Title
Introspective Theories and Geminal Categories

Author
Ramesh, Sridhar

Publication Date
2023

Introspective Theories and Geminal Categories

by

Sridhar Ramesh

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Logic and the Methodology of Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Thomas Scanlon, Chair
Professor Gabriel Goldberg
Professor Wesley Holliday
Professor Antonio Montalban
Professor Michael Shulman

Summer 2023

Introspective Theories and Geminal Categories

Abstract

Introspective Theories and Geminal Categories

by

Sridhar Ramesh

Doctor of Philosophy in Logic and the Methodology of Science

University of California, Berkeley

Professor Thomas Scanlon, Chair

In provability logic, a key principle is Löb's theorem, stating that if the provability of P provably entails P, then P itself is provable (in modal logic notation, $\Box P \vdash P$ has as a consequence $\vdash P$). This was first discovered in the follow-up work on Gödel's incompleteness theorems, with Gödel's results viewable as following from Löb's theorem. Later, it was also seen that the same formal pattern of Löb's theorem described certain fixed point constructions studied under the name of "guarded recursion".

The aim of these notes is to draw attention to a certain simple class of category-theoretic structures which serve as an abstract environment for deriving Löb's theorem and such fixed point constructions, allowing for a vastly generalized and unified understanding of the scope of applicability of such constructions. These are the structures we call "introspective theories".

This very minimal categorical definition nontrivially entails Löb's theorem and guarded recursion at both the term and type level. We also demonstrate how this abstraction offers a clean unification of the interpretation of the Gödel-Löb incompleteness theorems in traditional logic or via arithmetic universes a la Joyal, along with the interpretation by Birkedal et al of guarded recursion in presheaves over well-founded orders, along with the distinct classical interpretation of the GL modal logic in well-founded transitive Kripke frames.

We also explore free instances of our structure, which turn out to admit a tractable explicit description. The free introspective theory is what we call "the theory of geminal categories", and we explore also some further illuminating relationships between the concepts of introspective theories and geminal categories.

To Amma, who showed me why adding *this* many *that*s is the same as adding *that* many *this*es,

and Appa, who showed me what those $\sin$ and $\cos$ buttons on calculators mean.

Acknowledgments

The timeline on which this dissertation was written is rather unusual. I originally attended graduate school from 2006 through 2013, during which I had the bulk of the ideas which were to become this dissertation. But life took a circuitous path and I did not write them up until much later, now finally filing this dissertation in 2023.

First of all, I would like to thank Dana Scott who had been my advisor during the time I initially attended graduate school. His kindness, patience, and humor as an advisor for all those years were tremendous and I still feel his influence in all my mathematical work. Though I did not finish the dissertation at that time, I hope it can still be considered in some sense a part of the legacy Dana has fostered.

I am also profusely grateful to Michael Shulman, who in 2023 graciously took on the role of reading and advising on my dissertation. His extraordinarily detailed reading and his many helpful suggestions have improved this dissertation and will continue to greatly improve my future expositions of these ideas. I could not have completed this without him.

I also thank Tom Scanlon, for his help in navigating my process of returning to complete my PhD and for serving as the chair of my dissertation committee. Thank you also to the other members of my committee: Gabriel Goldberg, Wesley Holliday, and Antonio Montalban.

I am particularly indebted to my friends Reid Dale, Alex Kruckman, and Sarah Rovner-Frydman for their encouragement and feedback throughout my process of writing this dissertation. Reid introduced me to tools such as Overleaf and Quiver and has given me ceaseless encouragement throughout all my moments of doubt. Alex has listened to me explain the concepts of introspective theories and geminal categories more times over more years than anyone else. And Sarah's interest in the material, along with her suggestion to consider what introspective theories would look like from the perspective of indexed category theory, sparked the discovery of Theorem 2.5 and thus of Definition 2.1, setting off the momentum under which I ultimately finally managed to write the dissertation after so many years of not knowing how to start.

During my time in Berkeley, I also received lots of love, care, company, food, etc, from my family in the Bay Area. Thank you to Chitti (who showed me why the angles of a triangle sum to the angle across a straight line), Chittappa, Maya, Meera, Ramesh Mama, Sapna Mami, Deepa, Sheela, and Daniel.

Thank you also to my fellow grad student friends during my time at Berkeley and afterwards, including in rough chronological order Alex Rennet, Gwyneth Harrison-Shermoen, Justin Bledin, Melissa Fusco, Tamar Lando, Oran Gannot, Michael Wan, and Kevin Lin, in addition to the aforementioned Alex Kruckman and Reid Dale.

The diagrams in this document were made using the Quiver app created by Varkor, which is just such a tool as I spent my original years in Berkeley bemoaning the lack of existence of and fantasizing someone would someday make.

Finally, as noted, the journey to here was a very long one. I began graduate school in the summer of 2006. It is now the summer of 2023. Over those seventeen years, there have been so many friends in my life, and I've received help from so many people. So many that it is unfortunately not feasible for me to list them all here. But suffice it to say, I am grateful to all of you.

Contents

Chapter 0

Introduction

The aim of these notes is to identify and draw attention to a certain surprisingly simple and category-theoretically natural mathematical structure which both serves as an abstract environment for the reasoning used in establishing Gödel's incompleteness theorems and Löb's theorem in their traditional instances (as in [Göd31] and [Löb55]), and furthermore allows these and the further theorems and fixed-point results of the Gödel-Löb modal logic of provability (as in [Boo95]) to be vastly generalized.

Some such Löb-style fixed point phenomena have been explored in the literature before, but our abstraction is of note as a particularly simple and general one. This abstraction for the first time formally unifies three distinct threads of work in the literature, having as special cases the interpretation of the Gödel-Löb incompleteness theorems via the initial arithmetic universe a la Joyal (as discussed in [DO20]), the interpretation of Löb's theorem as a guarded fixed point combinator and associated work on guarded (co)inductive types via step-indexing in contexts such as the topos of trees (as in [Bir+11]), and the classical interpretation of the GL modal logic in well-founded transitive Kripke frames.

Our interest is in a *minimal* categorical structure which naturally reflects the abstract structure of the Gödelian argument. We emphasize that (as opposed to much of the literature on categorical abstractions of guarded recursion), our abstraction does not have Löb's theorem built into it directly as an assumption, but rather allows Löb's theorem to be derived from much more basic presumptions. Our abstraction is indeed so simple that it does not even make such common presumptions as cartesian closure, regularity, or coproducts, all of which turn out not to be needed for the derivation of Löb's theorem. (Indeed, not presuming cartesian closure is vital for allowing our abstraction to cover the initial arithmetic universe!)

The core idea is the identification of those essentially algebraic theories satisfying the property that every model of these theories contains also, as part of its structure, a homomorphism into an internal model of the same theory. We call these "introspective theories". This document is devoted to initiating the study of introspective theories.

We give two category-theoretic formalizations of the concept of an introspective theory (one directly corresponding to the above description (Definition 2.7), the other less so

(Definition 2.2)) and prove them equivalent (Theorem 2.5). We then derive a form of Löb's theorem, in terms of the existence of suitably guarded fixed points, for arbitrary introspective theories (Löb's Theorem for Introspective Theories (Theorem 4.19)). In this demonstration of Löb's theorem, the relationship between Löb's theorem and presheaves is also highlighted, including the applicability of Löb's theorem to non-representable presheaves, which has previously gone unremarked upon.

This derivation of Löb's theorem for introspective theories is our most important key result. The separate demonstrations of how each of the three traditional instances of Löb's theorem noted above correspond to certain constructions of introspective theories comprise other key results.

(Specifically, these three traditional instances are seen as instantiations of our abstract theory like so: An introspective theory corresponding to Joyal's work with the initial arithmetic universe is discussed in The initial arithmetic universe (Section 6.4). An introspective theory corresponding to step-indexing in the topos of trees is discussed in Presheaf example related to step-indexing in guarded recursion (Section 2.4.2.1). Introspective theories corresponding to the classical interpretation of GL modal logic in well-founded transitive Kripke frames are discussed in Presheaf examples related to Kripke frames (Section 2.4.2.2). These last two constructions are themselves unified and generalized much further in Models based on presheaf categories (Section 6.5).)

I believe this is the first formal demonstration of how traditional logical contexts such as the syntactic category of Peano Arithmetic (discussed as an introspective theory at Example based on a traditional logical theory (Section 2.4.1)) support guarded recursion not just at the level of propositions (where this amounts to Löb's theorem in its traditional sense), but also for general terms of arbitrary type, and also for types themselves. Similarly for contexts such as the initial arithmetic universe or the initial topos with natural numbers object (discussed in Self-initializing and super-initializing theories (Section 6.3)).

In addition to such traditional finitary logical theories, we give a similar demonstration of the initial topos with countable products as inducing another model of our formal abstraction (in A self-initializing theory with uncountable and uncomputable flavor (Section 6.3.3)). As this structure contains both uncomputable and uncountable data, yet is constructed in a very similar way to the traditional logical incarnations of the Gödel-Löb phenomenon, this should vividly dispel the oft-repeated canard that the Gödel-Löb phenomenon in logic is fundamentally about or constrained to computability. (As amounts to the same thing, this illustrates that the phenomenon is not constrained to structures internalizable in the initial arithmetic universe).

The concept of an introspective theory is itself essentially algebraic in nature, and thus admits free instances as well, and we give a tractable explicit description of the initial introspective theory in Geminal categories (Chapter 5). This explicit description of the initial introspective theory is another key result of ours. We also observe a remarkable surprising relationship between the initial introspective theory and the theory of introspective theories (Observation 5.21), and some dual co-free constructions of introspective theories (Co-free introspective theories and geminal categories (Section 5.12)).

Though introspective theories are our fundamental objects of interest, along the way, we consider also relaxations of the definition of introspective theories to encompass more general structures (in particular, the relaxation we call "locally introspective theories", defined at Definition 2.8) which, while not supporting the derivation of the Gödel-Löb phenomena, allow us to state other theorems and constructions in their natural generality and note broader connections with other mathematics.

0.1 Reading roadmap

The Preliminaries from Higher categorical terminology conventions (Section 1.1) through Self-indexing and slice categories (Section 1.5) cover conventions and material which are used throughout the entire document, which the reader will certainly want to familiarize themselves with. The remainder of the Preliminaries can be read on an as needed basis.

The first chapters Introspective theories (Chapter 2) and Modal logic (Chapter 3) establish the basic concepts of introspective theories, which all later chapters depend on. However, the later chapters Löb's theorem (Chapter 4), Geminal categories (Chapter 5), and Examples in the wild (Chapter 6) can be read essentially independently of each other, in any order or fashion the reader likes. The only dependence between these is that the concept of geminal categories from Geminal categories (Chapter 5) is invoked in one isolated section of Examples in the wild (Chapter 6), at The initial model as a geminal category (Section 6.3.1).

Chapter 1

Category-theoretic preliminaries

In this chapter, we set out terminology, conventions, concepts, lemmas, etc, which will be useful for us later on. We do not claim originality for any of the material in this Preliminaries chapter, though perhaps our choices for how to present this material may at times be idiosyncratic.

The Preliminaries from Higher categorical terminology conventions (Section 1.1) through Self-indexing and slice categories (Section 1.5) cover conventions and material which are used throughout the entire document, which the reader will certainly want to familiarize themselves with. The remainder of the Preliminaries can be read on an as needed basis.

1.1 Higher categorical terminology conventions

We assume familiarity with categories, functors, natural transformations, limits, presheaves, Set, Cat, all in the ordinary sense. At times, we may also call upon some comfort with concepts such as 2-categories, and abstract Kan extensions and comma objects within these. It will also be very useful to have some familiarity with functorial semantics and internal algebraic structures such as internal categories.

We will take all categories we work with to be locally set-sized (which is to say, we will take Set to be large enough to include the hom-set between any two objects of any category we work with). Generally speaking, we are interested in the categories we work with being overall set-sized as well, except for those particular large categories such as Set, Set^X, Cat, etc. Wherever paying explicit attention to such size issues is important, we will make some explicit note. Otherwise, we do not.

In particular, we do not bother explicitly stating size restrictions on an arbitrary category C before using the Yoneda embedding of C into $\mathrm{Set}^{C^{op}}$. Only in situations where there is some risk that it would not be possible to simply interpret Set as suitably large relative to C do we bother making explicit comment on size issues.

We write $m \circ n$ or just mn for composition of morphisms $n : X \to Y$, $m : Y \to Z$ in a category. We occasionally write $n; m$ to mean $m \circ n$. We write id_X or just id for the identity morphism on an object X. We also use parallel lines without arrowheads to denote identity morphisms (or canonical isomorphisms) in diagrams, like so:

$$A =\!=\!= B$$

When C is a category, we occasionally write $c \in C$ to mean that c is an object of C. We usually write $\mathrm{Hom}_C(a, b)$ to mean the morphisms from a to b in C, but we sometimes write $C(a, b)$ instead, especially when C is a 2-category so that $C(a, b)$ is not merely a set but a 1-category.

We use the term **lexcategory** for a category with finite limits. We use the term **lexfunctor** for a functor preserving finite limits, whose domain and codomain are both lexcategories. By LexCat, we mean the 2-category of lexcategories, lexfunctors, and natural transformations. We will not generally make distinctions between $f(a \times b)$ and $f(a) \times f(b)$, etc, when f is a lexfunctor, but shall instead write with the ordinary fluency for working with limit-preserving functors. (Similarly, we use $\mathrm{FinProdCat}$ for the 2-category of finite product categories (i.e., categories with finite products), functors preserving finite products, and natural transformations.)

We will speak frequently of category-valued presheaves (i.e., contravariant functors into the category of categories) and natural transformations between these. Technically, what we mean by these are not "functors" and "natural transformations" in the traditional sense, but what some call "pseudofunctors" and "pseudonatural transformations", or "2-functors" and "2-natural transformations", as the category of categories should be viewed as a 2-category (by which we mean the non-strict concept some call "bicategory"), lacking a notion of equality between its 1-cells and only having a notion of isomorphism between them instead. That is, wherever one might traditionally ask for an (automatically coherent) system of equalities, this is replaced by a coherent system of isomorphisms. We take the convention that this is what terminology such as "functor" and "natural transformation" already means, in such a context. But we will try our best to construct arguments in such a way as that this is not a bother that needs to be explicitly worried about.

Similarly, we do not worry about distinguishing between terms like "isomorphic" and "equivalent" in statements like "category C is isomorphic/equivalent to category D", always meaning by such a statement an adjoint equivalence. Everything always means the weakest thing it could mean, unless we explicitly say we are dealing with something stricter.

Similarly, if we ever describe diagrams involving functors between categories as commuting, we really mean that these diagrams commute up to natural isomorphism. If we make claims about uniqueness in such a context, we mean the space of choices with the relevant isomorphisms is contractible. And so on. Again, our convention is that this is what such terminology already means, in any categorical context where one has such concepts of isomorphism around, unless we have taken care to say we are working with stricter notions instead (see more on strictness below). Unless we have said we are talking about strict notions, we never distinguish between equivalent categorical structures.

(That all said, nothing we do is higher-dimensional than 2-categorical, so everything could in theory be strictified in some fashion, if so desired.)

1.2 Indexed sets and representability

We will now give a series of related definitions, concerning what are called indexed structures. The notions being described in this section are all old hat, none of them are newly invented by us, but we wish to pin them down with particular names to establish a language for easily talking about the things we wish to talk about in the rest of this document.

As we give these definitions, we will also observe a basic stock of theorems about them. Again, we make no claim to originality with these preliminaries. They simply may be useful to remind the reader of, or to give labels to in order to reference as we use them.

The reader who is already very familiar with these notions and just unfamiliar with our conventions of vocabulary is advised to just skim these preliminaries on initial read and then return as needed when faced with unfamiliar vocabulary. Frankly, the reader who is not very familiar with these notions is also given similar advice. No need to spend all one's time reading proofs and details of lemmas up front. It is probably best to read a bit of the preliminaries to get the lay of the land, then go off and read the actual content and come back as needed. But who knows? To each reader, their own reading style may be best.

The key notion upon which everything else builds is the following:

Definition 1.1 Let T be an arbitrary category. By a T-**indexed set**, we mean a presheaf on T; that is, a contravariant functor from T to Set. By a **function** or **map** or any such thing between T-indexed sets, we mean a natural transformation between the corresponding presheaves.

The category of T-indexed sets and maps between them is thus the presheaf category $\mathrm{Set}^{T^{\mathrm{op}}}$. We may also refer to this as $\mathrm{Psh}(T)$.

We may refer to the data of an indexed set at any object t of the category over which it is indexed as its data **defined over** t, or which is t-**indexed**, or as its t-**aspect**.[1] We can refer to the t-aspect of an indexed set P as the set $P(t)$ or P_t.

Note that data defined over t is automatically transferred to corresponding data defined over s by any morphism from s to t in T, by the action of the presheaf. More explicitly, given morphism $m : s \to t$ in T, we may write $P(m) : P(t) \to P(s)$ for the corresponding function in Set, or P_m.

