Modal type theories

Michael Shulman

University of San Diego

September 25, 2023

Outline

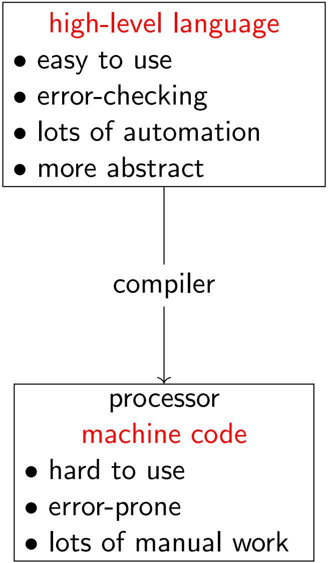
Towards synthetic mathematics

The fundamental objects of Coq, Agda, Lean, HOL, are types.

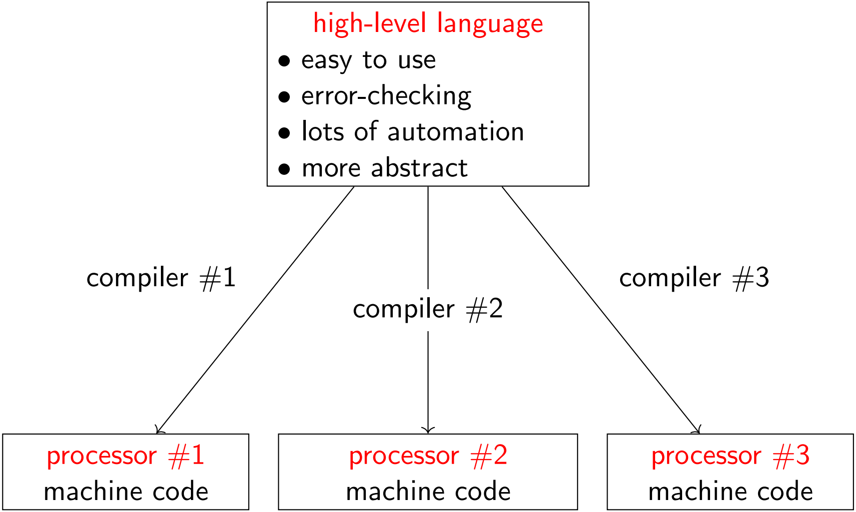
We usually think of types as similar to sets.

However, one of the most powerful aspects of type theory is that types can also be interpreted to have many other structures, just as a high-level programming language can be compiled to run on many different architectures.

