Logical Foundations for the Infrastructure of the Information Market

Heather, Michael, Livingstone, David, &
Rossiter, Nick,

CEIS, Northumbria University, Newcastle, UK

Outline of Talk

- Unstructured and Structured Data
- Implications of Gödel
  - Relational Model
  - Interoperability
- Difficulties of SQL with Gödel
- Need in interoperability
  - higher order formalism without axiom or number
- Applied category theory
  - topos (with composition, adjointness)

Simple data is well-handled

- Information Systems
  - Well established for simple data
    - Unstructured
      - e.g. web pages, natural language, images
    - Structured
      - e.g. relational database

Natural and Structured Data Types

Type	Structure	Examples	Applications
Images	Structured	Graphics	Business graphics
	Natural ('Unstructured')	Photographs	Publishing
Text	Structured	Relational databases	Business data
	Natural ('Unstructured')	The web	Google
Intermediate Data	Meta-structure added to unstructured data	Semi-structured data (as in XML) Semantic Web CAD/CAM Engineering Drawing with instructions	Fitting spare parts
Process Data	Dynamic	Banking Transactions Biometric identity	ATM accounting Iris data

Fig. 1. Natural and Structured Data Types

Interoperability

- Relatively easy between natural (unstructured) data:
  - natural language + translators
  - images + human eye
- Difficult between structured data:
  - schema is reductionist
  - inter-communication is problematical

Underlying Difficulties

- Relational Databases
  - based on first-order predicate calculus (FOPC)
- Efforts by Codd and Date
  - to keep narrowly within FOPC
    - atomic data (first normal form)
    - nested data is encapsulated
    - operations are within standard first-order set theory
      - arguments are sets
- But interoperability requires higher-order operations

Effect of Gödel

- Gödel showed that:
  - Both intensional and extensional systems that rely on axiom and number are undecidable
  - But FOPC is complete
- Therefore
  - Strict relational model and calculus is complete and decidable
  - Higher order systems that rely on axiom and number are not complete and decidable

Inherent Difficulty of Interoperability

- As interoperability is always higher order
  - from mapping functions to functions
- Interoperability is outside natural applicability of set theoretic methods

Particular Problems with SQL

- SQL has compromised the pure relational model
- So SQL poses special problems in interoperability
  - Variants in its implementation
  - Not faithful to relational model
  - Closed world assumption
  - Nulls

Variants of SQL

Features	Achievements	Problem in interoperability
Full facilities	Not achieved by MySQL	Very difficult between MySQL and other DBMS
Hierarchies, manipulation	Peculiar to Oracle	Difficult between DB2 and Oracle in network/hierarchical structures
Recursive union, assembling networks	Peculiar to DB2	
Implementation of integer type	Oracle treats as numeric(38)	Difficult between Oracle and other DBMS in formatting and rounding numbers
Dates	Different logical formats	Difficult between all systems in reliable data format recognition

Fig. 2. Effects of Variants of SQL on Interoperability

Not Faithful to Relational Model

Feature in SQL	Feature in relational model	Consequence for SQL
Default structure is bag	Default structure is set	Duplicate rows permitted, inconsistency in updates
Row identifier	No identifiers	Physical bias to extension
Rows may be sequenced	No sequencing of rows	Data is apparently ordered
Set operations such as union based on column position	Set operations such as union based on column name	Set operations are based on physical, not logical, ordering of columns
Duplicate column names allowed in output	No duplicates allowed	Confusing output

Fig. 3. Differences between SQL and the Relational Model

Problems with Nulls

Case	Result	Problem
Creation of nulls whether value is missing-but-applicable or missing-and- inapplicable	No distinction	Semantic simplification
Use of null to represent maybe in the Boolean type	Three values for Boolean logic	Contrary to normal view of Boolean logic as binary valued
Comparing a null value with a null value	maybe with join/restrict, true in set operations	Difference in outcome between set operations and other operations such as join
Split table into a set of sub-tables using restrict; union resulting sub-Tables	No guarantee that this will be the original table	Restrict only returns rows where the comparison returns true; hence those returning null are ignored and lost
Aggregation operators applied to columns containing some nulls	Count includes them; others ignore them	Arbitrary application
Aggregation operators applied to columns containing all nulls	Count returns zero; others return null	Arbitrary application
Second order distributivity (e.g. fuzzy sets)	Logical equivalences are not true	Inconsistent treatment of nulls

Fig. 4. Problems with Handling of Nulls by SQL

Nulls offend Gödel

- Use of nulls gives
  - *maybe* outcome to some queries
  - so result from query is not decidable
- Codd persisted with nulls
- Date has more recently removed them from the ‘pure’ relational model
  - not offend Gödel
  - keep within FOPC

Formalism for Interoperability

- If set theory in general is undecidable and not complete, what might be used?
- Category Theory (CT) has its focus and strengths in higher order logic e.g. functors
  - Pure CT is though axiomatic
  - n-categories rely on number
  - so both offend Gödel
  - Applied CT, based on a process view and of composition, appears to not offend Gödel