In contexts where it is clear what presheaf P we have in mind, we may also write m^* for $P(m)$. Also, in contexts where it would cause no confusion to speak in this way, given

[1]Those who prefer to talk in terms of fibered structures rather than indexed structures would call this the "fiber" at t.

some t-indexed datum $d \in P(t)$ and a morphism $m : s \to t$, we use the same name d also to refer to the corresponding s-indexed datum which more explicitly would be called $P(m)(d)$ or m^*d. It will be especially common for us to abuse language in this name-reusing way when t is a terminal object.

In the particular case where t is a terminal object, we may refer to the aspect at t of an indexed set as its **global aspect**. By the Yoneda lemma, this global aspect data $P(1)$ of a presheaf P on category T is the same as the data of a map from the terminal object 1 to P, which is the same as the data of a map from the constantly 1 presheaf to P. This is also the same as the data of the limit of P, thought of a T^{op}-indexed diagram. In this way, even if T does not have a terminal object, we may still speak of the global aspect of T-indexed sets. We sometimes use the notation $\mathrm{Glob}(P)$ to refer to the global aspect of an indexed set P.

Definition 1.2 We say an indexed set is **representable** (or T-representable, when we wish to emphasize which indexing category we are talking about) if the corresponding presheaf is representable.

Remark This notion is also sometimes referred to with the word "small" in the literature on indexed or fibered categories (especially in derived senses like describing an indexed category as "small" or "locally small"). Cf. Definition 7.3.3 in [Jac99], which uses the word "small" in essentially this way. When the indexing category T has finite limits (or even just splittings of idempotents), note that representability is equivalent also to the standard notion "tiny"; see [nLa23].

But it is perhaps a bit misleading to use "small"-derived terminology here, as this notion is not closed under subobjects. Indeed, what might be considered the smallest indexed set, the one which constantly takes the value of the empty set, is never "small" in the sense of being representable!

The analogy of representability to the familiar distinction between "small" sets and non-"small" proper classes is often a fruitful one, motivating this terminology. However, to avoid confusion, we stick with only referring to this notion as "representable" for now.

When we wish to refer to the distinction between set-sized and proper-class-sized, we will say "set-sized", rather than saying "small". We will shy away from using the word "small" altogether.

Convention 1.3 Via the Yoneda embedding (which we denote yoneda), we identify T itself as the full subcategory of representable T-indexed sets within the category of all T-indexed sets. In this way, we may speak, for example, of functions from objects of T to T-indexed sets. That is, when t is an object of T, we will readily write t also to mean the Yoneda embedding of t, when we wish to treat it as a representable T-indexed set; we will usually not explicitly write $\mathrm{yoneda}(t)$. And conversely, given a representable T-indexed set P, we freely write also P to name an object in T representing P, rather than explicitly writing $\mathrm{yoneda}^{-1}(P)$.

Via the Yoneda lemma, we frequently also identify $P(t)$ with $\mathrm{Hom}(t, P)$.

THEOREM 1.4 Note that representable sets, construed as objects of $\mathrm{Psh}(T)$, are closed under any limits which exist in T. In particular, if T is a lexcategory, representable sets are closed under finite limits. (This is essentially the observation that the Yoneda embedding preserves limits.)

DEFINITION 1.5 Note that given an arbitrary functor $f : S \to T$, this induces by composition a functor $f^* : \mathrm{Psh}(T) \to \mathrm{Psh}(S)$.[2]

That is, from a T-indexed set P, we may construct the following S-indexed set f^*P:

$$S^{\mathrm{op}} \xrightarrow{f^{\mathrm{op}}} T^{\mathrm{op}} \xrightarrow{P} \mathrm{Set}$$

THEOREM 1.6 Given $f : S \to T$ and an object s in S and a T-indexed set P, we have that $\mathrm{Hom}(s, f^*P) = \mathrm{Hom}(f(s), P)$, with this correspondence being natural in both s and P.

Proof. Keep in mind that in these Hom-expressions, s and $f(s)$ have implicitly been construed as S-indexed sets via the Yoneda embedding. That is, more explicitly, our claim is $\mathrm{Hom}(\mathrm{yoneda}(s), f^*P) = \mathrm{Hom}(\mathrm{yoneda}(f(s)), P)$. To establish this claim, we can apply the Yoneda lemma to both sides of the equation to reduce it to $(f^*P)(s) = P(f(s))$, which is the definition of f^*.

This completes the proof. (In fancy categorical jargon, we have demonstrated that $\mathrm{yoneda} \circ f : S \to \mathrm{Psh}(T)$ is the relative left adjoint of $f^* : \mathrm{Psh}(T) \to \mathrm{Psh}(S)$, relative to the Yoneda embedding $\mathrm{yoneda} : S \to \mathrm{Psh}(S)$.) ∎

LEMMA 1.7 Given f, s, and P as in Theorem 1.6, we have that every morphism $m : s \to f^*P$ factors through a morphism in the range of f^*. That is, $m = f^*(m') \circ \eta$ for some $m' : f(s) \to P$ and $\eta : s \to f^*(f(s))$.

Proof. This is corollary to Theorem 1.6 by the general yoga of relative adjoints.

Specifically, consider the following naturality diagram for the correspondence in Theorem 1.6, where m' is the morphism in $\mathrm{Hom}(f(s), P)$ corresponding to $m \in \mathrm{Hom}(s, f^*P)$ and η is the morphism in $\mathrm{Hom}(s, f^*f(s))$ corresponding to $\mathrm{id}_{f(s)} \in \mathrm{Hom}(f(s), f(s))$.

[2] I apologize for re-using this f^* notation both for the action of a presheaf on a morphism f, *and* for composition of a presheaf with a functor f, but this re-use of notation seems to be relatively standard. We can think of the second use of this notation as a kind of instance of the first, for the category-valued $\mathrm{Hom}(-, \mathrm{Set})$ presheaf on Cat.

$$\begin{array}{ccc} \mathrm{Hom}(f(s),f(s)) & \xrightarrow{\quad m'\circ - \quad} & \mathrm{Hom}(f(s),P) \\ \Big\| & \begin{array}{ccc} \mathrm{id}_{f(s)} & \longmapsto & m' \\ \downarrow & & \downarrow \\ \eta & \longmapsto & f^*(m')\circ\eta = m \end{array} & \Big\| \\ \mathrm{Hom}(s,f^*f(s)) & \xrightarrow{\quad f^*(m')\circ - \quad} & \mathrm{Hom}(s,f^*P) \end{array}$$

■

Theorem 1.8 Let Σ be the forgetful functor from a slice category T/t to its ambient category T. Then the t-aspect of a T-indexed set P is in correspondence with the global aspect of Σ^*P.

Proof. This is corollary to Theorem 1.6, which tells us $\mathrm{Hom}_{\mathrm{Psh}(T/t)}(1_{T/t}, \Sigma^*P)$ is in correspondence with $\mathrm{Hom}_{\mathrm{Psh}(T)}(\Sigma 1_{T/t}, P)$, where $1_{T/t}$ is the terminal object in T/t. As this terminal object is given by the identity morphism into t, we have that $\Sigma 1_{T/t} = t$. Thus, this equation is telling us that the global aspect of Σ^*P corresponds to the t-aspect of P, as desired. ■

Theorem 1.9 If $f \dashv g$, then $f^* \dashv g^*$. Thus $f^* = \mathrm{Lan}_{g^{\mathrm{op}}}$, while $g^* = \mathrm{Ran}_{f^{\mathrm{op}}}$.

Proof. This is simply the fact that adjunction is preserved by 2-functors, and reversed (in the sense of swapping left and right adjoints) by each of co and op. Thus, adjunction is preserved by $\mathrm{Hom}(-^{\mathrm{op}}, C)$ for any fixed C. In particular, adjunction is preserved by $\mathrm{Set}^{-^{\mathrm{op}}}$ within Cat. ■

Theorem 1.10 If $f : S \to T$ has a right adjoint $g : T \to S$, then $f^* : \mathrm{Psh}(T) \to \mathrm{Psh}(S)$ takes representable sets to representable sets. Specifically, $f^*(t) = g(t)$.

Proof. f^* takes any representable presheaf with representing object t in T to the representable presheaf $\mathrm{Hom}_T(f(-), t) = \mathrm{Hom}_S(-, g(t))$. ■

Theorem 1.11 Any functor of the form f^* preserves finite limits.

Proof. This can be seen in several ways. Perhaps most familiarly, this can be seen from the fact that (co)limits in a a functor category are computed pointwise where the pointwise (co)limits exist, and of course set-sized (co)limits all exist in Set. Secondly, when the domain of f is a set-sized category, it can be seen from the fact that f^* has left and right adjoints (the left and right Kan extensions along f^{op}), so that f^* in fact preserves ALL (co)limits that happen to exist, regardless of size. We can also note that $f^*(P) = \mathrm{Hom}(f(-), P)$, which is manifestly limit preserving (though this argument does not generalize as easily to colimit-preservation). ■

We also define more generally the concept of a function between indexed sets having representable fibers:

DEFINITION 1.12 A function $f : A \to B$ between T-indexed sets has **representable fibers** if the pullback of f along any map into B from a representable set is itself representable (thus, lives within a slice category of T). That is, we say f has representable fibers just in case for every pullback diagram of the following sort within the category of T-indexed sets, if t is representable, then so is s:

$$\begin{array}{ccc} s & \longrightarrow & A \\ \downarrow & \lrcorner & \downarrow{\scriptstyle f} \\ t & \longrightarrow & B \end{array}$$

(Beware that when T does not have finite limits, this definition does not have all the properties which might be expected. For example, we might expect that any morphism between T-representable sets should have T-representable fibers, which would not be true if T itself did not have pullbacks. If T does not have binary products, it will not even be true that a map into the terminal object 1 has representable fibers whenever its domain is representable.)

THEOREM 1.13 If $f : A \to B$ has representable fibers and B is representable, then A is representable too.

Proof. Apply the definition of representable fibers to the trivial case of pulling f back along id_B. ∎

The following two theorems follow from the composition of pullback squares into larger pullback squares (or pullback rectangles, one might say):

THEOREM 1.14 Maps with representable fibers are closed under composition.

Proof.

$$\begin{array}{ccc} s & \longrightarrow & A \\ \downarrow & \lrcorner & \downarrow{\scriptstyle f} \\ t & \longrightarrow & B \\ \downarrow & \lrcorner & \downarrow{\scriptstyle g} \\ u & \longrightarrow & C \end{array}$$

When we presume g to have representable fibers, we find that t is representable. Then when we presume f to have representable fibers, we find that s is representable. The composition of the individual pullback squares yields a pullback rectangle, which allows us to conclude that the composition $g \circ f$ has representable fibers.

The above illustrates the argument for binary composition, by simply composing pullbacks. The argument for n-ary composition for any finite n works inductively in the same way (note that the base 0-ary case works in the same way as well; the pullback of an identity morphism is an identity morphism, and an identity morphism with representable codomain has representable domain). ∎

THEOREM 1.15 Maps with representable fibers are closed under pullback along arbitrary maps.

Proof.

$$\begin{array}{ccccc} s & \longrightarrow & D & \longrightarrow & A \\ \downarrow & \lrcorner & \downarrow{\scriptstyle f'} & \lrcorner & \downarrow{\scriptstyle f} \\ t & \longrightarrow & C & \longrightarrow & B \end{array}$$

Any pullback of f' (along some arbitrary map) is a pullback of f itself (along an extended map with the same domain). Thus, if f has yields representable objects whenever pulled back along a map with representable domain, so does its pullback f'. ∎

THEOREM 1.16 If $L : Q \to T$ is a functor with a right adjoint, on a category Q with pullbacks, and f is a map between T-indexed sets with T-representable fibers, then L^*f has Q-representable fibers.

Proof. Let us say $f : A \to B$, and let an arbitrary map $m : q \to L^*(B)$ be given, where q is an object of Q. We must show that the pullback of L^*f along m also lies within Q. For sake of a name, let us call the domain of this pullback P.

$$\begin{array}{ccc} P & \longrightarrow & L^*A \\ \downarrow & \lrcorner & \downarrow{\scriptstyle L^*f} \\ q & \xrightarrow[m]{} & L^*B \end{array}$$

First, observe via Lemma 1.7 that m factors as $L^*(m') \circ \eta$ for some $m' : L(q) \to B$ and $\eta : q \to L^*L(q)$.

$$q \xrightarrow[\eta]{} L^*L(q) \xrightarrow[L^*m']{} L^*B, \qquad L^*m' \circ \eta = m$$

Thus, the pullback yielding P we are interested in can be decomposed as follows:

$$\begin{array}{ccccc} P & \longrightarrow & L^*(A \times_B L(q)) & \longrightarrow & L^*A \\ \downarrow & \lrcorner & \downarrow & \lrcorner & \downarrow{\scriptstyle L^*f} \\ q & \xrightarrow[\eta]{} & L^*L(q) & \xrightarrow[L^*m']{} & L^*B \end{array} \qquad L^*m' \circ \eta = m$$

The right half of the above diagram is L^* (known to preserve pullbacks by Theorem 1.11) applied to the following pullback diagram in $\mathrm{Psh}(T)$:

$$\begin{array}{ccc} A \times_B L(q) & \longrightarrow & A \\ \big\downarrow & \lrcorner & \big\downarrow f \\ L(q) & \xrightarrow[m']{} & B \end{array}$$

Note that, as f has T-representable fibers and $L(q)$ is an object of T (i.e., T-representable), we find that $A \times_B L(q)$ is also T-representable.

By Theorem 1.10, it follows that $L^*(A \times_B L(q))$ is Q-representable, as is $L^*L(q)$.

Thus, the left half of our above diagram is a pullback of morphisms within Q:

$$\begin{array}{ccc} P & \longrightarrow & L^*(A \times_B L(q)) \\ \big\downarrow & \lrcorner & \big\downarrow \\ q & \xrightarrow[\eta]{} & L^*L(q) \end{array}$$

As Q is closed under pullbacks, it follows that P is Q-representable, completing our proof. ■

Definition 1.17 We can talk about any kind of T-indexed structure or T-indexed maps between such structures, as the appropriate diagram of T-indexed sets and functions. For example, we can talk about T-indexed groups and group homomorphisms between them. When the T-indexed sets involved (the sorts within the structure, including the domains and codomains of all the maps defining the structure) are all representable, we say the entire structure is **representable**, or equivalently, we say it is **internal** to T[3]. By the Yoneda lemma, this amounts to a diagram of objects and morphisms within T itself.

Observe that, as f^* for an arbitrary functor $f : S \to T$ preserves finite limits (by Theorem 1.11), it not only takes T-indexed sets to S-indexed sets but also acts as a functor from T-indexed structures to S-indexed structures more generally, for any notion of structure definable using finite limits. For example, f^* takes T-indexed groups to S-indexed groups, and so on. Furthermore, by Theorem 1.10, if f has a right adjoint, then f^* will take representable structures to representable structures.

[3]This "T-internal gadgets" terminology makes most sense when T is thought of as a kind of structure such that structure-preserving maps from T to S take T-internal gadgets to S-internal gadgets. Thus, if the definition of gadgets invokes maps whose domains are defined using finite limits, we will use this terminology of T-internal gadgets only in contexts where we are taking T as a category with finite limits (for example, when speaking of internal categories). If the definiton of gadgets invokes maps whose domains are defined using finite products, we will use this terminology of T-internal gadgets only in contexts where we are taking T as a category with finite products (for example, when speaking of internal groups). If the definition of gadgets invokes maps whose domains are defined using countably infinite products, then to speak of T-internal gadgets, T must be carrying countably infinite product structure, etc.

Observe also that any T-indexed structure S has a global aspect $\mathrm{Glob}(S)$ which is an ordinary (that is, non-indexed) structure of the same sort (as well as aspects at each object of T).

1.3 Indexed categories

Definition 1.18 In the same vein as all this, by a T-**indexed category**, we mean a category-valued presheaf on T; that is, a contravariant functor from T to Cat, and by an **indexed functor** (or simply **functor**) between T-indexed categories, we mean a natural transformation between such presheaves.[4]. (In keeping with our general convention, note that "functor to Cat" and "natural transformation between functors to Cat" here really refer to pseudofunctors and pseudonatural transformations, respectively, as Cat is a 2-category). We say this indexed category is an **indexed lexcategory** (aka, **has finite limits**) if this presheaf factors through the inclusion of LexCat into Cat; that is, if it takes every object to a lexcategory and every morphism to a lexfunctor. We say an indexed functor between indexed lexcategories **preserves finite limits** if it arises from a natural transformation between the corresponding LexCat-valued presheaves. And in the same way as all this, we can speak of **natural transformations** between functors between indexed categories, or any other familiar categorical structure or property.

One might have thought our definition of T-indexed category-like structures would simply be a special case of our previous definition of T-indexed set-like structures as suitable diagrams within $\mathrm{Psh}(T)$ (that is, as suitable diagrams of Set-valued functors). That is indeed the essence of this definition. However, the fact that we take indexed categories to be given by pseudofunctors into the 2-category Cat, instead of treating Cat as a 1-category, provides a subtle but technically convenient generalization beyond directly demanding mere diagrams of Set-valued functors.