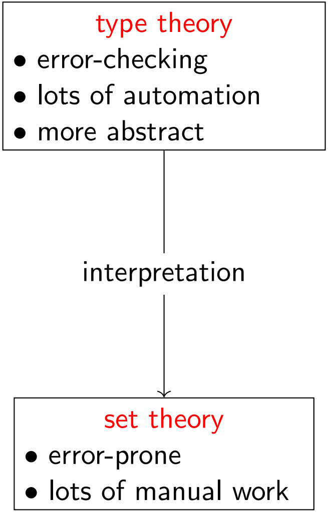
High level programming



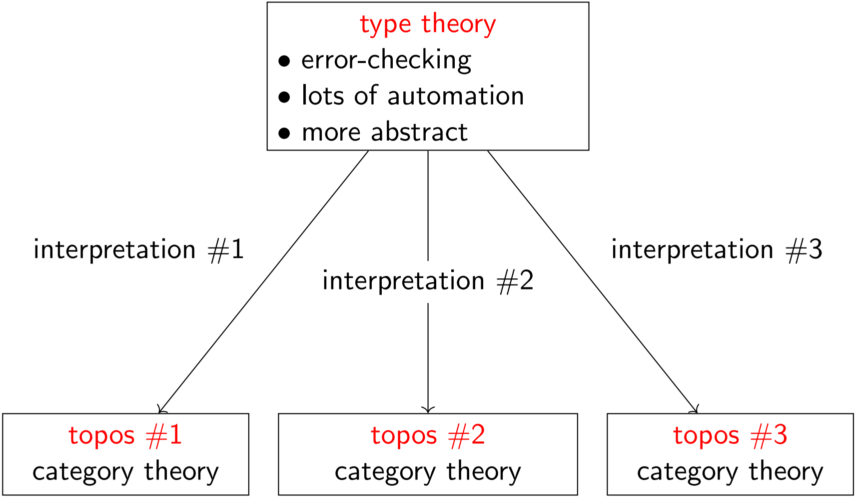
High level programming



High level mathematics



High level mathematics



The type/category dictionary

Syntax	Interpretation in a category $\mathcal{E}$
Type A	Object A
Product type $A \times B$	Cartesian product $A \times B$
Function type $A \to B$	Exponential object B^A
Function $f : A \times B \to C$	Morphism $f : A \times B \to C$
Term $f(x, g(y)) : C$ in context $x : A,\ y : D$	Composite morphism $A \times D \xrightarrow{1 \times g} A \times B \xrightarrow{f} C$
Dependent type $B(x)$ in context $x : A$	Object $B \to A$ of $\mathcal{E}/A$

A plethora of exotic models

Thus, types can potentially represent many kinds of things, like

- sets (classical mathematics)
- ∞-groupoids (homotopy type theory)
- topological spaces (synthetic topology)
- smooth spaces (synthetic differential geometry)
- computable spaces (synthetic domain theory)
- simplicial sets/spaces (synthetic category theory)
- sheaves

Outline

Incorporating exotic structure

The generic interpretation of type theory implies that a theorem in plain type theory is automatically true about any model.

How can we incorporate specifics of one model into type theory?

1. By assuming axioms, e.g.
   - The Axiom of Choice
   - The Law of Excluded Middle: every proposition is true or false.
   - "Brouwer's Theorem": every function $\mathbb{R} \to \mathbb{R}$ is continuous.
   - "Church's Thesis": every function $\mathbb{N} \to \mathbb{N}$ is computable.

   Note in particular that AC and LEM are not true in every model, so in general we must argue constructively.

2. By adding new type-formers...

Adding homotopy to type theory

Ordinary type theory

- Intuition: types as sets, terms as functions.

Homotopy type theory

- New intuition: types as ∞-groupoids, terms as functors.
- Detect their ∞-groupoid structure with the identity type.
- The old intuition is still present in the 0-types.
- Some types that already existed turn out "automatically" to have nontrivial ∞-groupoid structure (e.g. $\mathcal{U}$ is univalent).

Cubical type theory, simplicial type theory are similar.

Adding homotopy to type theory

Ordinary type theory

- Intuition: types as sets, terms as functions.

Homotopy type theory

- New intuition: types as ∞-groupoids, terms as functors.
- Detect their ∞-groupoid structure with the identity type.
- The old intuition is still present in the 0-types.
- Some types that already existed turn out "automatically" to have nontrivial ∞-groupoid structure (e.g. $\mathcal{U}$ is univalent).

Cubical type theory, simplicial type theory are similar.

Adding topology to type theory

Ordinary type theory

- Intuition: types as sets, terms as functions.

Synthetic topology

- New intuition: types as spaces, terms as continuous maps.
- Detect their topological structure... how?
- The old intuition is still present in the discrete spaces.
- Some types that already existed turn out "automatically" to have nontrivial topological structure (e.g. the real numbers $\mathbb{R}$ have their usual topology).

Adding topology to type theory

Ordinary type theory

- Intuition: types as sets, terms as functions.

Synthetic topology

- New intuition: types as spaces, terms as continuous maps.
- Detect their topological structure... how?
- The old intuition is still present in the discrete spaces.
- Some types that already existed turn out "automatically" to have nontrivial topological structure (e.g. the real numbers $\mathbb{R}$ have their usual topology).

The need for discontinuity

In classical mathematics, we have the Intermediate Value Theorem:

Theorem (in classical mathematics)

For any continuous function $f : [a, b] \to \mathbb{R}$ *and point* c *with* $f(a) < c < f(b)$*, there exists* $x \in [a, b]$ *with* $f(x) = c$*.*

In synthetic topology, where all functions are continuous, we expect to drop the adjective:

Theorem? (in synthetic topology)

For any function $f : [a, b] \to \mathbb{R}$ *and point* c *with* $f(a) < c < f(b)$*, there exists* $x \in [a, b]$ *with* $f(x) = c$*.*

The need for discontinuity

In classical mathematics, we have the Intermediate Value Theorem:

Theorem (in classical mathematics)

For any continuous function $f : [a, b] \to \mathbb{R}$ and point c with $f(a) < c < f(b)$, there exists $x \in [a, b]$ with $f(x) = c$.