Composition in CT
a) with Gödel; b) against Gödel

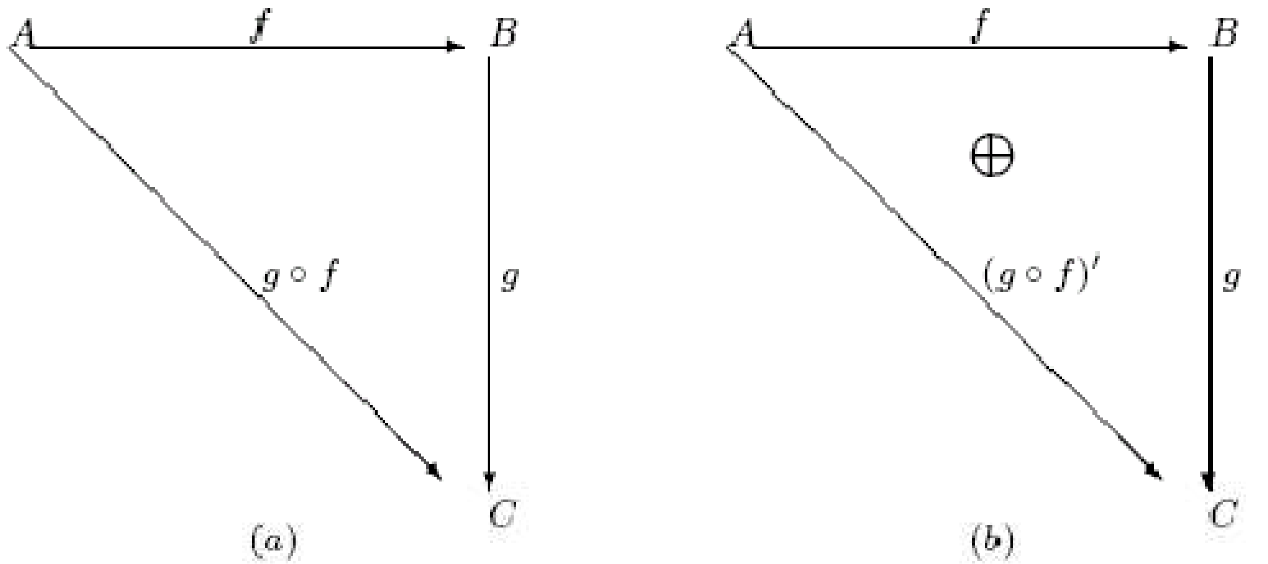

Fig. 5. Commuting Diagrams for a) Composition, b) Punctured Composition: $(g \circ f) \neq (g \circ f)'$