Still, all the same notational conventions apply to indexed categories. E.g., given a T-indexed category C, we write $C(t)$ (or C_t, or even Yoneda-style $\mathrm{Hom}(t, C)$) for the category which is the t-aspect of C at an object t of T, we write $C(m) : C(t) \to C(s)$ (or C_m) for the functor induced by a morphism $m : s \to t$ in T, we may write m^* instead of $C(m)$ in contexts where it is clear that we are referring to the action of C, etc.

We now might like to speak about an indexed category being representable, in the sense that its collection of objects and its collection of morphisms are both representable. This is the essence of the definition we will indeed adopt (at Definition 1.24) but there is one pitfall to be aware of here, related to the just mentioned subtlety. We generally speak

[4]The machinery of indexed categories is equivalent to the machinery of fibered categories, a presentation some prefer, but we refrain from that presentation for now. Many of the features which make fibered categories most useful do not strongly apply to our ultimate interest largely in internal structures, while adding distracting complexity to the exposition. The current choice of presentation seemed the simpler one for our purposes, but the reader who disagrees may translate everything into the language of fibered categories if they prefer.

about categories in such a way as that they do not come with a particular notion of their set of objects, as such. That is, two categories may be equivalent (in the technical sense of "equivalent" within the 2-category Cat) though presented with different ostensible sets of objects. For example, a category presented as comprised of one terminal object, and a category presented as comprised of two isomorphic terminal objects, are equivalent categories; there is no pseudofunctor from the 2-category of categories, functors, and natural isomorphisms to Set which would send the first of these to a one-element set and the second to a two-element set. We are to treat them as the "same" category. So to speak about a category as having a particular set of objects, we must imagine it as carrying more fine-grained equality structure on its objects than we normally do.

Though a category does not have a well-defined set of objects, it *does* have a well-defined set of morphisms between any two given objects. Thus, there is no such difficulty in defining when an indexed category is locally representable.

DEFINITION 1.19 Given a T-indexed category C, an object t of T and any two objects a and b in $C(t)$, we can define a T-indexed set whose aspect at objects r of T is the set $\{(m, n) \mid m \in \mathrm{Hom}_T(r, t), n \in \mathrm{Hom}_{C(r)}(m^*a, m^*b)\}$, with the obvious corresponding action on morphisms of T. If the T-indexed set defined in this way is representable for every object t of T and objects a and b in $C(t)$, then we say C is **locally representable**.

Note that this is the same as saying that $\langle \mathrm{cod}, \mathrm{dom} \rangle : \mathrm{Mor}(C) \to \mathrm{Ob}(C) \times \mathrm{Ob}(C)$ has representable fibers in the sense of Definition 1.12, except for that we do not need to think of $\mathrm{Ob}(C)$ as carrying an equality relation as such.

In particular, we have the following result:

THEOREM 1.20 If $L : Q \to T$ is a functor with a right adjoint, on a category Q with pullbacks, and C is a T-indexed locally representable category, then the composition of C with L to yield a Q-indexed category is also locally representable.

Proof. By the reasoning of Theorem 1.16, as applied to the map $\langle \mathrm{cod}, \mathrm{dom} \rangle : \mathrm{Mor}(C) \to \mathrm{Ob}(C) \times \mathrm{Ob}(C)$, although $\mathrm{Ob}(C)$ need not itself be thought of as carrying an equality relation as such. ■

1.4 Strict categories and internal categories

These bothers around the ill-defined set of objects of a general indexed category shall take us down some technical digressions for a bit, before we return to our big picture ideas. (Please keep in mind, the nuances of this section mostly do not matter for a big picture understanding. The main part of this document where such details might matter is in being rigorous in Geminal categories (Chapter 5). We recommend that on a first read, the reader ignore all discussion of strictification or distinction between strict and non-strict concepts, in order to pick up the big picture ideas. The reader can then pay attention to these details on later more scrupulous re-reads as desired.)

Definition 1.21 Specifically, let us say a **strict category** is a set of objects (including the ability to speak about equality of objects in a potentially finer-grained sense than isomorphism) and a set of morphisms, with the usual operations and satisfying the usual equations. We may also speak of a **strict functor**, meaning a homomorphism of such structure that preserves all of it on-the-nose. Strict categories and the strict functors between them comprise the 1-category $\mathrm{StrictCat}$.

We can speak straightforwardly of natural transformations between strict functors, and with those in mind we can also create a 2-category $\mathrm{StrictCat}_2$ of strict categories, strict functors, and natural transformations.

But when we speak of equality of parallel strict functors, we will always mean equality in the 1-category $\mathrm{StrictCat}$, not merely isomorphism in the 2-category $\mathrm{StrictCat}_2$.

Every strict category [or functor or etc], gives rise to a category [or functor or etc] in whatever ordinary sense one would like to think of these. We may say the strict category [or etc] presents the category [or etc] which results. Beware, non-isomorphic strict categories can both present the same (up to equivalence) category!

Just as every strict category presents a non-strict category, conversely, one would ordinarily say every category is presented by at least one strict category.[5] One might, if one likes, say that the only distinction between categories and strict categories is that we gather categories up into a 2-category and speak of categories up to equivalence in such, while we gather strict categories up into a 1-category and speak of strict categories up to isomorphism in such.

Definition 1.22 We now go further in defining a **strict lexcategory**. Here, we mean more than just a strict category for which finite limits exist. We also mean that, when taking special "basic limits", the relevant limit is not merely defined up to isomorphism, but is given as a particular object (in keeping with the fact that objects can be distinguished more finely-grained than up to isomorphism, within a strict category). A **strict lexfunctor** is accordingly one which preserves these chosen basic limits not merely up to isomorphism, but on-the-nose. Strict lexcategories and the strict lexfunctors between them comprise the 1-category $\mathrm{StrictLexCat}$. Strict lexcategories, the strict lexfunctors between them, and the natural transformations between those comprise the 2-category $\mathrm{StrictLexCat}_2$.

In the same way, we can also speak of a **strict category with finite products**, or any similar such categorical structure.

It is important for us to make this demand of chosen basic limits and their preservation on the nose in order to ensure that $\mathrm{StrictLexCat}$ is the category of models and homomorphisms

[5] In certain non-traditional foundations, this may not be true. For example, in Homotopy Type Theory, a groupoid may come primitively with no particular discrete set of objects. If said groupoid was found in the wild instead of constructed by hand and there is furthermore no presumption of an Axiom of Choice, there may be no way to turn it into a strict category. But for our purposes, this sort of thing does not matter. Even in set-theoretic foundations without the Axiom of Choice, the situation becomes more nuanced for turning functors between arbitrary categories equivalent to given strict categories into strict functors between the given strict categories, but again, that will not concern us for now.

of an essentially algebraic theory (a fact we will make use of in Geminal categories (Chapter 5)).

This business of **basic limits** will require more explanation, another technical subtlety. What I mean by this is like so: Consider for example the concept of a category with a terminal object. And now consider the concept of a category with a pair of terminal objects, a terminal object A and a terminal object B. Ordinarily, we would like to say these are equivalent concepts or equivalent theories. They give rise to equivalent 2-categories (of categories with terminal objects, functors taking terminal objects to terminal objects, and natural transformations between these). However, the concept of a strict category with a single chosen terminal object, and the concept of a strict category with two chosen terminal objects A and B, are not equivalent concepts. We can ask questions in the one case that we cannot in the other; for example, in the latter case, we can distinguish between those models in which A and B are equal objects and those models in which A and B are not equal objects, merely isomorphic. This is reflected also in these giving rise to non-equivalent 1-categories of models (in which the objects are strict categories with the designated terminal objects, and the morphisms are functors preserving designated terminal objects on the nose). So when we seek to strictify the concept of a category with a terminal object, we really must make a choice as to how we choose to designate the terminal object; once or multiply.

This issue was illustrated above for terminal objects, but arises again, perhaps even more perniciously, for categories with finite products or finite limits or the like. Here, we find that the essentially algebraic theory of "A strict category with a chosen terminal object and a binary operation sending any pair of objects to a chosen product" is not precisely the same as the essentially algebraic theory of "A strict category with an n-ary operation on objects assigning chosen n-ary products, for each finite n". Or the essentially algebraic theory of "A strict category with a chosen terminal object and chosen (binary) pullbacks" is not precisely the same as the essentially algebraic theory of "A strict category with a chosen terminal object, chosen binary products, and chosen (binary) equalizers", particularly when we ask for homomorphisms between such structures which preserve their operations on-the-nose.

So in general, when we wish to talk about the appropriate notion of "strict lexcategory" (or "strict category with finite products" or "strict cartesian closed category" or any such thing), we must make some decision as to how exactly to formalize this. We must make some choice of a basic stock of limit operations (or representing object operations more generally) of the desired sort, such that all the other desired limits (or representing objects) can be constructed from these basic operations. Different choices will yield slightly different strict concepts, albeit equivalent for all non-strict purposes.

None of the results in this work are ever particularly sensitive to what choice of basic such operations we take in strictifying a categorical concept. We shall simply suppose some such choice has been made whenever needed, and refer to its operations as our basic limits (or basic representing object operations more generally). The one notable presumption we will make is that there are only finitely many basic limit operations involved in defining a strict lexcategory (or any such finitely axiomatizable thing); beyond that, any choice is fine.

If the reader insists that we commit to a specific choice, let us for harmony with [PV07] say a strict lexcategory is defined by having a chosen terminal object and a chosen (binary) pullback operator.

DEFINITION 1.23 Of course, we can speak of **indexed strict categories** now (or indexed strict lexcategories, indexed strict categories with finite products, etc), straightforwardly via Definition 1.17, as the appropriate diagram of indexed sets and functions between them. And we can speak of such indexed strict categories as being representable, just in case their indexed sets of objects and of morphisms are both representable.

DEFINITION 1.24 We will now say an indexed category is **representable** if it is equivalent to some indexed strict category which is representable. Note that we do not demand that, as part of its structure, any particular such strict category is selected; merely, that it is possible to do so. However, we may use the terminology **internal category**, to mean the selection of a specific representable indexed strict category; similarly, an **internal lexcategory** will mean the selection of a specific representable indexed strict lexcategory (including chosen basic limits), and so on for any such notion.

In this way, the terminology of internality always comes with the presumption of strictness. (In particular, an **internal lexfunctor** between internal lexcategories means an indexed strict lexfunctor between them, preserving basic limits on the nose). As T-internal structures are both strict and representable, they can not only be viewed as living within $\mathrm{Psh}(T)$ but can also be viewed as just suitable diagrams within T.

DEFINITION 1.25 We also say an indexed strict category is **locally representable** if the map $\langle \mathrm{dom}, \mathrm{cod} \rangle$ from its set of morphisms to its set of pairs of objects has representable fibers (in other words, though its set of objects may not be representable, everything that exists between any two particular objects is representable).

We can repeat in this language the observation made at the end of Definition 1.19. Given an indexed category C which is equivalent to some indexed strict category C', we have that C is locally representable just in case C' is locally representable. Note that, although an indexed category may be equivalent to non-isomorphic indexed strict categories, they will all agree on whether they are locally representable.

Note that a representable strict category indexed over a category with finite limits is a fortiori locally representable, as expected, as the collection of morphisms between any particular pair of objects is given by an equalizer between sets already presumed representable in a representable strict category.

In the same way as all the above, we adopt the following convention even for non-indexed categories, which can be thought of as categories indexed by the terminal category 1.

DEFINITION 1.26 We say **representable category** to mean a category which is presented by some (set-sized) strict category. And we say **representable lexcategory** to mean a lexcategory which is presented by some (set-sized) strict lexcategory. In particular, it is for us true by definition that all representable lexcategories admit presentations with chosen basic limits.

Note that all the ordinary constructions of category theory which produce set-sized categories or lexcategories from other such data in fact furthermore produce representable categories or representable lexcategories from analogous inputs. So this convention does not change very much, except it saves us some pedantry in the scope of some claims we make which might otherwise fail in contexts without the Axiom of Choice (where a random category found in the wild, instead of constructed by hand, could conceivably have finite limits without admitting any presentation with chosen basic limits, or some such pathology).

We note without detailed proof some strictification results which will be useful to us later.

THEOREM 1.27 Given any lexcategory T, there is a strict lexcategory T_{strict} which presents T, such that furthermore, for any lexfunctor $f : T \to L$ into a strict lexcategory L, there is a strict lexfunctor $f_{strict} : T_{strict} \to L$ which presents f.

THEOREM 1.28 Any indexed category is presented by some indexed strict category.

THEOREM 1.29 Any indexed lexcategory is presented by some indexed strict lexcategory.

THEOREM 1.30 Any internal category which has finite limits (qua indexed category) can be further equipped as an internal lexcategory (without modifying the internal category structure).

Proof. Let the internal category C, internal to T, be given, and suppose its t-aspect has finite limits for each object t of T. That is, the category whose objects are $\mathrm{Hom}(t, \mathrm{Ob}(t))$ and whose morphisms are $\mathrm{Hom}(t, \mathrm{Mor}(t))$, with suitable composition structure from the diagram internal to T defining C, has finite limits.

Then in particular, for each basis finite limit shape, we can consider the case where t is taken to be the set of diagrams of such shape within C (for example, for binary products, we can consider $t = \mathrm{Ob}(C) \times \mathrm{Ob}(C)$, or for binary equalizers, we can consider t taken to be the kernel pair (that is, pullback along itself) of $\langle \mathrm{cod}, \mathrm{dom} \rangle : \mathrm{Mor}(C) \to \mathrm{Ob}(C) \times \mathrm{Ob}(C)$). There will then be, within the t-aspect of C, a corresponding generic diagram of this shape, which will have some limit within C as C has finite limits. The selection of any particular such limit (that is, a particular value in $\mathrm{Hom}(t, \mathrm{Ob}(C))$ to serve as the apex of the limit cone, and particular further values in $\mathrm{Hom}(t, \mathrm{Mor}(C))$ to serve as the projection morphisms of the limit cone) gives us the morphisms within T which serve as a limit-assigning operation on C for this particular shape of basic limit. After making such a choice for each of the basic limit operations (of which we can presume there are only finitely many), we ultimately have equipped C as an internal lexcategory. ∎

Note that it is NOT true that any indexed strict category which has finite limits (qua indexed category) can furthermore be equipped as an indexed strict lexcategory (without modification to the indexed strict category structure)! The former has reindexing functors which need only preserve finite limits in a non-strict-sense, while the latter's chosen basic limits must be such that all reindexing functors preserve basic limits on-the-nose. So it is rather remarkable that we get this for free once our indexed strict category is furthermore representable.

1.5 Self-indexing and slice categories

DEFINITION 1.31 Note that, from any lexcategory T (or even just a category with pullbacks), we obtain a T-indexed lexcategory by considering the functor $T/-$ which assigns to each object t of T the slice category T/t, and whose action on morphisms is given by pullback. We refer to this as the **self-indexing** of T.

Note in the above that our flexibility in considering an indexed category as a pseudo-functor into Cat, rather than a strict functor into $\mathrm{StrictCat}$, pays off in letting us not worry about how to choose specific pullbacks in a strictly functorial way.

The self-indexing $T/-$ of a lexcategory T is not in general representable, nor even locally representable. Given two globally defined objects A and B of the self-indexed category, their corresponding hom-set $\mathrm{Hom}_{T/-}(A, B)$ amounts to the presheaf $\mathrm{Hom}_T(A \times -, B)$ on T, which is to say, the exponential B^A within $\mathrm{Psh}(T)$. This indexed set is representable just in case an exponential object B^A already exists within T. This extends in the same way to non-globally-defined objects of the self-indexed category (considered as globally defined over some slice category of T instead, a la Theorem 1.8), and so we have the following:

THEOREM 1.32 The self-indexing of a lexcategory T is locally representable just in case T is locally cartesian closed.

Even when we do not have local cartesian closure in full, note that when $A = 1$, the exponential B^A always is given by B itself, so that hom-sets whose domain is 1 are always representable within the self-indexed category, with $\mathrm{Hom}_{T/-}(1, B)$ being the same as B itself.

DEFINITION 1.33 In the same way, we can also speak of an **indexed category with finite products**, and indeed, from any category with finite products T (or even just a category with binary products), we obtain a T-indexed category with finite products by considering the functor $T//-$ which assigns to each object t of T the full subcategory of T/t consisting of projections (slice category objects given by the projection : $t \times s \to t$ for some object s of T), and whose action on morphisms is again given by pullback (the pullback of a projection being another projection in a canonical way). We refer to this as the **simple self-indexing** of T. Note that $T//t$ can also be thought of as the Kleisli category for the $t \times -$ comonad;

that is, the objects of $T//t$ are the same as the objects of T, while a morphism : $s_1 \to s_2$ in $T//t$ is the same as a morphism : $t \times s_1 \to s_2$ in T, with suitable composition structure.

For a category with finite limits (or just pullbacks and binary products), the simple self-indexing can be thought of as a full subcategory of the self-indexing; specifically, the full subcategory whose objects in each aspect are restricted to those of T itself.

By analogous reasoning to before, the simple self-indexing $T//-$ of a category with finite products T is locally representable just in case T is cartesian closed.

Observation 1.34 Given any functor $f : A \to B$ between arbitrary categories A and B, we get an induced functor $f' : A/a \to B/f(a)$ between slice categories for any object a of A. If f is a lexfunctor between lexcategories, then this induced functor f' is also a lexfunctor between lexcategories.