In synthetic topology, where all functions are continuous, we expect to drop the adjective:

Non-theorem (in synthetic topology)

For any function $f : [a, b] \to \mathbb{R}$ and point c with $f(a) < c < f(b)$, there exists $x \in [a, b]$ with $f(x) = c$.

But then not only would f be continuous as a function of its input, the theorem itself would be continuous as a function of its input f. And even classically, the x cannot be chosen continuously.

Discontinuity

Thus, in synthetic topology we have primitive notions of both (continuous) function and also discontinuous function.

The former form the usual function-types $A \to B$ and $(x : A) \to B$; the latter form a new type $(x :^\flat A) \to B$.

Theorem (in (one version of) synthetic topology)

$$(f :^\flat [a, b] \to \mathbb{R}) \to (c :^\flat \mathbb{R}) \to (f(a) < c < f(b)) \\ \to \exists (x \in [a, b]).\ f(x) = c.$$

I'll sketch a proof of this, after introducing more structure.

Modal operators

We can reify discontinuous functions in two ways:

1. $(x :^\flat A) \to B$ is equivalent to $(x : \flat A) \to B$.
   - $\flat A$ is A "retopologized discretely".
   - $\flat$ is a coreflection into the subcategory of discrete types.
2. $(x :^\flat A) \to B$ is also equivalent to $(x : A) \to \sharp B$.
   - $\sharp B$ is B "retopologized indiscretely".
   - $\sharp$ is a reflection into the subcategory of indiscrete types.
3. It follows that $\flat \dashv \sharp$.

Such unary type operators are called modalities, after the classical $\Box$ ("It is necessary that...") and $\Diamond$ ("It is possible that...") from modal logic.

Internal modalities

A monadic modality like $\sharp$, acting on one universe, can simply be axiomatized inside ordinary MLTT.

$$\begin{aligned}
\sharp &: \mathsf{Type} \to \mathsf{Type} \\
\eta_\sharp &: (A : \mathsf{Type}) \to A \to \sharp A \\
\mu_\sharp &: (A : \mathsf{Type}) \to \sharp\sharp A \simeq \sharp A \\
&\vdots
\end{aligned}$$

But this is not possible for a comonadic modality like $\flat$. The only internalizable comonadic modalities are slicing over a proposition.

So we must modify the judgmental structure in some way, such as with our discontinuous function-types.

Crisp variables

What kind of arguments can $f : (x :^\flat A) \to B$ be applied to? Intuitively, only elements of $\flat A$, not A.

We mark some variables in the context as crisp, written $x :^\flat A$, and say the argument of f can only use those.

Semantically, $x :^\flat A$ is equivalent to $x : \flat A$. Syntactically, we have

$$\frac{\Gamma \vdash f : (x :^\flat A) \to B \qquad \Gamma/\flat \vdash a : A}{\Gamma \vdash f\, a : B}$$

where $\Gamma/\flat$ prevents us from accessing the non-crisp variables.

$$(x : A, y :^\flat B, z : C, w :^\flat D)/\flat \;\cong\; (y :^\flat B, w :^\flat D).$$

We call a term crisp if it is defined in context $\Gamma/\flat$.

(Demo)

Axioms of synthetic topology

We need the following axioms:

1. Crisp LEM: For any crisp proposition P, we have $P \vee \neg P$.
    - Full LEM is non-topological: the union of a subspace and its complement has all the points, but the disjoint union topology.
    - Crisp LEM implies that crisp statements can be proven by contradiction.

Axioms of synthetic topology

We need the following axioms:

1. Crisp LEM: For any crisp proposition P, we have $P \vee \neg P$.
   - Full LEM is non-topological: the union of a subspace and its complement has all the points, but the disjoint union topology.
   - Crisp LEM implies that crisp statements can be proven by contradiction.
2. $\mathbb{R}$ is contractible: For any discrete A (i.e. $A \simeq \flat A$), every map $\mathbb{R} \to A$ is constant, i.e. $(\mathbb{R} \to A) \simeq A$.
   - Intuitively, if A is discrete, a continuous $\mathbb{R} \to A$ must factor through $\pi_0(\mathbb{R}) = 1$.
   - Will see later this is equivalent to $\int \mathbb{R} = 1$.

