
- “logical forms” obtained are useful for **reasoning**.

Questions:

- which kind of logic on the target?
- how do we know when we’re done?
- how do we measure quality of results?





What makes language so hard?

- Problems with **non-existence**
“Negotiations prevented a strike”
- Bad idea: $\exists x \exists y \text{Strike}(x)$ and Negotiations (y) and prevented(x,y)
- Could do existence predicates, but entailments don't work.
Instead: Do **concepts!** see “Preventing Existence”, Crouch et al, FOIS2001
- Problems with intensionality...
Anecdotal evidence at least 453 sentences out of 1586 on 100 tips photocopier repair will have a propositional complement
- Must deal with intensionality, use **Contexts!** See “Entailment, Intensionality and Text Understanding”, Condoravdi et al, 2003
- Ambiguity is rampant and pervasive!
- PACK and/or CHOOSE!





NOT ONLY A TOY PLEASE...

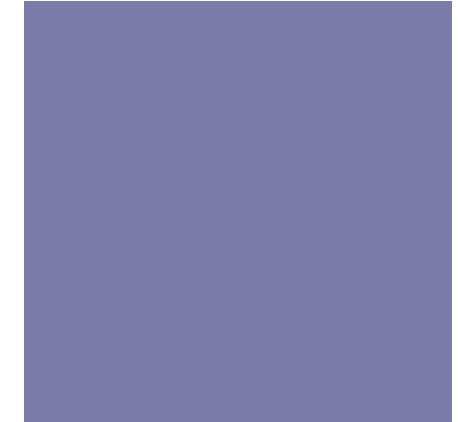
Need to deal with huge amounts of text

Need to be robust

Need to move easily between genres, subject domains

...

Where's that bridge? Where does it lead to?





NLP: Symbolic or Data Driven?

- Knowledge-based representations of meaning
- Deep/logical representations allow high precision and recall, but
- Typically on restricted domains
- Hard for users to read/interpret output of the system
- **Very hard for systems to build up knowledge**
- Shallow, open-domain representations
- Broad-coverage
- Fails “gracefully”
- Lower precision and recall
- Hard to compare systems
- Too sensitive to form of sentences

Obvious that we need a mixture, which one is the hard question..

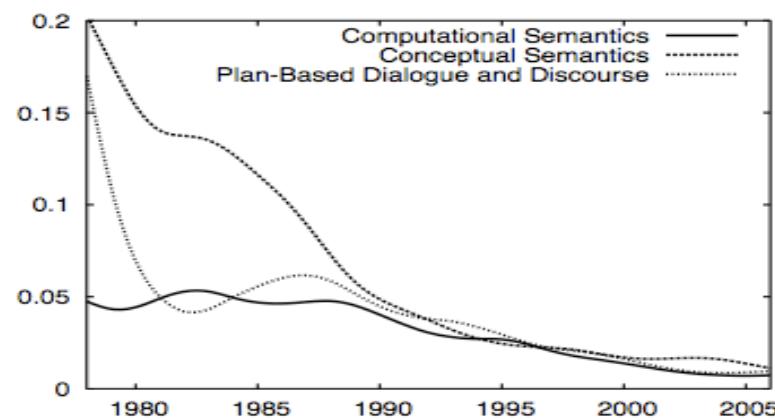
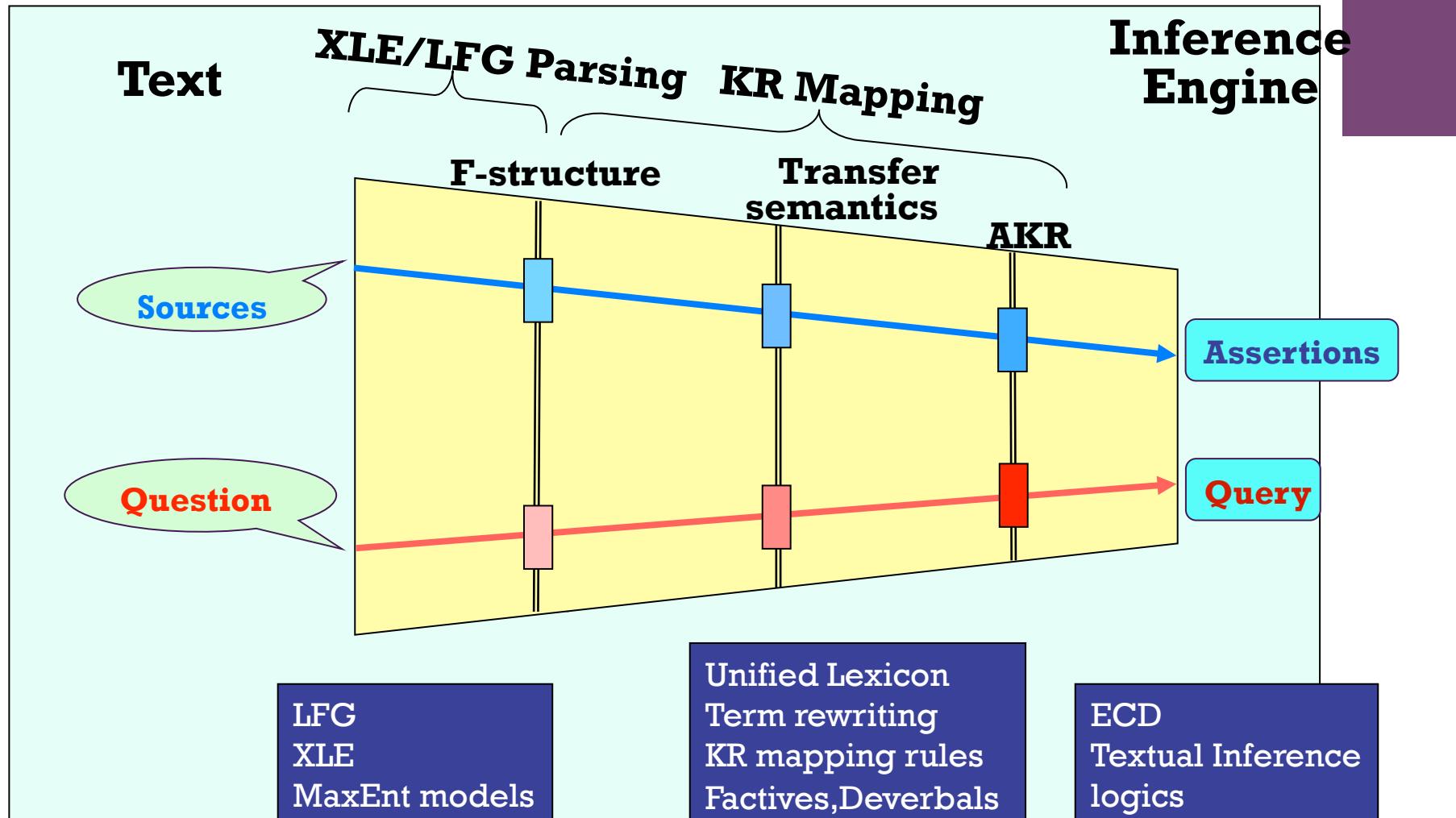


Figure 2: Topics in the ACL Anthology that show a strong decline from 1978 to 2006.

+

PARC Layered Architecture



Basic idea: canonicalization of meanings