In the same way, there is also an induced functor $f'' : A//a \to B//f(a)$ between slice categories restricted to projections, and if f is a finite product preserving functor between categories with finite products, then so is this f''.

Observation 1.35 For any category T with a terminal object 1, we have an equivalence between T and its slice category $T/1$.

Lemma 1.36 If Y is a category with initial object 0 and X is a (2-)category, then to any functor $f : Y \to X$, we can associate a corresponding functor f' from Y to the slice category $f(0)/X$.

Furthermore, if D and C are parallel functors from Y to X, then a natural transformation from D to C amounts to the same thing as a map $\mathcal{S}$ from $D(0)$ to $C(0)$ along with a natural transformation from D' to $\mathcal{S}^* \circ C'$, where $\mathcal{S}^* : C(0)/X \to D(0)/X$ is the functor between these slice categories given by composition with $\mathcal{S}$.

(Dually, for contravariant functors $f : Y^{\mathrm{op}} \to X$ (such as with indexed structures), acting on a category Y with a terminal object 1, we obtain a corresponding contravarint functor f' from Y to the co-slice category $X/f(1)$. And then the dual further result as well.)

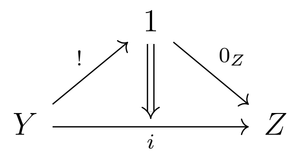
Proof. The first half of the lemma is just the combination of Observation 1.34 and Observation 1.35.

The second half is also straightforward to mechanically verify when X is a 1-category. This lemma should be understood as a triviality. But we will take some care to write out in detail an abstract demonstration that works just as well when X is a 2-category (or indeed, when all categories involved are of whatever higher dimension), such that (in keeping with our linguistic convention) the functors involved are pseudofunctors, the natural transformations are pseudonatural transformations, etc, without having to get our hands dirty manually fussing about higher-dimensional coherence data.

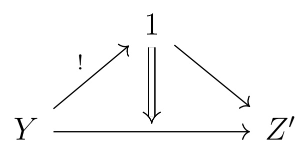
Throughout the remainder of this proof, all references to "category", "functor", etc, are in the sense of whatever dimension of higher-categories encapsulates both Y and X.

Let Z be the category obtained by augmenting Y with a new object 0_Z and unique maps from 0_Z to each object of Y. We have an inclusion functor $i : Y \to Z$, and this inclusion is fully faithful, in the sense that the induced map $\mathrm{Hom}_Y(y_1, y_2) \to \mathrm{Hom}_Z(i(y_1), i(y_2))$ is an equivalence for all $y_1, y_2 \in \mathrm{Ob}(Y)$.

The unique maps from 0_Z to each object in the range of i constitute a diagram of this form:

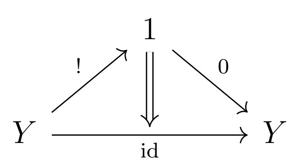


What's more, because of how Z was constructed by freely augmenting Y with a new object and cone from it to the inclusion of Y, this diagram satisfies the universal property that for any other similar diagram



there is a unique functor from Z to Z' commutatively relating the two diagrams. In jargon, this universal property is summarized by saying Z (along with the data of 0_Z and i) is the co-comma of the unique functor from Y to 1 and the identity functor from Y to Y.

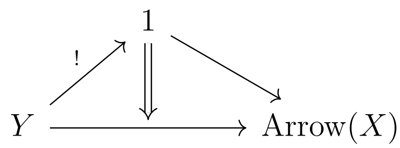
Now, observe that i has a left adjoint, the functor $q : Z \to Y$ such that $q \circ i$ is the identity on Y and such that q of the initiality co-cone for 0_Z in Z is the initiality co-cone for 0 in Y. That is, q is the functor obtained by the co-comma property for Z as applied to this diagram expressing the initiality co-cone of 0 in Y:



It is straightforward to verify that this q is indeed left adjoint to i, as any data in Z is either from the fully faithful inclusion of Y or from the initiality co-cone for 0_Z, and $\mathrm{Hom}_Y(q(i(y_1)), y_2) \simeq \mathrm{Hom}_Y(y_1, y_2) \simeq \mathrm{Hom}_Z(i(y_1), i(y_2))$ naturally in y_1, y_2 from Y, and $\mathrm{Hom}_Y(q(0_Z), y) = \mathrm{Hom}_Y(0, y) \simeq 1 \simeq \mathrm{Hom}_Z(0_Z, i(y))$ naturally in y from Y.

Now consider any two parallel functors $D, C : Y \to X$. Because $q \circ i$ is the identity on Y, we have that $\mathrm{Nat}(D, C) \simeq \mathrm{Nat}(D \circ q \circ i, C)$, where Nat denotes the space of natural transformations between these functors. But because $q \dashv i$, we in turn have that $\mathrm{Nat}(D \circ q \circ i, C) \simeq \mathrm{Nat}(D \circ q, C \circ q)$.

Finally, let us consider what a natural transformation between $D \circ q$ and $C \circ q$ amounts to. This is the same as a functor from Z to the arrow category of X whose domain and codomain projections to X yield $D \circ q$ and $C \circ q$. But by the co-comma property of Z, this functor out of Z corresponds to data of the following form:



such that the rightmost arrow of this diagram corresponds to some arrow $\mathcal{S}$ in X whose domain is $(D \circ q)(0_Z) = D(0)$ and whose codomain is $(C \circ q)(0_Z) = C(0)$, and such that the bottom arrow of this diagram corresponds to a natural transformation from $D \circ q \circ i \simeq D$ to $C \circ q \circ i \simeq C$. The 2-cell in the above diagram then corresponds to the remaining data necessary for us to construe this natural transformation from D to C as simply the codomain projection of a natural transformation between D' and $\mathcal{S}^* \circ C'$, the functors from Y to $D(0)/X$ as mentioned in the statement of this lemma. ■

In order to state the next theorem, some terminology:

Definition 1.37 If T is a lexcategory, then for each object of t, we can construct the free lexcategory extending T with a global element of t. Call this $T[1 \to t]$. Also, for any $f : s \to t$ in T, we can get a map from $T[1 \to t]$ to $T[1 \to s]$ by sending the generic global element of t in $T[1 \to t]$ to the result of applying f to the generic global element of s in $T[1 \to s]$. This action is clearly functorial. Thus, $T[1 \to -]$ comprises a T-indexed object of T/LexCat.

We can replace all references to finite limit structure above with finite product structure. In this case, let us use the name $T[[1 \to -]]$ for the resulting T-indexed object of $T/\mathrm{FiniteProductCat}$.

By Lemma 1.36, we can see $T/-$ as a contravariant functor from a lexcategory T to LexCat/T. And similarly for $T//-$ in terms of finite product structure.

Theorem 1.38 $T[1 \to -]$ is equivalent to $T/-$, when the latter is viewed as a contravariant functor from a lexcategory T to T/LexCat via Lemma 1.36.

(And in just the same way, for a category with finite products T, we have that $T[[1 \to -]]$ is equivalent to $T//-$.)

Proof. This is a standard observation (see 1.10.15 of [Jac99], although this claims it without proof).

It is also simple enough to show, so we will write out the argument:

Applying $T/- : T^{\mathrm{op}} \to \mathrm{LexCat}$ to the unique functor from object t in T to 1, we get a lexfunctor R from $T = T/1$ to T/t given by pullback along the unique map t to 1; more explicitly, $R(x)$ is the projection slice from $t \times x$ to t, which projects out the first component. We also have a left adjoint to this, the forgetful functor $L : T/t \to T$.

Breaking down what our proposed theorem says, the claim we must show is that there is a morphism $g : R(1) \to R(t)$ in T/t such that, for any lexcategory X, lexfunctor $F : T \to X$, and morphism $h : F(1) \to F(t)$ in X, there is a unique lexfunctor $E : T/t \to X$

such that $E \circ R = F$ and $E(g) = h$. (More precisely, there is a contractible space of such lexfunctors such that $E \circ R$ is naturally isomorphic to F, and the induced action of that isomorphism as a map from $\mathrm{Hom}_X(E(R(1)), E(R(t)))$ to $\mathrm{Hom}_X(F(1), F(t))$ takes $E(g)$ to h. By a contractible space, we mean that such a lexfunctor E exists, and for any two such lexfunctors E_1 and E_2, there is a unique natural isomorphism between them which, when whiskered along R and then composed with the given isomorphism from $E_2 \circ R$ to F, yields the given isomorphism from $E_1 \circ R$ to F.) Furthermore, we must show that reindexing within $T[1 \to -]$ corresponds to reindexing within $T/-$ (i.e., to pullback).

We will take our g to be the map from $R(1)$ to $R(t)$ given by the diagonal map $\Delta : t \to t \times t$ (that is, such that $L(g) = \Delta$).

As for the existence aspect of the claim, suppose given an arbitrary lexfunctor $F : T \to X$ and also a morphism $h : 1_X \to F(t)$, where 1_X is any terminal object of X. This F induces a lexfunctor $F' : T/t \to X/F(t)$ via Observation 1.34. Composing this with the pullback action $h^* : X/F(t) \to X/1_X$ and the equivalence of $X/1_X$ with X, we get a lexfunctor $E : T/t \to X$ such that $E \circ R = F$ and $E(g) = h$.

As for uniqueness, it will suffice to show that every diagram in T/t is the pullback along g of some diagram in the range of R (thus, every diagram in T/t has its image under a lexfunctor determined by the lexfunctor's behavior on g and on the range of R). We show this now.

By the observation of Observation 1.34, our R induces also a lexfunctor R' from T/t to $(T/t)/R(t)$.

Observe also that iterated slice categories can be reduced to slice categories. That is for any object x of T/t, we have that the iterated slice category $(T/t)/x$ is equivalent to the slice category $T/L(x)$. In particular, $(T/t)/R(t) = T/L(R(t)) = T/(t \times t)$. Thus, the observation of our last paragraph is that the action of R induces a lexfunctor $R' : T/t \to T/(t \times t)$.

Let $\pi_2 : t \times t \to t$ be the projection of the second component. Note that the pullback action $\pi_2^* : T/t \to T/(t \times t)$ is the same as R'. These both send objects f of T/t to objects $t \times f$ of $T/(t \times t)$, as in the following diagram (and act accordingly on morphisms as well):

$$\begin{array}{ccc} t \times s & \longrightarrow & s \\ \big\downarrow{\scriptstyle t \times f} & \lrcorner & \big\downarrow{\scriptstyle f} \\ t \times t & \xrightarrow[\pi_2]{} & t \end{array}$$

Finally, recall that $L(g) = \Delta : t \to t \times t$ is the diagonal map. Observe that $\pi_2 \circ \Delta = \mathrm{id}_t$. Therefore, the pullback action $(\pi_2 \circ \Delta)^* = \Delta^* \circ \pi_2^* = \Delta^* \circ R' : T/t \to T/t$ is equivalent to the identity.

Thus, every object or morphism in T/t is given (as a diagram in T) by Δ^* applied to some object or morphism in the range of R'. Which is to say, every diagram in T/t is given by g^* applied to some diagram in the range of R, as desired.

Thus, we have the uniqueness to complement existence, and have established that T/t is the free augmentation of T with a global element of t.

Finally, it is easy to verify that the pullback actions from T/t to T/s for arbitrary morphisms $m : s \to t$ correspond to the reindexings from $T[1 \to t]$ to $T[1 \to s]$ along m.

This completes the proof that $T/-$ is the same as $T[1 \to -]$. Note that the same argument, restricted to only those slice category objects which are projections, also shows that when T is a category with finite products, the simple self-indexing $T//-$ is the same as $T[[1 \to -]]$. ■

Observation 1.39 The analogues of Theorem 1.38 automatically also follow for $T[1 \to -]$ for any categorical structure extending the structure of a lexcategory which is automatically transferred to slice categories and preserved by pullback; that is, any structure which automatically transfers from an instance of that structure also to its self-indexing (e.g., for the concepts of locally cartesian closed categories, or for elementary toposes, or for categories with finite and countably infinite limits).

And in just the same way also for $T[[1 \to -]]$ for any categorical structure extending the structure of a category with finite products which automatically transfers from an instance of that structure to its simple self-indexing (e.g., for cartesian closed categories).

1.6 Double or multiple indexing

At this point, for any algebraic-categorical notion S (e.g., the notion of a commutative ring, or the notion of a lexcategory), we also have a definition of the notion of a pair of a category and an instance of notion S indexed over that category.

We can thus iterate this process. In particular, we can speak of a T-indexed (category C and C-indexed set P). We can call this also a (T, C)-indexed set P. Let us observe in more detail what this amounts to.

What this means is that, in addition to having a category T and a T-indexed category C, we also have for every object t in T, some corresponding $C(t)$-indexed set. Thus, we obtain for each t-indexed object c of C a corresponding set we may denote $P(t)(c)$ or $P(t, c)$ or $P_t(c)$ (the t-defined c-defined elements of P). And for each morphism $n : c \to d$ in $C(t)$, we have a reindexing function $P(t, n) : P(t, d) \to P(t, c)$. These reindexings act functorially in that $P(t, n_1 \circ \ldots \circ n_k) = P(t, n_k) \circ \ldots \circ P(t, n_1)$ for any sequence of composable morphisms $n_1, \ldots, n_k$ in $C(t)$.

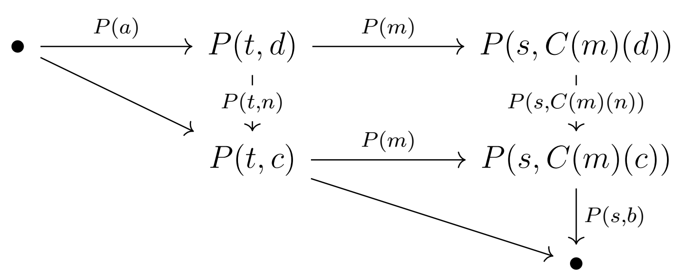
But furthermore, we must have functorial reindexing maps for P along morphisms of T. This means, for any map $m : s \to t$ in T, we must have for every t-defined object c of C a reindexing function $P(m, c) : P(t, c) \to P(s, C(m)(c))$. We may just write $P(m)$ to refer generically to any $P(m, c)$. These reindexings act functorially in that $P(m_1 \circ \ldots \circ m_k) = P(m_k) \circ \ldots \circ P(m_1)$ for any sequence of composable morphisms $m_1, \ldots, m_k$ in T.

Finally, the reindexings along morphisms of T must preserve in a suitable sense the reindexings along morphisms of C. This means the following square of reindexings commutes, for any morphisms $m : s \to t$ in T and $n : c \to d$ in $C(t)$:

$$\begin{array}{ccc} P(t,d) & \xrightarrow{P(m)} & P(s, C(m)(d)) \\ {\scriptstyle P(t,n)}\big\downarrow & & \big\downarrow{\scriptstyle P(s,C(m)(n))} \\ P(t,c) & \xrightarrow[P(m)]{} & P(s, C(m)(c)) \end{array}$$

Using this coherence condition, any reindexing in C followed by a reindexing in T (the left-bottom path) can be turned into an equivalent reindexing in T followed by a reindexing in C (the top-right path). Thus, for any string of reindexings (alternating between reindexings in C and reindexings in T), there is a unique reindexing in T followed by a reindexing in C which it is forced equivalent to by the coherence condition and functoriality.

Thus, we can resummarize all of these conditions like so: We create a category denoted $\int_T C$ (or just $\int C$) whose objects are pairs (t, c) where t is an object in T and c is an object in $C(t)$. A morphism in $\int C$ from (s, c) to (t, d) is given by a pair (m, n) where $m : s \to t$ in T and $n : c \to C(m)(d)$ in $C(s)$. This represents a reindexing along m followed by a reindexing along n, and so by consideration of the previous paragraph, we get also the appropriate composition rule validating our desired coherence condition and automatically ensuring associativity. Specifically, the appropriate composition rule is that $(a, n) \circ (m, b) = ((a \circ m), (C(m)(n) \circ b))$, as can be visualized from our above-noted coherence condition like so:



Then, a (T, C)-indexed set is just the same as as a $(\int_T C)$-indexed set. This also gives us easily the right notion of maps between (T, C)-indexed sets. They are just maps between the corresponding $(\int_T C)$-indexed sets (i.e., natural transformations between presheaves on $\int_T C$). In this way, we can speak about (T, C)-indexed structures in general.

Given a (T, C)-indexed set P, and a globally defined object c of C, it will often be convenient for us to speak of $P(c)$, meaning the T-indexed set which takes t to $P(t, c)$. [More generally, given any t-defined object c of C, we can in the same way consider $P(c)$ as a (T/t)-indexed set.]. This might be seen as conflicting with the natural notation $P(t)$ for objects t of T to denote the t-aspect of P (a presheaf on the t-aspect of C). Generally this will not cause ambiguity, except if there is some existing ambiguity where objects of T and globally defined objects of C have been given the same name (as could happen with terminal objects named 1). If there is ever any such ambiguity, it should be cleared up by context.