Axioms of synthetic topology

We need the following axioms:

1. Crisp LEM: For any crisp proposition P, we have $P \vee \neg P$.
   - Full LEM is non-topological: the union of a subspace and its complement has all the points, but the disjoint union topology.
   - Crisp LEM implies that crisp statements can be proven by contradiction.
2. $\mathbb{R}$ is contractible: For any discrete A (i.e. $A \simeq \flat A$), every map $\mathbb{R} \to A$ is constant, i.e. $(\mathbb{R} \to A) \simeq A$.
   - Intuitively, if A is discrete, a continuous $\mathbb{R} \to A$ must factor through $\pi_0(\mathbb{R}) = 1$.
   - Will see later this is equivalent to $\int \mathbb{R} = 1$.
3. Analytic Markov's Principle: If $a, b : \mathbb{R}$ satisfy $a \neq b$, then either $a < b$ or $a > b$.
   - Markov's Principle says that if an algorithm doesn't run forever, then it eventually halts.
   - Think of an algorithm computing better and better approximations to a and b, halting if it finds a difference.

Connectedness of $\mathbb{R}$

Lemma ($\mathbb{R}$ is connected)

If $\mathbb{R} = U \cup V$ with $U \cap V = \emptyset$, then either $\mathbb{R} = U$ or $\mathbb{R} = V$.

Proof.

Given the assumption, we can define $f : \mathbb{R} \to \mathrm{Bool}$ by

$$f(x) = \begin{cases} \mathrm{true} & \text{if } x \in U \\ \mathrm{false} & \text{if } x \in V. \end{cases}$$

But Bool is discrete, since crisp discrete types are coreflective and hence closed under colimits, and $\mathrm{Bool} \simeq \top \sqcup \top$.
Thus, since $\mathbb{R}$ is contractible, f is constant. If it is constant at true, then $\mathbb{R} = U$; if it is constant at false, then $\mathbb{R} = V$. □

A similar argument applies to any interval $[a, b] \subseteq \mathbb{R}$.

The Intermediate Value Theorem

Theorem (IVT)

Let $f :^\flat [a, b] \to \mathbb{R}$ and $c :^\flat \mathbb{R}$ be crisp, and suppose $f(a) < c < f(b)$. Then there exists $x \in [a, b]$ with $f(x) = c$.

Proof.

By Crisp LEM, we may assume for contradiction that $f(x) \neq c$ for all $x \in [a, b]$.
Let $U = \{ x \mid f(x) < c \}$ and $V = \{ x \mid f(x) > c \}$.
Our assumption, plus Analytic Markov's Principle, gives $[a, b] = U \cup V$, and clearly $U \cap V = \emptyset$.
So, by the lemma, either $[a, b] = U$ or $[a, b] = V$.
But this contradicts $f(a) < c < f(b)$. □

(Demo)

Outline

Real-cohesive HoTT

Here we think of types as topological ∞-groupoids.

Every type has both ∞-groupoid structure and topological structure. Either, both, or neither can be trivial.

Example

- The higher inductive S^1 has nontrivial higher structure ($\Omega S^1 = \mathbb{Z}$), but is cohesively discrete (no topology).
- $\mathbb{S}^1 = \{ (x, y) : \mathbb{R}^2 \mid x^2 + y^2 = 1 \}$ has trivial higher structure (is a 0-type), but nontrivial cohesion (its "usual topology").

However, S^1 is the "shape" of $\mathbb{S}^1$, written $S^1 = \int \mathbb{S}^1$.
Then $\int \dashv \flat \dashv \sharp$.

We can use this to prove synthetic theorems that relate point-set topology to homotopy theory, such as Brouwer's fixed-point theorem or the Borsuk-Ulam theorem.

Shulman, "Brouwer's fixed-point theorem in real-cohesive homotopy type theory"

Cohesive type theory

Many other interpretations also support the modalities $\flat, \sharp, \int$. We call this cohesive type theory, after Lawvere.