Adjointness between two Composition Triangles

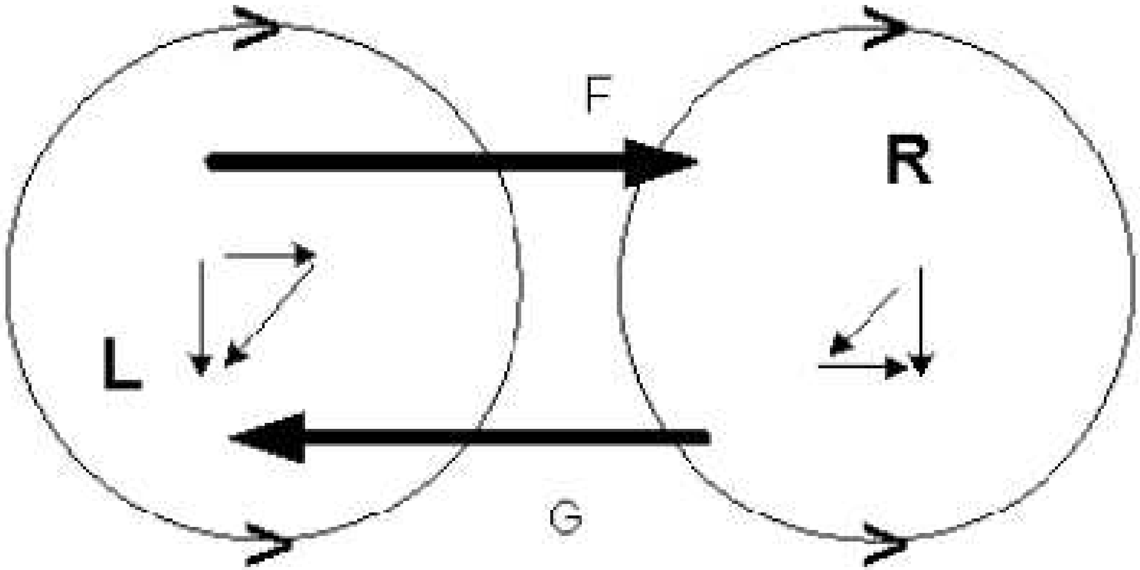

Fig. 6. Adjointness between two Composition Triangles

Composition Triangles in Detail

a) unit of adjunction η; b) co-unit of adjunction ε

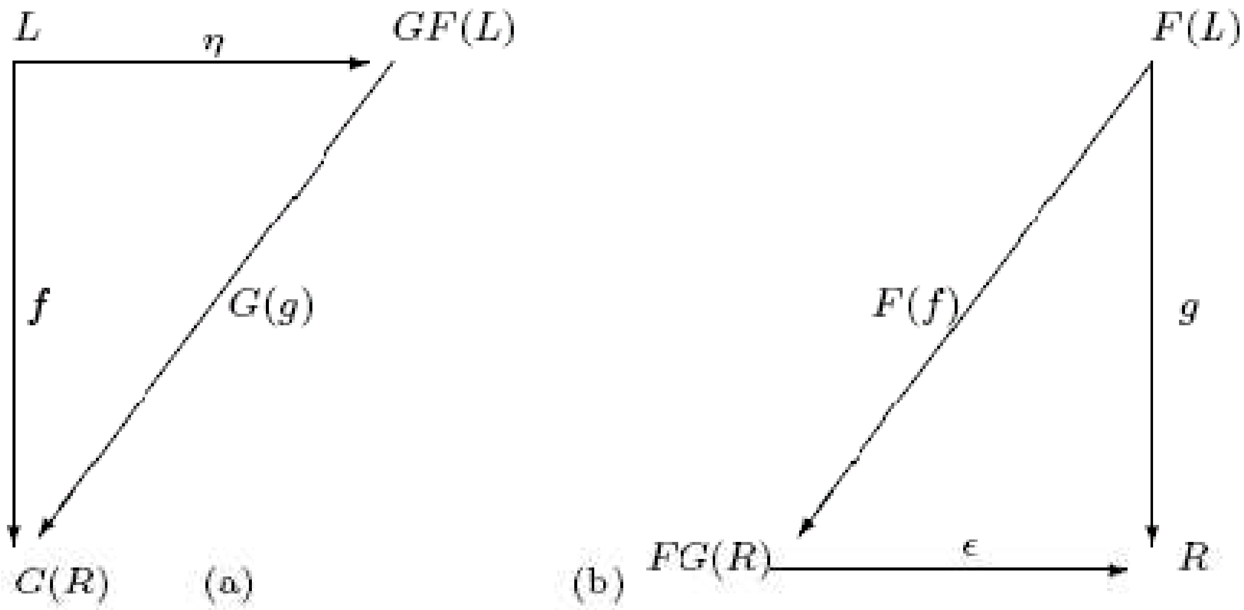

Fig. 7. Roles in Adjointness of a) η, b) ϵ

Architecture for Interoperability

Effective Topos T with interoperability between categories L and R in context of category C

T = SoS (system of systems)

L, R are interoperating systems

C is context of interoperation

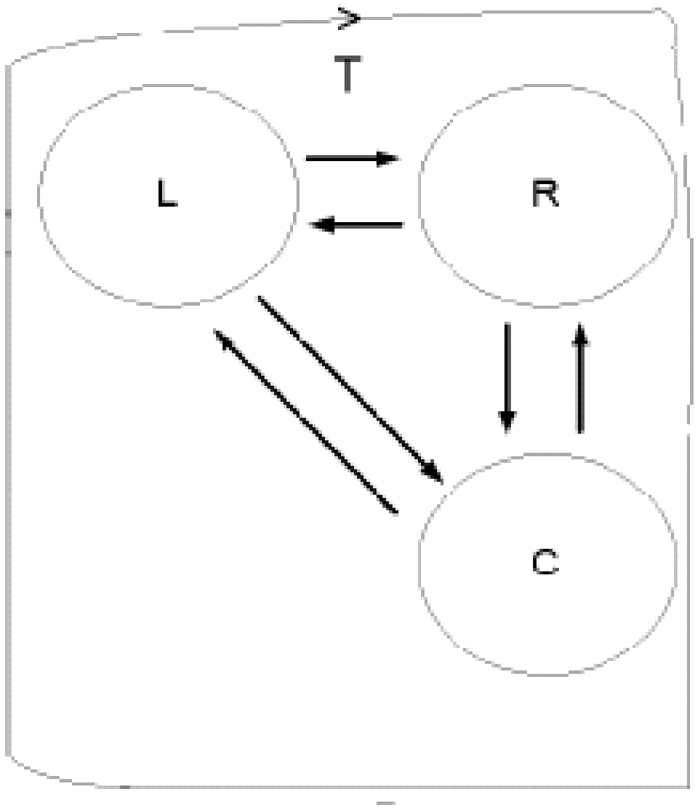

Figure 8

Arrows represent adjointness

Summary

- Implications of Gödel
  - Pure relational model in itself, as first-order predicate calculus, is complete and decidable
  - Interoperability is though higher order
  - Set theory, as defined with axiom and number, is not complete and decidable for higher order
  - Applied category theory, without axiom or number, seems appropriate
- Example architecture given for applied category theory with topos and composition