We say a (T, C)-indexed structure P is T-representable if, for any t-defined object c of C, the corresponding (T/t)-indexed structure $P(c)$ is (T/t)-representable. When C is presented by a strict category C_{strict}, this is equivalent to saying that the corresponding map into $\mathrm{Ob}(C_{strict})$ has T-representable fibers. Put yet another way, a T-representable (T, C)-indexed set amounts to a T-indexed functor from C^{op} to the self-indexing $T/-$.

Definition 1.40 (Presheaves over Indexed Categories) More generally we have a T-indexed category $\mathrm{Psh}(C)$, acting as the category of T-representable (T, C)-indexed sets.

In more detail, the T-indexed category C gives rise in an obvious way, by reversing arrows, to another T-indexed category C^{op}. We then have that the two T-indexed categories C^{op} and $T/-$ are objects of $\mathrm{Cat}^{T^{\mathrm{op}}}$. But $\mathrm{Cat}^{T^{\mathrm{op}}}$ is a cartesian closed 2-category (in much the same way that $\mathrm{Set}^{T^{\mathrm{op}}}$ is a cartesian closed category, at least when Set or Cat are interpreted expansively enough to include sets or categories of comparable size to T), and thus we can form within it the exponential object given by $T/-$ raised to the power C^{op}. This exponential object is the T-indexed category we call $\mathrm{Psh}(C)$.

The objects of the global aspect of this $\mathrm{Psh}(C)$ correspond to the T-representable (T, C)-indexed sets.

We will only rarely need to consider any of this multi-indexing, and to the extent we do, almost always will really only care about (T, C)-indexed structures P in cases where C is in fact T-representable and P is also T-representable, thus simply given by a suitable diagram in T. However, there is one important instance in which we will need to explicitly consider a doubly-indexed set without any presumed representability properties (in Bootstrapping to Löb's theorem for introspective theories (Section 4.5)).

Further discussion

The construction $\int_T C$ is called the **Grothendieck construction**. By projecting out first coordinates, we get a functor from $\int_T C$ to T; functors which arise in this way are called Grothendieck fibrations, or just fibrations. It turns out, given merely the data of a fibration as a functor between categories, one can recover the indexed category which gave rise to it.

Thus, the data of an indexed category is equivalent to the data of a fibration. The entire machinery of indexed categories can therefore equivalently be presented in terms of fibrations. For this reason, fibrations are also called fibered categories. In particular, one can give a more intrinsic account of the conditions under which an arbitrary functor is a fibration. Furthermore, any natural transformation between T-indexed categories induces a corresponding map between the corresponding fibrations in Cat/T, and again the natural transformation can be recovered from this map, and again a more intrinsic account can be given of which maps arise in this way. Some things are easier to describe in a fibration-based presentation. Other things are more difficult. For our purposes (using the general language of indexed structures ultimately for the goal of understanding specifically representable or internal structures), we felt the indexed category presentation was the most apt. Thus, we will not describe the theory of fibered categories further. We use the Grothendieck construction only for the correspondence between $(\int_T C)$-indexed sets and (T, C)-indexed sets.

Of course, this construction can be iterated further now. A (T, C)-indexed category D is a $(\int_T C)$-indexed category (i.e., a contravariant functor from $\int_T C$ to Cat), and thus gives rise to another category $\int_{\int_T C} D$. Structures indexed over $\int_{\int_T C} D$ can be called (T, C, D)-indexed structures. And so on ad infinitum. But do not worry, we will not need to explicitly consider any further depth of indexing than double-indexing.

Note also that any structure singly-indexed over T can automatically be thought of as doubly-indexed over T and C, where the indexing over C is trivial. This is basically by the fact that $\int_T C$ comes with a projection functor to T, so that each T-indexed structure thus induces, via this functor, a (T, C)-indexed structure. Thus, we can readily speak of maps between T-indexed structures and (T, C)-indexed structures, by treating the former as implicitly (T, C)-indexed themselves.

Indeed, more generally in the multiply indexed context, any structure indexed over some prefix of a string of categories is automatically indexed over the full string of categories. And in the same way, this allows us to speak of maps between structures indexed by different strings of categories. This is perhaps the main reason for us to bring all this up, just so that we can speak of maps between structures at different levels of indexing.

(Keep in mind also that an honest-to-goodness actual structure, living in Set, is like the zero-ary case of indexing; it's indexed by the empty string of categories $()$, but can be seen in a trivial way as T-indexed for any category T).

Note that a map from a T-indexed structure A to a (T, C)-indexed structure B thus amounts to a map from A to $\mathrm{Hom}_C(1, B)$, whenever C has a terminal object. So all this high-faluting multiply indexed stuff just amounts to another way of thinking about maps into global aspects.

1.7 Arithmetic universes, toposes, and other special kinds of category

In addition to categories simpliciter and lexcategories, we have various other augmentations of the basic notion of category which are occasionally of interest. All of the following notions are standard in the literature. We note them here simply for reference (particularly the notion of "arithmetic universe", which is perhaps less well known than the others).

Definition 1.41 A **regular** category is a lexcategory with pullback-stable image factorization. If furthermore every congruence is a kernel pair, we call it **exact** (this is called "effective regular" by some authors).

DEFINITION 1.42 An **extensive** category is a category C with finite coproducts, on which the coproduct operation induces an equivalence between $C/a \times C/b$ and $C/(a \times b)$ for all objects $a, b \in C$.

DEFINITION 1.43 A **natural numbers object** (abbreviated **NNO**) in a category with a terminal object is an object N along with maps $z : 1 \to N$ and $s : N \to N$, which is initial among all objects with such structure, in the sense that for any other object N' with maps $z' : 1 \to N'$ and $s : N' \to N'$, there is a unique map $m : N \to N'$ satisfying $m \circ z = z'$ and $m \circ s = s' \circ m$.

More generally, a category with finite products is said to have "list objects" if for every object X, there is an object L with maps $z : 1 \to L$ and $s : X \times L \to L$, which is initial among all objects with such structure, in the sense that for any other object L' with maps $z' : 1 \to L'$ and $s' : X \times L' \to L'$, there is a unique map $m : L \to L'$ satisfying $m \circ z = z'$ and such that $m \circ s = s' \circ (\mathrm{id}_X \times m)$.

DEFINITION 1.44 An **arithmetic universe** is a category which is exact and extensive with pullback-stable list objects (see [Mai10]). An **arithmetic functor** is one which preserves all this structure.

We denote the initial arithmetic universe as IAU.

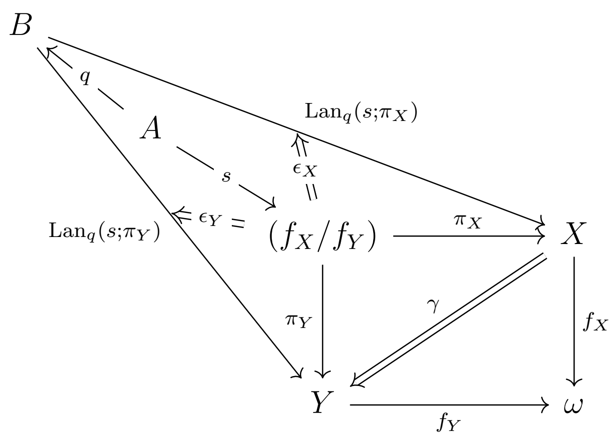
OBSERVATION 1.45 Structures internal to IAU can be thought of as codes for computably enumerable structures. Any computer program describing some computably enumerable structure induces a structure of that sort internal to IAU. (That said, be cautioned that non-isomorphic structures in IAU can become isomorphic after applying the unique arithmetic functor to Set, non-equal morphisms in IAU can become equal in Set, etc! This is like how the same function can be computed by different computer programs, sometimes so different that there is no proof (in whatever fixed proof system) that they compute the same function.)

There are many alternative characterizations of IAU. For example, [Mai10] observes that IAU is also the initial exact and extensive category with a pullback-stable NNO (thus, without presuming list objects in general). I believe the argument given in that paper actually shows furthermore that IAU is also the initial exact category with a pullback-stable NNO (thus, without presuming coproducts).

DEFINITION 1.46 A **topos** is a cartesian closed lexcategory with a subobject classifier. (This is sometimes called an "elementary topos" in the literature, to distinguish from the notion of a "Grothendieck topos"). If a topos has an NNO, we call it an an **NNO-topos**.

1.8 Comma objects and their interaction with Kan extensions

THEOREM 1.47 (THE COMMA-KAN LEMMA) Suppose, within some 2-category, we have the following comma object and left Kan extensions:



Furthermore, suppose f_X preserves the Kan extension $\mathrm{Lan}_q(s;\pi_X)$. (We notably do NOT make any such assumption on f_Y).

Then $\mathrm{Lan}_q\, s : B \to (f_X/f_Y)$ exists and furthermore is preserved by both π_X and π_Y.

(Dually, we can turn all 2-cells around in this theorem, replacing the left Kan extensions with right Kan extensions and moving the preservation condition so that the functor on the codomain side of our comma category must preserve the corresponding Kan extension.)

Proof. We may compute as follows: By the universal property of the comma object (f_X/f_Y), the set of 1-cells from B to this comma object whose projections match our two Kan extensions is given by the set of 2-cells between the top and bottom path from B to ω in the above diagram. Since f_X preserves the top Kan extension, the top path is itself a left Kan extension, and using its universal property, we find that the 2-cells from the top path to the bottom path are the same as 2-cells between two different paths from A to ω; specifically, from $s;\pi_X; f_X$ to $q;\mathrm{Lan}_q(s;\pi_Y); f_Y$. Such a 2-cell is given by the composition of γ and ϵ_Y.

Thus, we do indeed get a 1-cell $m : B \to (f_X/f_Y)$ whose composition with each projection π matches $\mathrm{Lan}_q(s;\pi)$. What remains is only to show that m is indeed $\mathrm{Lan}_q\, s$.

Let an arbitrary $k : B \to (f_X/f_Y)$ be given. By the universal property of the comma category again, we have that 2-cells from m to k are in correspondence with choices of 2-cells from $m;\pi$ to $k;\pi$ for each projection π, such that both resulting composite 2-cells from $m;\pi_X; f_X$ to $k;\pi_y; f_Y$ are equal. But each $m;\pi = \mathrm{Lan}_q(s;\pi)$, so a 2-cell from this to $k;\pi$ amounts to a 2-cell from $s;\pi$ to $q;k;\pi$. A choice of such 2-cells satisfying the coherence condition is, again by the universal property of the comma category (f_X/f_Y), the same thing as a 2-cell from s to $q;k$. Thus, we have shown $\mathrm{Hom}(m,k) = \mathrm{Hom}(s,(q;k))$, which establishes m as satisfying the universal property defining $\mathrm{Lan}_q\, s$. This completes the proof.

(The dualized result of course follows automatically.) ■

Note 1.48 The Comma-Kan Lemma (Theorem 1.47) is perhaps best understood more modularly as the combination of two results concerning Kan extensions: One on the interaction of Kan extensions with products, and another on the interaction of Kan extensions with inserters. Comma objects can then be understood as a particular combination of products and inserters. However, as nearly all the use we will make of this idea is specifically concerning comma categories, we have written the lemma in this combined way, instead of breaking it down into those two steps.

Corollary 1.49 Given a cospan of functors f_X, f_Y from respective categories X and Y, if X and Y both have colimits of a particular shape and f_X preserves colimits of that shape, then the two projections out of (f_X/f_Y) jointly create colimits of that shape.

That is, (f_X/f_Y) has colimits of that shape, both projections out of this comma category preserve colimits of that shape, and these projections jointly reflect colimits of that shape (i.e., a functor into the comma category preserves colimits of that shape whenever its compositions with both projections preserve colimits of that shape).

(Dually, we have the corresponding statements where all instances of “colimit” are turned into “limit” and the first statement’s limit preservation condition is put on f_Y rather than f_X.)

Proof. The existence of such colimits in (f_X/f_Y), along with their preservation by both projections, follows from The Comma-Kan Lemma (Theorem 1.47) within Cat by taking A to be the generic category of the indicated shape and taking B to be the terminal category, considering how colimits correspond to left Kan extensions along functors to the terminal category.

The final claim (the joint reflection of such colimits) then follows from the fact that the forgetful functor from (f_X/f_Y) into $X \times Y$, like any forgetful functor from a comma category to the corresponding product category, is conservative (that is, if the image of a morphism under this functor is invertible, the morphism was already invertible in the comma category). A conservative functor which preserves colimits, on a category which has those colimits, automatically also reflects colimits. ■

Corollary 1.50 Comma objects exist in LexCat, constructed in the same way as in Cat (thus, preserved by the forgetful functor into Cat).

(This corollary is so ubiquitously useful for us that we will not explicitly cite it each time we implicitly invoke it, but rather trust the reader to have absorbed it.)

Proof. From Corollary 1.49, we see that when f_X, f_Y are a co-span of finite limit preserving functors between categories which have finite limits, then the comma category (f_X/f_Y) (the comma object in Cat) is also a lexcategory and its projections are lexfunctors. Thus, this (f_X/f_Y) and its projections exist within LexCat. That these continue to comprise a comma object span within LexCat follows immediately from the fact that the forgetful functor $|-|$ from LexCat to Cat induces bijections between the sets of 2-cells $\mathrm{Hom}(f, g)$ and $\mathrm{Hom}(|f|, |g|)$ for any parallel 1-cells f and g in LexCat (that is, the 2-cells in LexCat between lexfunctors are just ordinary natural transformations, with no further property or structure). ■

The above is paradigmatic of a situation which comes up often, for which it will be convenient to have terminology:

DEFINITION 1.51 Let D be a 2-category and let $Special$ be a sub-2-category of D. We will say a cell of D is special if it lies in $Special$. We presume any invertible cell in D whose domain or codomain is special is itself special (thus, membership in $Special$ is invariant under isomorphism in D), and we also presume that any 2-cell between special 1-cells is special.

Suppose furthermore that D has comma objects. And suppose for any comma object (f_X/f_Y) in D where f_X is special and the domain of f_Y is special, the comma object has special structure jointly created by its two projections. That is, (f_X/f_Y) is special, its two projections are special, and for any 1-cell in D from a special object to (f_X/f_Y), we have that this 1-cell is special whenever both of its compositions with the projections out of (f_X/f_Y) are special.

In this case, we say that the restriction of D to $Special$ is **left comma-stable**, or that $Special$ is left comma-stable within D. (Dually, if this property holds when we instead demand f_Y to be special while allowing f_X to be an arbitrary morphism of D with special domain, then we say $Special$ is **right comma-stable** within D.)

In this language, Corollary 1.49 shows that the restriction of Cat to categories having, and functors preserving, finite colimits is left comma-stable within Cat. And dually, LexCat is right comma-stable within Cat.

We note here a number of left comma-stable sub-2-categories of LexCat, all similarly demonstrable as corollaries of The Comma-Kan Lemma (Theorem 1.47):

COROLLARY 1.52 If X, Y, and ω are lexcategories with finite coproducts, $f_X : X \to \omega$ is a lexfunctor which preserves finite coproducts, and $f_Y : Y \to \omega$ is a lexfunctor (not necessarily preserving coproducts), then the projections out of the comma category (f_X/f_Y) jointly create finite coproducts.

If furthermore, finite coproducts are pullback-stable in X and Y, then finite coproducts are pullback-stable in (f_X/f_Y).

If furthermore finite coproducts are disjoint in X and Y (i.e., these are extensive categories), then this is the case in (f_X/f_Y) as well.

In other words, the restriction of LexCat to lexcategories with finite coproducts, and lexfunctors preserving finite coproducts, is left comma-stable. And the further restriction of LexCat to extensive categories and such functors is also left comma-stable.

Proof. The joint creation of finite coproducts is just a special case of Corollary 1.49. What remains to be shown are the inheritance of exactness properties by the comma objects.

Let Q be the initial lexcategory-with-finite-coproducts generated by morphisms $A_1, \dots, A_n$ into an object A, along with another morphism f into A. To say that n-ary coproducts are pullback stable in a lexcategory C with finite coproducts is to say that every lexfunctor preserving finite coproducts from Q to C sends a particular morphism in Q (the comparison morphism between the coproduct of the pullback and the pullback of the coproduct, for the generic objects $A_1, \ldots, A_n$ of Q/A as pulled back along f into $Q/\operatorname{dom}(f)$) to an isomorphism. Since the projections out of our comma category jointly preserve and reflect finite limit structure, finite coproduct structure, and invertibility, this property will hold in the comma category just in case it holds in both X and Y.

Disjointness is characterized by a similar invertibility condition and thus can be proven to hold in the comma category once it holds in both X and Y in the same way. ■

Corollary 1.53 The restriction of LexCat to regular categories and regular functors is left comma-stable. Furthermore, the restriction of LexCat to effective regular categories and regular functors is left comma-stable.

Corollary 1.54 The restriction of LexCat to lexcategories with pullback-stable natural numbers objects, and lexfunctors preserving NNOs, is left comma-stable.

Similarly, the restriction of LexCat to lexcategories with pullback-stable list objects, and lexfunctors preserving list objects, is left comma-stable.

Putting these all together, we have:

Corollary 1.55 The restriction of LexCat to arithmetic universes and arithmetic functors is left comma-stable.