- Smooth spaces (synthetic differential geometry)
- Simplicial spaces (shape is geometric realization)
- Globally equivariant spaces
- ...

Commuting cohesions

Just as real-cohesive HoTT combines cohesion with homotopy theory, we can combine cohesions, e.g. thinking of types as simplicial topological spaces.

- Modalities $\int_{\heartsuit}, \flat_{\heartsuit}, \sharp_{\heartsuit}, \int_{\clubsuit}, \flat_{\clubsuit}, \sharp_{\clubsuit}$.
- $\flat_{\heartsuit}, \flat_{\clubsuit}$ are idemp. comonads, $\int_{\heartsuit}, \sharp_{\heartsuit}, \int_{\clubsuit}, \sharp_{\clubsuit}$ are idemp. monads.
- Adjunctions $\int_{\heartsuit} \dashv \flat_{\heartsuit} \dashv \sharp_{\heartsuit}$ and $\int_{\clubsuit} \dashv \flat_{\clubsuit} \dashv \sharp_{\clubsuit}$.
- $\flat_{\heartsuit} \circ \flat_{\clubsuit} = \flat_{\clubsuit} \circ \flat_{\heartsuit}$, etc.

Myers, Riley, "Commuting Cohesions", 2301.13780

Stable homotopy type theory

Spectra are like ∞-groupoids with abelian group structure. Spectra alone don't have a lot of type-theoretic structure, but we can think of types as families of spectra $(E_x)_{x:A}$ indexed by some ∞-groupoid A (varying with the type).

We have a single modality $\natural$ that zeroes out the spectra, remembering only the indexing space: $\natural(E_x)_{x:A} = (0)_{x:A}$. This is both a monad and a comonad, and self-adjoint $\natural \dashv \natural$.

Riley, Finster, Licata, "Synthetic Spectra via a Monadic and Comonadic Modality", 2102.04099

Synthetic guarded domain theory

Think of types as time-varying sets. For example, objects of the "topos of trees", $\mathsf{Set}^{\omega^{\mathrm{op}}}$.

The "later" modality, defined by $\triangleright A(n) = A(n-1)$ for $n > 0$, and $\triangleright A(0) = 1$, marks

> *"the types of data that may be used only if some 'computational progress' has taken place, thereby enforcing productivity at the level of types" (GKNB).*

Gratzer, Kavvos, Nuyts, Birkedal, "Multimodal Dependent Type Theory", 2011.15021

Directed type theory

Directed type theory envisions types as categories rather than sets or groupoids. (Unlike Rzk, all types are categories, not just those satisfying a condition.)

Since Cat is not locally cartesian closed, Π-types exist only sometimes, for some combinations of fibrational, opfibrational, groupoidal dependency.

We can track these dependencies using modalities:

- A^{op} = the opposite category
- $\mathsf{core}(A)$ = the core (maximal subgroupoid)

Cisinski, Nguyen, Walde, "Univalent Directed Type Theory", CMU Seminar 2023

Outline

What is modal type theory?

A modal type theory consists of

1. One or more ordinary type theories.
2. New unary type formers acting on or between them.
   (Higher-ary type formers make a "substructural" type theory.)
3. Functions relating these type formers and their composites.

Licata, Shulman, Riley, "A Fibrational Framework for Substructural and Modal Logics", FSCD'17

What is modal type theory?

A modal type theory consists of

1. One or more ordinary type theories.
2. New unary type formers acting on or between them.
   (Higher-ary type formers make a "substructural" type theory.)
3. Functions relating these type formers and their composites.

Accordingly, it is specified by a 2-category $\mathcal{M}$, with

1. Objects $p, q, r, \ldots$ called modes.
2. Morphisms $\mu : p \to q, \ldots$ called modalities.
3. 2-cells $\alpha : \mu \Rightarrow \nu, \ldots$ which today I will call laws.

Licata, Shulman, Riley, "A Fibrational Framework for Substructural and Modal Logics", FSCD'17

What is modal type theory?

A modal type theory consists of

1. One or more ordinary type theories.
2. New unary type formers acting on or between them.
   (Higher-ary type formers make a "substructural" type theory.)
3. Functions relating these type formers and their composites.