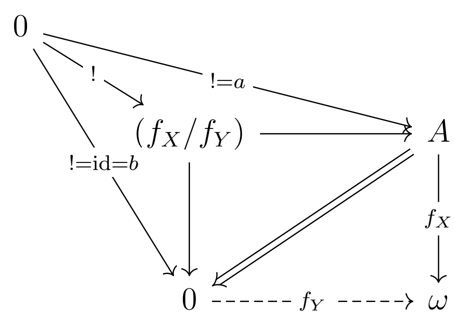
Left comma-stability is of interest to us because of the following abstract theorem, which will be useful to us later in establishing Construction 6.1:

Theorem 1.56 Let D be a 2-category and let $\mathit{Special}$ be a left comma-stable sub-2-category of D. We will say a cell of D is special if it lies in $\mathit{Special}$.

Furthermore, suppose $\mathit{Special}$ has an initial object 0.

Then for any special object ω, we have that the unique special map from 0 to ω is furthermore initial within $D(0, \omega)$. In particular, it follows that the identity on 0 is initial within $D(0, 0)$.

Proof. Let $f_Y : 0 \to \omega$ be an arbitrary map in D. Let $f_X : A \to \omega$ be any special map into ω (two choices always available are the identity from ω to itself, or the unique special map from 0 to ω). Consider the following diagram, illustrating the comma object (f_X/f_Y) of D, as well as unique special maps from 0:



Here, we use solid arrows to denote special 1-cells. The dashed arrow f_Y is a 1-cell of D, not presumed special. The bottom right square is of course the one corresponding to the comma object. Special 1-cells out of 0 are denoted ! to indicate their uniqueness.

Because the comma object is special (by left comma-stability), it has a unique special map from 0. By the universal property of a comma object, along with our presumption that the projections jointly create special structure, this means there is a unique triple of special $a : 0 \to A$, special $b : 0 \to 0$, and 2-cell : $f_X \circ a \to f_Y \circ b$. Because of the initiality of 0, we know that a must be the unique special map from 0 to A, while b must be the identity on 0. Thus, we conclude there is a unique 2-cell from this $f_X \circ a$ to f_Y. As $f_X \circ a$ must be the unique special map from 0 to ω, while f_Y was an arbitrary map in D, we can conclude that the unique special map from 0 to ω is initial within $D(0, \omega)$. ■

1.9 Initial models

Definition 1.57 Given a lexfunctor $F : T \to S$, we may sometimes call this a **model** or **internal model** of T in S. If F is initial within the category $\mathrm{LexCat}(T, S)$, then we naturally call this an **initial model** of T in S. Of course, as with any initial objects, if there is any initial model of T in S, then there is a unique one up to unique isomorphism.

If S_1 has an initial model of T given by $F : T \to S_1$, and $H : S_1 \to S_2$ is a lexfunctor such that $H \circ F$ is an initial model of T in S_2, then we say H preserves initial models of T.

Lemma 1.58 For any lexcategory T, its global sections lexfunctor $\mathrm{Hom}_T(1, -) : T \to \mathrm{Set}$ is an initial model of T in Set.

Proof. By the Yoneda lemma, for any functor $f : T \to \mathrm{Set}$, the natural transformations from $\mathrm{Hom}_T(1, -)$ to f are in correspondence with the elements of $f(1)$. But if f is a lexfunctor, then $f(1)$ has a unique element. ■

Observation 1.59 Let 0 be the initial lexcategory (i.e., the terminal category), and let T and S be arbitrary lexcategories. Then an initial model of T in S is the same thing as a left Kan extension of the unique lexfunctor $: 0 \to S$ along the unique lexfunctor $: 0 \to T$.

Proof. This is immediate upon unraveling definitions. ■

Theorem 1.60 Let T be a fixed lexcategory, and throughout the following, take "initial model" to mean specifically "initial model of T".

If X, Y, and ω are lexcategories with initial models, $f_X : X \to \omega$ is a lexfunctor preserving the initial model, and $f_Y : Y \to \omega$ is an arbitrary lexfunctor, then the comma lexcategory (f_X/f_Y) has an initial model, jointly created by its two projections. (That is, for any lexcategory Z with an initial model, a lexfunctor from Z to (f_X/f_Y) preserves the initial model if and only if both of its compositions with these projections preserve the initial model.)

In other words, the restriction of LexCat to lexcategories having, and lexfunctors preserving, initial models of T is left comma-stable.

Proof. Thanks to Observation 1.59, this is the instance of The Comma-Kan Lemma (Theorem 1.47) within LexCat where A is taken to be the initial lexcategory and B is taken to be T. ■

Definition 1.61 A **finitely axiomatized lex theory** is a lexcategory T which is generated from the initial lexcategory by finitely many operations of freely adjoining a new object, freely adjoining a new morphism between existing objects, or freely making two existing parallel morphisms equal (all these free constructions are meant in the sense of free lexcategories, of course).

Theorem 1.62 Any NNO-topos has an initial internal model of any finitely axiomatized lex theory. Such initial internal models are furthermore preserved by functors preserving NNO-topos structure.

Proof. This is a standard result, shown by carrying out in the internal logic of an NNO-topos the ordinary mathematical construction of an initial model of an essentially algebraic theory. The finite sets of defining sorts, operations, equations, etc, of the finitely axiomatized lex theory are given by finite coproducts of the terminal object. Beyond this, the key ingredient in the construction is the existence in the internal logic of an NNO-topos of suitable sets of well-founded branching trees (so-called W-types).

In particular, we use such trees whose nodes are labelled by operations of the theory, and whose edges out of a node are labelled in correspondence with the arity of this operation, to represent the terms of the free model. Such trees can be represented in an NNO-topos by suitable partial functions from finite sequences of edge labels to node labels (finite sequences in turn can be represented by suitable partial functions on the natural numbers object). By impredicative quantification over powersets as is available in the internal logic of a topos, we may also restrict to the well-founded such trees, and subquotient these to give the well-defined terms of the algebraic theory modulo provable equality. Using similar techniques, we may build up the appropriate homomorphisms and uniqueness proofs to establish this as an initial model of the essentially algebraic theory.

As all of this is standard, we omit the details. ■

There is a stronger generalization of the above available to us, though it takes rather more care to prove:

Theorem 1.63 Any arithmetic universe has an initial internal model of any finitely axiomatized lex theory. Such initial internal models are furthermore preserved by arithmetic functors.

Proof. This standard result is the motivation for most of the interest in arithmetic universes. Details for the special case of finite product theories are given in [Mai05], which notes also that the same argument can be extended to lex theories. This is also implicit in the constructions of [Mai03] and [Mai10], and is noted explicitly in [Vic20]. ■

There is also a variant of Theorem 1.62 of note:

DEFINITION 1.64 By **countable products**, we mean products of collections of objects indexed by the natural numbers. A **finitely axiomatized countably lex theory** is, a la Definition 1.61, a lexcategory with countable products which is generated from the initial lexcategory with countable products by finitely many operations of freely adjoining a new object, freely adjoining a new morphism between two existing objects, or freely making two existing parallel morphisms equal (all these free constructions are meant in the sense of free within the context of lexcategories with countable products and lexfunctors preserving countable products, of course).

THEOREM 1.65 If S is a topos with countable products, and T is a finitely axiomatized countably lex theory, there is an initial lexfunctor preserving countable products from T to S. Such initial internal models are furthermore preserved by functors preserving the structure of a topos with countable products.

Proof. This is by the same techniques as in the proof of Theorem 1.62. The key observations are the following:

We can carry out in the internal logic of any NNO-topos S the standard mathematical construction of the initial model of any internal finitely axiomatized countably lex theory. This will be initial in an internal sense of structures having, and homomorphisms preserving, operations of countable arity; that is, operations of countable arity are interpreted by using the slice category S/N where N is the natural numbers object of S. (Note that this is potentially distinct from the category $S^{\mathbb{N}}$ which is the product of countably many copies of S. Thus, this is not strong enough to guarantee initiality in the "external" sense of having an initial lexfunctor preserving countable products from T to S, as there may be such lexfunctors from T to S which are not "visible" internally.)

Next, we note that a topos with countable products also has countable coproducts (by Paré's theorem on the construction of colimits from similarly shaped limits in a topos), and thus also has a natural numbers object given as the coproduct of countably many terminal objects. For such a topos with countable products S, we actually will have that S/N and $S^{\mathbb{N}}$ coincide. Thus, in such a context, the construction from the above paragraph yields an initial model of T in the same sense as required to establish the noted theorem. ■

We make also the following observations:

OBSERVATION 1.66 The theory of (strict) NNO-toposes and the theory of (strict) arithmetic universes are finitely axiomatized lex theories.

OBSERVATION 1.67 The theory of (strict) toposes with countable products is a finitely axiomatized countably lex theory.

1.10 Quasi-equational theories

The significance of lexcategories with respect to logic is that they capture, in a presentation-independent way (a la Lawvere's functorial semantics), those logical theories given by various sorts, partially defined finitary operations on those sorts (whose domain of definition is given by the constraint that some finitely many other compositions of operations are simultaneously equal), and universal entailments between equations between compositions of such operations (more precisely, entailments from finite conjunctions of equations to equations).

One can often make different choices of primitive sorts, operations, and laws for presenting ultimately the same essentially algebraic theory, in the broad sense of theory equivalence we are most ultimately interested in (having the same composite operations and derived laws, thus the same models, etc). The presentation-freeness of lexcategories means that lexcategories up to categorical equivalence are the same concept as such theories up to this broad sense of theory equivalence.

However, sometimes it is useful to draw finer distinctions between different presentations of such a theory, or at any rate to be able to talk about compact syntactic presentations of these theories more directly.

Many slightly different formalizations of such theory presentations have been given, but one of the cleaner approaches seems to be the notion of **quasi-equational theory** proposed in [PV07]. (The interested reader can find comparison to several other approaches spelled out in this paper.) We shall not here spell out the definition of quasi-equational theories but simply refer the reader to [PV07]. We shall note in this subsection our notational conventions and the useful constructions or theorems we will need for the uses we will make of quasi-equational theories. Essentially all concepts and results in this subsection are taken from [PV07], although rephrased into the language of our own conventions around strict and non-strict categories.

In [PV07], the concept of a **model** of a quasi-equational theory $\mathbb{T}$ within a lexcategory or strict lexcategory S is defined, as well as the notion of **homomorphism** between such models.

These are given by suitable diagrams within S; specifically, a model is given by the selection of an object in S for each sort in $\mathbb{T}$, along with the selection of morphisms in S with appropriate domain and codomain for each operation in $\mathbb{T}$, required to satisfy a corresponding condition on such morphisms for each law of $\mathbb{T}$. A homomorphism between two such models is given by the selection of a morphism in S with appropriate domain and codomain for every sort in $\mathbb{T}$, required to satisfy an appropriate condition for each operation of $\mathbb{T}$. In particular, if $\mathbb{T}$ is finitely specified, then specifying a model or homomorphism only involves specifying finitely much data subject to finitely many conditions. (And similarly if "finite" is replaced by "less than κ" for some infinite cardinal κ.)

The collection of models of $\mathbb{T}$ within lexcategory S, along with homomorphisms between them, comprises a category, which we will call $\mathbb{T}-\mathrm{Mod}(S)$. When S is a strict lexcategory, then in the same way we get a strict category $\mathbb{T}-\mathrm{Mod}(S)$. (Indeed, if strict lexcategory S presents the non-strict lexcategory S', then the strict category $\mathbb{T}-\mathrm{Mod}(S)$ will present the non-strict category $\mathbb{T}-\mathrm{Mod}(S')$. That is, any model of $\mathbb{T}$ in non-strict S' can be presented by some model of $\mathbb{T}$ in strict S, regardless of what choices S makes for how to impose or refrain from imposing equations on its objects.)

This is all suitably functorial. Any lexfunctor $F : S_1 \to S_2$ takes models M of $\mathbb{T}$ in S_1 to models $F(M)$ of $\mathbb{T}$ in S_2, and similarly for actions of F upon homomorphisms between such models. Furthermore, natural transformations between lexfunctors also induce homomorphisms between the corresponding models. In this way, $\mathbb{T}-\mathrm{Mod}(-)$ acts as a 2-endofunctor on LexCat, as well as a 1-endofunctor on $\mathrm{StrictLexCat}$, as well as a 2-endofunctor on $\mathrm{StrictLexCat}_2$.

It is shown in [PV07] how there is a quasi-equational theory (which they call $\mathbb{T}_{\mathrm{cart}}$ but which we shall call $\mathbb{T}_{\mathrm{lex}}$) such that models of $\mathbb{T}_{\mathrm{lex}}$ in S correspond to internal lexcategories in S, and homomorphisms between these correpond to internal lexfunctors.

More generally, they show how to associate to any quasi-equational $\mathbb{T}$ another corresponding quasi-equational theory (which they call $\mathrm{Cart}\,\varpi\mathbb{T}$ but which we shall call $\mathrm{Lex}\varpi\mathbb{T}$) such that a model of $\mathrm{Lex}\varpi\mathbb{T}$ in S corresponds to an S-internal lexcategory L along with a model of $\mathbb{T}$ in the global aspect of L.

It is shown in [PV07] how to associate to any quasi-equational theory $\mathbb{T}$ a corresponding strict lexcategory $\mathcal{C}_{\mathbb{T}}$ called its classifying category. We may also use the name $\mathcal{C}_{\mathbb{T}}$ again for the non-strict lexcategory this presents.

This has the property that there is a natural correspondence between $\mathrm{LexCat}(\mathcal{C}_{\mathbb{T}}, -)$ and $\mathbb{T}-\mathrm{Mod}(-)$. There is also a natural correspondence between $\mathrm{StrictLexCat}_2(\mathcal{C}_{\mathbb{T}}, -)$ and $\mathbb{T}-\mathrm{Mod}(-)$ (this now meaning the strict version of $\mathbb{T}-\mathrm{Mod}(-)$). And finally, this last correspondence respects strict equality as well, in that there is a natural bijection between the sets $\mathrm{StrictLexCat}(\mathcal{C}_{\mathbb{T}}, -)$ and $\mathrm{Ob}(\mathbb{T}-\mathrm{Mod}(-))$.

As noted above, whenever strict lexcategory S presents non-strict lexcategory S', we have that $\mathbb{T}-\mathrm{Mod}(S)$ presents $\mathbb{T}-\mathrm{Mod}(S')$. In conjunction with the correspondences of the last paragraph, this means $\mathrm{StrictLexCat}_2(\mathcal{C}_{\mathbb{T}}, S)$ will present $\mathrm{LexCat}(\mathcal{C}_{\mathbb{T}}, S')$. That is, any functor out of $\mathcal{C}_{\mathbb{T}}$ which preserves limits but not necessarily on-the-nose is naturally isomorphic to some functor which preserves limits on-the-nose.

WARNING 1.68 Note that this last property does not hold when $\mathcal{C}_{\mathbb{T}}$ is replaced by an arbitrary strict lexcategory C! For an arbitrary strict lexcategory C may impose equations on its objects (e.g., it may demand equality of $1 = 1 \times 1$ for the canonical terminal object 1 and not mere isomorphism) which highly constrain the existence of strict lexfunctors out of C, while not so constraining functors whose limit preservation needn't be on the nose.

Finally, we note that [PV07] shows us how to take any strict lexcategory T which presents a lexcategory T' to some quasi-equational theory $\mathrm{Th}(T)$ such that $\mathrm{LexCat}(T', -)$ is in natural correspondence with $\mathrm{Th}(T)-\mathrm{Mod}(-)$.

More generally, given an arbitrary lexcategory T', if we make a choice of strict lexcategory T which presents T', then we may for convenience use the name $\mathrm{Th}(T')$ to refer to $\mathrm{Th}(T)$, even though this strictly speaking depends on the choice of T and not merely on T'. Regardless of the exact quasi-equational theory produced, we will in any case still have the property that $\mathrm{LexCat}(T', -)$ and $\mathrm{Th}(T')\mathrm{-Mod}(-)$ are in correspondence.

Observation 1.69 All of the concepts and results in this subsection generalize completely smoothly to theories in correspondence with categories which not only have finite limits but furthermore have k-ary products, for all k drawn from some fixed set of infinite cardinalities K. (Such categories in fact have all limits of diagrams whose object and morphism cardinalities are in K or finite, by the usual reduction to products and binary equalizers.)

1.11 Localization

The reader is advised that we only ultimately make one use of the results in this section (at Theorem 6.14 calling upon Construction 6.13). Thus, the reader may skip this section unless and until interested in the details of that particular result.

Definition 1.70 For any category C, and any set of morphisms M of C, we may consider freely adjoining inverses to the morphisms in M. This process is called **localization**, and the resulting category is denoted $C[M^{-1}]$. Thus, we have a localization functor $f : C \to C[M^{-1}]$ such that f sends every morphism in M to an isomorphism, and for any $g : C \to D$ which also sends every morphism in M to an isomorphism, there is a unique functor $h : C[M^{-1}] \to D$ such that $h \circ f = g$.

Lemma 1.71 Any localization functor is essentially surjective on objects (eso).

Lemma 1.72 Let C be a category, let M be some set of morphisms of C, and let $f : C \to C[M^{-1}]$ be the corresponding localization. Then given any category D and parallel functors $g_1, g_2 : C[M^{-1}] \to D$, we have that each natural transformation from $g_1 \circ f$ to $g_2 \circ f$ is the whiskering along f of a unique natural transformation from g_1 to g_2. Thus, $\mathrm{Cat}(C[M^{-1}], D)$ comprises a full subcategory of $\mathrm{Cat}(C, D)$.

Proof. Consider the comma category $(\mathrm{id}_D/\mathrm{id}_D)$. A functor $k : C \to (\mathrm{id}_D/\mathrm{id}_D)$ corresponds (via its composition with the two projections out of the comma category) to two functors $h_1, h_2 : C \to D$, along with a natural transformation from h_1 to h_2. As a morphism in a comma category is invertible just in case both of its projections are, we find that such k sends all of M to isomorphisms just in case each of h_1 and h_2 do. In this case, k factors uniquely through f, providing us with a unique corresponding functor $k' : C[M^{-1}] \to (\mathrm{id}_D/\mathrm{id}_D)$ such that $k = k' \circ f$. But such a k' corresponds to any natural transformation between the unique factorizations of h_1 and h_2 through f, whose whiskering along f yields our original natural transformation from h_1 to h_2. ■

DEFINITION 1.73 Let C be a lexcategory. If there is some lexfunctor on C (with any codomain), such that M is the set of morphisms taken to isomorphisms by this lexfunctor, then we say $C[M^{-1}]$ is a **lex localization**.

Lex localizations admit a tractable explicit construction using the "calculus of right fractions", which in particular decomposes each morphism in $C[M^{-1}]$ as the inverse of a morphism from M followed by a morphism from C. The details of this calculus of fractions construction are given in [GZ67], among other references (see also [Bor94] and [KS06]).

This calculus of fractions construction immediately gives us the following result:

LEMMA 1.74 If $f : C \to C[M^{-1}]$ is a lex localization, then for each object $c \in C$, the induced functor from C/c to $C[M^{-1}]/f(c)$ is essentially surjective on objects.

We also have the following result:

LEMMA 1.75 Lex localizations are not just localizations qua category but also qua lexcategory, in that if $f : C \to C[M^{-1}]$ is a lex localization, then $C[M^{-1}]$ is a lexcategory, f is a lexfunctor, and any lexfunctor out of C which factors through f (necessarily uniquely) is such that this factorization is itself a lexfunctor.

(There is of course a precisely dual result for categories with, and functors preserving, finite colimits.)

Proof. This is given by the combination of Propositions 3.1, 3.2, and 3.4 from [GZ67]. ■

LEMMA 1.76 If a lexfunctor is conservative (i.e., any morphism it sends to an isomorphism is already an isomorphism), then it is faithful.

Proof. Let lexfunctor $f : C \to D$ and $m_1, m_2 : c_1 \to c_2$ in C be given such that $f(m_1) = f(m_2)$. The equalizer of $f(m_1)$ and $f(m_2)$ is therefore an isomorphism. But this is the same as f applied to the equalizer of m_1 and m_2, which therefore (by the conservativity of f) must already be an isomorphism. Thus m_1 and m_2 equal. ■

LEMMA 1.77 Let $h : C \to D$ be a lexfunctor, let M be the set of morphisms sent to isomorphisms by h, let $f : C \to C[M^{-1}]$ be the corresponding lex localization, and let $g : C[M^{-1}] \to D$ be the uniquely determined lexfunctor such that $h = g \circ f$.

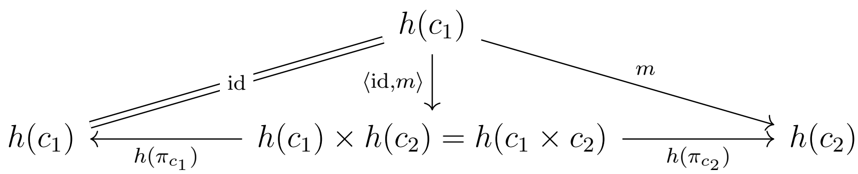
Then g is conservative (and thus faithful, by Lemma 1.76). Furthermore, g is an equivalence of categories whenever, for every object $c \in C$, every object in $D/h(c)$ is isomorphic to h applied to some object in C/c (i.e., the action of h between slice categories is eso).

Proof. First, we show that g is conservative. By the calculus of right fractions construction, each morphism x in $C[M^{-1}]$ is of the form $f(m)^{-1}; f(y)$, where $m \in M$ and thus $f(m)$ is an isomorphism. So if $g(x)$ is an isomorphism, so is $g(f(m)); g(x) = g(f(m); x) = g(f(m); f(m)^{-1}; f(y)) = g(f(y)) = h(y)$. Thus, $y \in M$ as well, and thus $f(y)$ is an isomorphism as well, so that $f(m)^{-1}; f(y) = x$ is an isomorphism as well. This completes the proof that g is conservative (and thus faithful).

Now, suppose h is such that its action between slice categories is eso. We already know that g is faithful, so in order to show that g is an equivalence, we must show that g is eso and full.

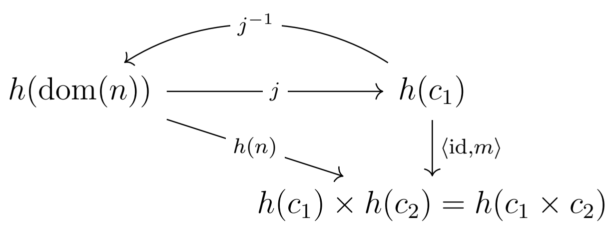
That g is eso follows from the fact that $h = g \circ f$ is eso when acting on $C/1$.

Since f is eso and $h = g \circ f$, to show that g is full, it suffices to show that for every $c_1, c_2 \in C$, each $m \in \mathrm{Hom}_D(h(c_1), h(c_2))$ is given by g applied to some morphism in $\mathrm{Hom}_{C[M^{-1}]}(f(c_1), f(c_2))$. Begin by considering $\langle \mathrm{id}, m\rangle : h(c_1) \to h(c_1) \times h(c_2)$, as in the following commutative diagram:

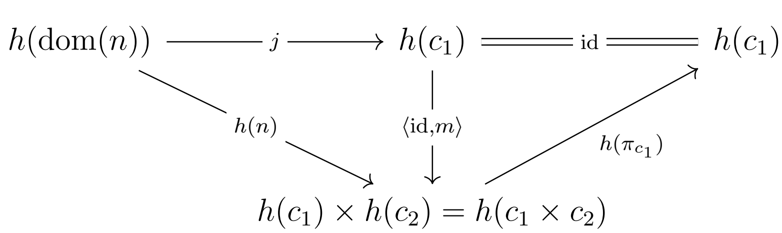


Here, π_{c_1} and π_{c_2} are the projections out of $c_1 \times c_2$, and we keep in mind that as a lexfunctor, h preserves this product cone.

Next, by presumption, this $\langle \mathrm{id}, m\rangle$ is isomorphic, as an object of $D/h(c_1 \times c_2)$, to $h(n)$ for some $n \in C/(c_1 \times c_2)$. Thus, we obtain the following commutative diagram, for some isomorphism j:

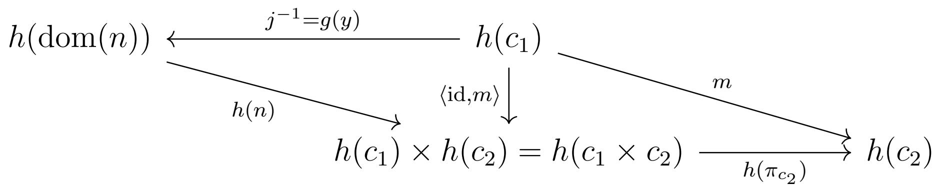


Putting these together, we find that $j = h(n); h(\pi_{c_1}) = h(n; \pi_{c_1})$, as in the following commutative diagram:

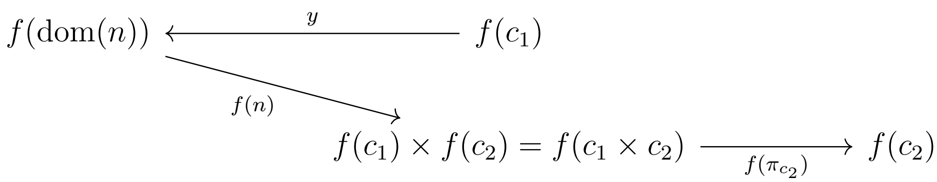


Thus, j is an isomorphism in the range of h. Our setup is such that whenever any morphism of the form $h(x) = g(f(x))$ has an inverse, this is given by applying g to an inverse of $f(x)$. Thus, $j^{-1} = g(y)$ where $y : f(c_1) \to f(\mathrm{dom}(n))$ is the inverse of $f(n; \pi_{c_1})$.

Finally, keeping in mind that $h = g \circ f$, we have by the following commutative diagram that m is given by g applied to a morphism from $f(c_1)$ to $f(c_2)$:



Specifically, m is given by g applied to the following composition:



This completes the demonstration that g is full, and thus completes the proof that g is an equivalence. ■

Lemma 1.78 Let C be an arithmetic universe, let some arithmetic functor from C to another arithmetic universe be given, and let M be the set of morphisms of C which are sent to isomorphisms by said arithmetic functor. Then letting $f : C \to C[M^{-1}]$ be the corresponding localization (an **arithmetic localization**), we have that this is also the localization qua arithmetic universe, in that $C[M^{-1}]$ is an arithmetic universe, f is an arithmetic functor, and any arithmetic functor out of C which factors (necessarily uniquely) through f is such that this factorization is itself an arithmetic functor.

Proof. This is all straightforward by the general techniques of [GZ67]. We omit the details. ■

1.12 Miscellaneous

Here we collect various lemmas and definitions which we call upon at some point later in the document, which did not seem to fit anywhere else.

Observation 1.79 In our metatheory, we have access to the following principle: If M is the term model of some finitely axiomatized lex theory of "gadgets" (for example, the lex theory of NNO-toposes), and Set is also a gadget (but a non-set-sized one), then M not only is initial with respect to set-sized gadgets but also there is a unique homomorphism from M to Set.

Proof. The proof is by the exact same proof that shows M's initiality with respect to set-sized gadgets.

Alternatively, the proof can be carried out like so: Firstly, for existence of a map from M to Set, we take the finitely many finitary operations of our theory and note that the hull definable from these within Set describes a set-sized (indeed, countable) subgadget of Set.

M will have by initiality a homomorphism into this subgadget, and thus into Set itself. As for uniqueness, consider any two homomorphisms from M to Set. Again, their ranges will be set-sized subgadgets of Set, and we can take the union of those ranges, close under the operations of our theory, and find some other set-sized subgadget of Set containing them both. M will have a unique homomorphism into this enveloping subgadget, and thus the parallel homomorphisms of M must have been equal. ■

Be cautioned, however, that the reasoning above only works in our metatheory, with typical principles available to us like the ability to reason about subcollections of Set. Analogous reasoning can fail internally; e.g., every topos T with NNO is such that its self-indexing $T/-$ is an indexed topos with NNO, and such that it has an initial internal topos with NNO T' constructed as a term model, and yet there need be no topos-with-NNO homomorphism from T' to $T/-$. (Indeed, in the initial topos-with-NNO, there will not be such a homomorphism, by Gödel's second incompleteness theorem considerations). Similarly for "arithmetic universe" in place of "topos with NNO".

THEOREM 1.80 The global sections functor $\mathrm{Hom}_{\mathrm{IAU}}(1, -)$ is the unique arithmetic functor from the initial arithmetic universe IAU to Set.

Proof. A unique arithmetic functor ! from IAU to Set is known to exist by the initiality of IAU (keeping in mind Observation 1.79). What remains is only to show that this ! is the same as the global sections functor. By Lemma 1.58, we know that the global sections functor is initial among lexfunctors from IAU to Set. But by Theorem 1.56 with Corollary 1.55, we know that ! is also initial among these. Thus, ! and the global sections functor are isomorphic (indeed, uniquely isomorphic), completing the proof. ■

With this last theorem, we must be careful. As it invoked Observation 1.79, its reasoning does not internalize. In particular, we do NOT have internal to IAU that the global sections functor from its internal initial arithmetic universe IAU' to the self-indexing $\mathrm{IAU}/-$ is arithmetic, or even that this preserves the initial object.[6]

DEFINITION 1.81 Given an endofunctor F on a category, we say a morphism of the form $M : F(m) \to m$ is an F-**algebra**, and dually, a morphism of the form $W : w \to F(w)$ is an F-**coalgebra**.

More generally, we say an F-algebra map from M_1 to M_2 is a commutative square of the following form:

$$\begin{array}{ccc} F(m_1) & \xrightarrow{F(x)} & F(m_2) \\ {\scriptstyle M_1}\downarrow & & \downarrow{\scriptstyle M_2} \\ m_1 & \xrightarrow[x]{} & m_2 \end{array}$$

[6]That this does not preserve the initial object can be seen from the combination of Löb's Theorem for Introspective Theories (Theorem 4.19) and the construction observed in Observation 6.12. Essentially, this would violate Gödel's second incompleteness theorem.

These compose in the obvious way, giving rise to a category of F-algebras, and in the obvious dual fashion, we also obtain a category of F-coalgebras.

Definition 1.82 Given an endofunctor F, we say that a map $y : w \to m$ is an F-**hylomorphism** from a coalgebra $W : w \to F(w)$ to an algebra $M : F(m) \to m$ just in case the following square commutes:

$$\begin{array}{ccc} F(w) & \xrightarrow{F(x)} & F(m) \\ {\scriptstyle W}\uparrow & & \downarrow{\scriptstyle M} \\ w & \xrightarrow[x]{} & m \end{array}$$

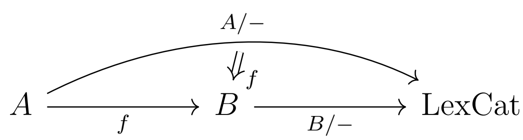
In other words, just in case x is a fixed point of $x \mapsto M \circ F(x) \circ W$.

Lemma 1.83 (Coalgebras As Strict Lexcategory) Let F be an arbitrary (not presumed strict) endolexfunctor on a strict lexcategory C. The category of F-coalgebras is a lexcategory, with its forgetful functor to C creating finite limits (by the analogous reasoning to Corollary 1.49 for inserters rather than comma categories; indeed, this category of coalgebras can be seen as a (non-full) sublexcategory of the comma category (id/F)). As C is in fact a strict lexcategory, we can thus equip the category of F-coalgebras as a strict lexcategory, with its forgetful functor to C creating basic limits (note that it is fine here if F does not preserve basic limits on the nose; all that mattered was that it preserves finite limits in the non-strict-sense).

Theorem 1.84 Let C be any set-sized category, and let P in $\mathrm{Psh}(\mathrm{Psh}(C))$ be given. This P is $\mathrm{Psh}(C)$-representable just in case P turns set-sized colimits in $\mathrm{Psh}(C)$ (i.e., set-sized limits in $\mathrm{Psh}(C)^{\mathrm{op}}$) into set-sized limits in Set.

Proof. The necessity of this condition is clear, from the fact that any representable functor preserves limits. As for its sufficiency, we may consider the map $P(\mathrm{yoneda}(-)) : C^{\mathrm{op}} \to \mathrm{Set}$. This is an object X of $\mathrm{Psh}(C)$, and thus also represents a presheaf $\mathrm{yoneda}(X) = \mathrm{Hom}_{\mathrm{Psh}(C)}(-, X)$ on $\mathrm{Psh}(C)$. By the Yoneda lemma, this $\mathrm{yoneda}(X)$ agrees with P on those objects of $\mathrm{Psh}(C)$ which are C-representable. As all objects in $\mathrm{Psh}(C)$ are set-sized colimits of such objects (by the so-called co-Yoneda lemma), and both P and any representable presheaf turn set-sized colimits into limits, it follows that this $\mathrm{yoneda}(X)$ agrees with P in general. ■

Construction 1.85 Given any pullback-preserving functor $f : A \to B$ between lexcategories, we have that the action of f on slice categories of A also acts as an indexed lexfunctor between the A-indexed lexcategories $A/-$ and $B/f(-)$. We may refer to this indexed lexfunctor also by the same name f, in slight abuse of language. Thus, we obtain a diagram of the following sort:



That is, given any object or morphism of a slice category A/a, we may apply f to it (considered as a morphism or commutative triangle in A) to get an object or morphism of the slice category $B/f(a)$. And because finite limits in slice categories are computed via pullbacks, this action preserves those finite limits.

Chapter 2

Introspective theories

2.1 Preview

In this chapter, we introduce the central object of our interest, the notion of an "introspective theory"[1].

An introspective theory is an essentially algebraic theory such that every model of the theory includes a lexcategory with an internal model of the same theory, as well as a homomorphism from the overall model into the global aspect of the internal model.

We will give two formal definitions of an introspective theory, and prove them equivalent. The second formal definition we give will directly correspond to the previous paragraph. The first formal definition we give will be a bit more compact, but framed in the language of indexed categories.

En route to discussing introspective theories, we also discuss some more general notions we call "pre-introspective theories", "locally introspective theories", and so on, which will be of some use to us as well.