Accordingly, it is specified by a 2-category $\mathcal{M}$, with

1. Objects $p, q, r, \ldots$ called modes.
2. Morphisms $\mu : p \to q, \ldots$ called modalities.
3. 2-cells $\alpha : \mu \Rightarrow \nu, \ldots$ which today I will call laws.

And it should have semantics in a (pseudo) 2-functor $\mathcal{M} \to \mathcal{C}at$:

1. Each mode represents a category.
2. Each modality represents a functor.
3. Each law represents a natural transformation.

Licata, Shulman, Riley, "A Fibrational Framework for Substructural and Modal Logics", FSCD'17

Modal dependence

- Each mode has its own ordinary type theory.
- For a p-type A and a q-type B, with $\mu : p \to q$,

  $$f : (x :^{\mu} A) \to B$$

  is a function associating, to any x in A, an element of B that depends on x through μ.
- Ordinary $(x : A) \to B$ coincides with $(x :^{1_p} A) \to B$.

Example

In synthetic topology, our 2-category $\mathcal{M}$ has one mode p, one nonidentity modality $\flat : p \to p$, with $\flat\flat = \flat$ and a law $\epsilon : \flat \Rightarrow 1_p$.

Then $(x :^{\flat} A) \to B$ is our discontinuous function-type.

Positive modalities

A modality $\mu : p \to q$ maps a p-type A to a q-type $\mu \boxdot A$, internalizing μ-dependence with a universal property:

$$(x :^{\mu} A) \to B \quad \simeq \quad (y : \mu \boxdot A) \to B$$

- Semantically, $x :^{\mu} A$ and $y : \mu \boxdot A$ are equivalent.
- Syntactically, we have a constructor $\mathsf{mod} : (x :^{\mu} A) \to \mu \boxdot A$ with an induction principle that any $y : \mu \boxdot A$ can be assumed to be $\mathsf{mod}(x)$ for some $x :^{\mu} A$.

Example

$\flat \boxdot A$ is the discrete coreflection $\flat A$, with $\mathsf{mod} : (x :^{\flat} A) \to \flat A$.

Negative modalities

A modality $\mu : p \to q$ can also map a q-type B to a p-type $\mu \diamond\!\!\to B$, with dual universal property:

$$(x :^{\mu} A) \to B \quad \simeq \quad (y : A) \to \mu \diamond\!\!\to B.$$

- Semantically, a right adjoint $(\mu \boxdot -) \dashv (\mu \diamond\!\!\to -)$.
- Syntactically, have a destructor $\mathsf{unmod} : (x :^{\mu} \mu \diamond\!\!\to B) \to B$ like a Σ-type, with an η-rule.

Example

$\flat \diamond\!\!\to A$ is the codiscrete reflection $\sharp A$, with $\mathsf{unmod} : (x :^{\flat} \sharp B) \to B$.

Dealing with modal contexts

Question

What kind of thing can a modal function be applied to?

E.g. the constructor $\mathsf{mod} : (x :^{\mu} A) \to \mu \boxdot A$ requires a rule

$$\frac{\mathbf{?} \vdash M : A}{\Gamma \vdash \mathsf{mod}(M) : \mu \boxdot A}$$

If $\mu : p \to q$, then Γ is a q-context, but **?** must be a p-context!

We allow variables annotated by general modalities in the context: $(\Gamma, x :^{\mu} A)$ is a q-context if $\mu : p \to q$ and A is a p-type. Then we need to "cancel out" the μ annotation on such a variable to use it.

Context division

The rules for mod, and more general modal function application:

$$\frac{\Gamma/_\mu \vdash M : A}{\Gamma \vdash \mathsf{mod}(M) : \mu \boxdot A} \qquad \frac{\Gamma \vdash M : (x :^\mu A) \to B \qquad \Gamma/_\mu \vdash N : A}{\Gamma \vdash M\, N : B}$$

where $\Gamma/_\mu$ (also written $\Gamma.🔒_\mu$ or $\Gamma.\{\mu\}$ or $\mu\backslash\Gamma$) is a context division or context lock, "only allowing access to μ-variables".

More precisely, $\Gamma/_\mu$ allows access to a variable $x :^\varrho A$ if we can specify a law (2-cell) $\alpha : \varrho \Rightarrow \mu$.

Multiple divisions accumulate: $\Gamma/_\mu/_\nu$ requires $\varrho \Rightarrow \mu \circ \nu$, etc.

This works syntactically, but what does $\Gamma/_\mu$ mean semantically?

Gratzer, Kavvos, Nuyts, Birkedal, "Multimodal Dependent Type Theory", 2011.15021

Division is an adjoint

Recall the introduction rule of $\mu\boxdot A$:

$$\frac{\Gamma/_{\mu} \vdash a : A}{\Gamma \vdash \mathsf{mod}(a) : \mu\boxdot A}$$

This suggests that $(-/_{\mu})$ is a left adjoint to $\mu\boxdot -$.

Theorem (∼GKNB)

MTT with mode theory $\mathcal{M}$ can be interpreted in any 2-functor $\mathscr{C} : \mathcal{M} \to \mathit{CwF}$ such that

- *Each category $\mathscr{C}_p$ models MLTT, and*
- *Each map $\mathscr{C}_\mu : \mathscr{C}_p \to \mathscr{C}_q$ is a dependent right adjoint.*

Left adjoints to modality functors

Thus, in any chain of adjoint functors, we can model all but the leftmost as modalities in MTT. Sometimes we can do even better:

Example

In a cohesive topos with $\int \dashv \flat \dashv \sharp$, we can model $\flat$ and $\sharp$ as MTT modalities. And since $\int$ is an idempotent monadic modality, we can axiomatize it internally.

But this doesn't always work:

Example

The category of condensed*/pyknotic sets has $\flat \dashv \sharp$ but not $\int$.
It seems we can only model $\sharp$, and $\flat$ is a comonad, so not internal.

Co-dextrification

Given $\mathscr{C} : \mathcal{M} \to \mathcal{C}at$, let an object of $\widehat{\mathscr{C}}_r$ consist of

1. For each $\mu : p \to r$ in $\mathcal{M}$, an object $\Gamma_{/\mu} \in \mathscr{C}_p$.
2. For each $\varrho : p \to q$ and $\alpha : \mu \Rightarrow \nu \circ \varrho$, a map $\Gamma_{/\nu} \to \mathscr{C}_\varrho(\Gamma_{/\mu})$.
3. Coherence axioms.

Theorem (S.)

Let $\mathscr{C} : \mathcal{M} \to \mathcal{C}at$, where each $\mathscr{C}_p$ has, and each $\mathscr{C}_\mu$ preserves, $\mathcal{M}$-sized limits. Then $\widehat{\mathscr{C}} : \mathcal{M} \to \mathcal{C}at$, each $\widehat{\mathscr{C}}_\mu$ has a left adjoint, and the types in $\widehat{\mathscr{C}}_p$ are those of $\mathscr{C}_p$.

Thus, we can interpret MTT in $\widehat{\mathscr{C}}$ to reason about $\mathscr{C}$, with modalities $\mu \boxdot -$ for each $\mu : p \to q$ in $\mathcal{M}$.

Moreover, if some $\mathscr{C}_\mu$ has a right adjoint, so does $\widehat{\mathscr{C}}_\mu$, so we can interpret negative modalities $\mu \diamond\!\!\rightarrow -$ for such μ.

Shulman, "Semantics of multimodal adjoint type theory", 2303.02572

Towards general modal proof assistants

Can we implement general modal type theories?

- Gratzer: MTT satisfies normalization
- SGB: Prototype implementation of locally posetal MTT

Potential issues:

- Substitutions in MTT have no "list of terms" canonical form: generated inductively by terms, divisions, composites, etc.
- When evaluating a variable x^{α} in an NbE environment, we have to substitute the resulting "value" along α.
- Co-dextrification with negatives has freely added adjoints. But such 2-categories can have undecidable equality (DPP).

Gratzer, "Normalization for multimodal type theory", 2106.01414
Stassen, Gratzer, Birkedal, "mitten: a flexible multimodal proof assistant", preprint 2022
Dawson, Paré, Pronk, "Undecidability of the Free Adjoint Construction", ACS 2003