2.2 First definition (indexed style)

Definition 2.1 A **pre-introspective theory** is a lexcategory T, a T-indexed lexcategory C, and a lexfunctor $\mathcal{F}$ from the self-indexing of T to C, like so:

$$T^{\mathrm{op}} \underset{C}{\overset{T/-}{\underset{\mathcal{F}\Downarrow}{\rightrightarrows}}} \mathrm{LexCat}$$

We write out the triple $\langle T, C, \mathcal{F} \rangle$ to refer to a pre-introspective theory when we wish to be fully explicit about its structure. But in typical abuse of language, we also often refer to it simply by the name of its underlying lexcategory T or of the pair $\langle T, C \rangle$, when this

[1] It was alternatively suggested by Alex Kruckman to use for this notion the name "SR-category", with the initials SR standing for... self-reference.

would not cause confusion. We will frequently use the same name $\mathcal{F}$ as though it applies to all introspective theories simultaneously, in the same way that notation like $+$ or $\times$ is overloaded as applying over all rings simultaneously.

Definition 2.2 An **introspective theory** is a pre-introspective theory $\langle T, C\rangle$ in which C is representable[2].

We shall show in later chapters how this simple concept of an introspective theory already suffices to exhibit and capture all the fundamental phenomena of Gödel codes, diagonalization, the Gödel incompleteness theorems, and Löb's theorem. And we shall show that all the traditional instances of Gödel's incompleteness phenomena arise from special cases of this purely algebraic structure. We will also demonstrate functorial fixed point results for this structure, and show some interesting applications of these.

We shall also introduce some further generalizations of this concept, in order to be able to state results along the way in their natural generality or point out connections to related work or interesting structures that are not quite introspective theories per se but are closely related. But throughout these notes, if at any time the abstractions seem daunting or distracting, remember that the concrete concept which matters most is the concept of an introspective theory as defined above.

The example-oriented reader may immediately demand some example of a pre-introspective theory, to orient themselves. Here is the simplest example (or class of examples) of a pre-introspective theory:

Example 2.3 Let T be any lexcategory. Then we have a pre-introspective theory $\langle T, T/-, \mathrm{id}\rangle$. That is, a pre-introspective theory in which C is taken to be the self-indexing itself, with $\mathcal{F}$ as the identity.

Alas, this simple example of a pre-introspective theory is almost never an introspective theory. That is to say, a lexcategory's self-indexing is almost never representable[3].

Here, then, is a simple example of an introspective theory:

Example 2.4 Let T be any lexcategory, and let C be any representable T-indexed lexcategory. Then we have an introspective theory $\langle T, C, \mathcal{F}\rangle$ where each aspect of $\mathcal{F}$ sends all objects to the terminal object.

This is indeed an introspective theory. But alas, although this last example can be as nontrivial as one likes in terms of the structure of T and C, it is of course trivial in all its further structure.

Nontrivial introspective theories do exist and we will give some archetypal examples of them soon enough. But in order to do so, it will be convenient to first develop some further machinery on how (pre-)introspective theories may be presented.

---

[2] We reminder the reader that this means C is presented by an internal category in T.

[3] Indeed, the only case in which this happens is the trivial one where T is the terminal category! We will ultimately establish this result at Theorem 4.23.

2.3 Second definition (non-indexed style)

We shall now make an observation about an alternative but equivalent way to specify the data of a pre-introspective theory.

THEOREM 2.5 Given a lexcategory T and a T-indexed lexcategory C, specifying a pre-introspective theory $\langle T, C, \mathcal{F} \rangle$ (i.e., specifying a T-indexed lexfunctor from the self-indexing $T/-$ to C) is equivalent to specifying a (non-indexed) lexfunctor $\mathcal{S} : T \to \mathrm{Glob}(C)$, along with specifying maps $\mathcal{N}_t$ from each $t \in T$ to $\mathrm{Hom}_C(1, \mathcal{S}(t))$, naturally in t.

That is, keeping in mind that $\mathrm{Hom}_C(1, \mathcal{S}(-)) : T \to \mathrm{Psh}(T)$, and recalling that we also identify T with a full subcategory of $\mathrm{Psh}(T)$ via the Yoneda embedding $\mathrm{yoneda} : T \to \mathrm{Psh}(T)$, the last part of the above is asking for a natural transformation $\mathcal{N} : \mathrm{yoneda} \to \mathrm{Hom}_C(1, \mathcal{S}(-))$.

Proof. Let T be a lexcategory, and let C be some T-indexed lexcategory. By Lemma 1.36 (keeping in mind the contravariance of the functors defining indexed structures), a map from the self-indexing $T/-$ to C as T-indexed lexcategories is the same as a lexfunctor $\mathcal{S}$ from T to the global aspect of C, along with a map from $T/-$ to C as T-indexed objects of T/LexCat (where the map $\mathcal{S}$ is used to treat C as a T-indexed object of T/LexCat).

Next we apply Theorem 1.38. The map from $T/-$ to C as T-indexed objects of T/LexCat is the same as choosing, in a natural way over all t in T, some t-defined value in $\mathrm{Hom}_C(1, \mathcal{S}(t))$. That is, maps from each $t \in \mathrm{Ob}(T)$ to $\mathrm{Hom}_C(1, \mathcal{S}(t))$, comprising a natural transformation. ■

REMARK It wasn't fundamentally important that we were dealing with lexcategories here. The use of Lemma 1.36 as applied to C^{op} only required a terminal object in C. And for the invocation of Theorem 1.38, we only needed that there is some free construction of adjoining global elements. (Even the role terminality plays here is to some degree eliminable, though we have no interest for now in eliminating it). In particular, we get a completely analogous result when lexcategories are replaced throughout by any of the structures noted in Observation 1.39, including for categories with finite products using the simple self-indexing.

As a result of Theorem 2.5, we can give an alternative definition equivalent to Definition 2.1:

DEFINITION 2.6 A **pre-introspective theory** is a lexcategory T, a T-indexed lexcategory C, a lexfunctor $\mathcal{S}$ from T to the global aspect of C, and a natural transformation $\mathcal{N}$ from each $t \in \mathrm{Ob}(T)$ to $\mathrm{Hom}_C(1, \mathcal{S}(t))$. (That is, $\mathcal{N} : \mathrm{yoneda} \to \mathrm{Hom}_C(1, \mathcal{S}(-))$, where yoneda and $\mathrm{Hom}_C(1, \mathcal{S}(-))$ are parallel functors from T to $\mathrm{Psh}(T)$.)

Much as before, we may write out $\langle T, C, \mathcal{S}, \mathcal{N} \rangle$ to be fully explicit, but in typical abuse of language, will refer to a pre-introspective theory by simply naming T or the pair $\langle T, C \rangle$. We will frequently use the same names $\mathcal{S}$ and $\mathcal{N}$ as though they apply simultaneously to all such structures (in the same way that notation like $+$ and $\times$ is overloaded as applicable to separate rings simultaneously).

The definition of an introspective theory remains exactly as before (Definition 2.2) regardless of how one thinks of pre-introspective theories, but for reminder's sake:

DEFINITION 2.7 An **introspective theory** is a pre-introspective theory $\langle T, C\rangle$ in which C is representable.

While it may sometimes be easier to prove theorems about (pre-)introspective theories by using Definition 2.1, it will often be easier to show structures actually are (pre-)introspective theories by using Definition 2.6. But this is not the only benefit of Definition 2.6. The reduction of the full indexed lexfunctor $\mathcal{F}$ to just its global aspect ($\mathcal{S}$) and a natural transformation between 1-functors means much less data around to explicitly fuss about. In particular, when we wish to strictify this into a lex definition eventually at Definition 5.10, we will find the appropriate coherence conditions much easier to manage. It will also be easier to define the appropriate notion of homomorphisms between (pre-)introspective theories by thinking about Definition 2.6.

Definition 2.6 also allows us to quickly appreciate the significance of introspective theories from a functorial semantics point of view. An introspective theory is precisely an essentially algebraic theory (this is the role of T) extending the theory of lexcategories (this is the role of C), such that every model of the theory (which thus has an underlying lexcategory as its interpretation of C) is equipped with a designated homomorphism (this is the role of $\mathcal{N}$) into an internal model of that same theory in its underlying lexcategory (this is the role of $\mathcal{S}$). In short, every model has a homomorphism into a further internal model.

It will be useful for us also to consider sometimes the following concept, intermediate between pre-introspective theories and introspective theories:

DEFINITION 2.8 A **locally introspective theory** is a pre-introspective theory $\langle T, C\rangle$ in which C is locally representable.

Almost all the results we discuss for introspective theories admit straightforward generalization to locally introspective theories. The sole major exception is the derivation of Löb's theorem for introspective theories in Löb's Theorem for Introspective Theories (Theorem 4.19). However, because that one result is so important to us, our main interest in this document is in discussing introspective theories, rather than locally introspective theories more generally.

2.4 Archetypal examples

Let us now finally give the example-oriented reader a nontrivial example of an introspective theory by which to orient themselves. (On the other hand, the reader who prefers to consider abstract definitions without immediately diving into worked out examples of a highly concrete flavor may skip any or all of this section at this introductory time if they find its details a distraction. To each reader at their own taste!).

2.4.1 Example based on a traditional logical theory

In this subsection, we will give two closely related examples. The first example we present is somewhat atypical of general introspective theories, but important nonetheless. It is very similar to the arithmetic universe constructions considered by Joyal in his account of Gödel's incompleteness theorem and by others following up on this (Joyal himself never published this work, but a detailed account has been given in [DO20], building off the formalization of the initial arithmetic universe given in [Mai10]).

Although very similar, the category we use in this first example is not exactly the same as the initial arithmetic universe considered in [DO20] and [Mai10]. The variant and presentation we give is intended to feel natural to an audience of traditional logicians. The connection of this construction to the initial arithmetic universe will be discussed in more detail later at Theorem 6.14.

After having given this first example, we will then tweak it slightly into another introspective theory which provides much better intuition for the general nature of introspective theories.

Construction 2.9 Let us start with the first-order logical theory ZF-Finite: This is the theory ZF but with the axiom of infinity replaced by its negation[4]. The universe this theory describes is the hereditarily finite sets V_ω. Throughout this construction, whenever we speak of formulae, we mean formulae in the language of ZF-Finite, and whenever we speak of provability, we mean provable within ZF-Finite.

Certain formulae are Σ_1. These are the formulae which consist of an initial string of unbounded existential quantifiers (ranging over the entire universe), after which all other quantifiers are bounded (ranging only over the elements of some particular already introduced hereditarily finite set).

Put another way, which may be more comfortable for some readers, the Σ_1 formulas ϕ are precisely those for which there is a computer program P outputting a (possibly empty, possibly finite, possibly infinite) stream of tuples of hereditarily finite sets such that ZF-Finite proves that the tuples which ϕ holds of are precisely the ones output by P. That is, the Σ_1 formulas describe the computably enumerable relations.

(The equivalence between these two accounts of the Σ_1 formulas of ZF-Finite is well known, and we will not go over its details. At any rate, the reader may pick whichever account they like with which to think about the following.)

[4]This theory happens to be bi-interpretable with Peano Arithmetic, but it will be more convenient for us to speak in terms of ZF-Finite so as not to fret about codings of a sort every modern mathematician readily takes for granted in a ZF-style context. Pedantically, we must also make sure to take the Axiom of Foundation in the definition of ZF-Finite to be suitably phrased, e.g. in terms of $\in$-induction, or else we will not have this bi-interpretability.

Now let us define a category whose objects are the Σ_1 formulae with one free variable. Such formulae amount to certain definable subsets of the universe V_ω; that is, they describe classes of hereditarily finite sets. (Note that the classes these formulae describe may themselves be infinite! For example, the tautologically true formula describes the class of all hereditarily finite sets.)

Given two such objects $\phi(n)$ and $\psi(m)$, we take as morphisms between these any Σ_1 formula $F(n,m)$ which provably acts as the graph of a function between the corresponding classes. That is, such that both $\forall n, m.F(n,m) \Rightarrow (\phi(n) \wedge \psi(m))$ and $\forall n.\phi(n) \Rightarrow \exists! m.F(n,m)$ are provable.

Two such morphisms $F(n,m)$ and $G(n,m)$ are considered equal just in case the bi-implication $\forall n, m.F(n,m) \iff G(n,m)$ is provable.

Finally, morphisms compose in the expected way for graphs of functions; that is, the composition of $F(n,p)$ with $G(p,m)$ is given by $(G \circ F)(n,m) = \exists p(F(n,p) \wedge G(p,m))$.

We omit here the straightforward details of verifying that this structure we have just described does indeed satisfy the rules to be a category. Indeed, it is furthermore a regular category (that is, it has finite limits and pullback-stable image factorization; it has finite products because of the definability of ordered pairs in ZF-Finite, and it furthermore has equalizers and image factorization using suitable instances of Separation in ZF-Finite). However, it is not an exact category (that is, not every equivalence relation in this category admits a corresponding quotient). Let Z_{Σ_1} be its ex/reg completion.

(There is not in general any need for the categories involved in an introspective theory to be exact, or even regular. They need only have finite limits. However, for the particular construction we are outlining now, this ex/reg completion is the Z_{Σ_1} we need to look at.)

More explicitly, we can describe Z_{Σ_1} like so:

Its objects are the Σ_1 binary relations $\phi(n,m)$ which can be proven to be partial equivalence relations (i.e., symmetric and transitive), thus corresponding to certain subquotients of the universe of all hereditarily finite sets.

Given any two such formulae $\phi(n_1, n_2)$ and $\psi(m_1, m_2)$, a morphism in Z_{Σ_1} from ϕ to ψ is a Σ_1 formula $F(n,m)$ which provably corresponds to the graph of a function between the corresponding subquotients of the universe. That is, such that the universal closures of all the following are provable:

$F(n,m) \Rightarrow \phi(n,n) \wedge \psi(m,m)$

$\phi(n_1,n_2) \wedge \psi(m_1,m_2) \wedge F(n_1,m_1) \Rightarrow F(n_2,m_2)$

$\phi(n,n) \Rightarrow \exists m[F(n,m)]$

$F(n,m_1) \wedge F(n,m_2) \Rightarrow \psi(m_1,m_2)$.

Two such formulae $F(n,m)$ and $F'(n,m)$ are considered to be equal as morphisms from ϕ to ψ if they are provably equivalent (that is, if both $\forall n, m.F(n,m) \Rightarrow F'(n,m)$ and $\forall n, m.F'(n,m) \Rightarrow F(n,m)$ are provable).

Given morphisms $F : \phi \to \psi$ and $G : \psi \to \chi$ of this sort, we again define their composition in the usual way of composing functions represented as graphs, as $(G \circ F)(n,m) = \exists p[F(n,p) \wedge G(p,m)]$.

This all describes the category Z_{Σ_1}, which one can verify is indeed a category and moreso, an exact category.

Note that our construction of Z_{Σ_1} is such that the objects of Z_{Σ_1}, the morphisms of Z_{Σ_1}, the equality relation on morphisms of Z_{Σ_1}, the composition structure of Z_{Σ_1}, the finite limit structure of Z_{Σ_1}, etc, are all definable within the language of ZF-Finite; indeed, all definable by Σ_1 formulae. (In particular, keep in mind that provability in ZF-Finite is itself a Σ_1 property). Thus, there is a lexcategory Z'_{Σ_1} internal to Z_{Σ_1} which corresponds to this very same construction of Z_{Σ_1} we have just described. And we have a lexfunctor $\mathcal{S}$ from Z_{Σ_1} to the global aspect of Z'_{Σ_1} which sends each piece of the construction of Z_{Σ_1} to the corresponding piece of the construction of Z'_{Σ_1}. This is all straightforward.

As the last bit of introspective theory structure, we must build a natural transformation $\mathcal{N}$ from the identity endofunctor to the endofunctor $\mathrm{Hom}_{Z'_{\Sigma_1}}(1, \mathcal{S}(-))$ on Z_{Σ_1}. The core idea behind this $\mathcal{N}$ is simple. Essentially, to every hereditarily finite set x, we can assign it a code $\ulcorner x \urcorner$, which is an explicit term in the language of ZF-Finite denoting that set. The easy way to do this is to recursively assign to each set $\{a, b, c, \ldots\}$ the term describing a finite set whose members are explicitly enumerated by the terms assigned to $a, b, c, \ldots$. We thus send a set such as $\{\{\}, \{\{\}\}\}$ to the term in the language of ZF-Finite which might be called "$\{\{\}, \{\{\}\}\}$" within quotation marks, and so on.

This gives us a function $\ulcorner - \urcorner$ from hereditarily finite sets to terms in the language of ZF-Finite which describe hereditarily finite sets. This function $\ulcorner - \urcorner$ is definable by a Σ_1 formula and thus gives a morphism in Z_{Σ_1}. This serves as the component of $\mathcal{N}$ at the object of Z_{Σ_1} describing the collection of ALL hereditarily finite sets.

(The categorically oriented reader may think of this recursive definition of $\ulcorner - \urcorner$ as a catamorphism, where the collection of all hereditarily finite sets is understood as the initial algebra for the covariant finite powerset functor.)

All the other objects of Z_{Σ_1} are subquotients of that object (and similarly for the objects of Z'_{Σ_1}), and therefore the components of the natural transformation $\mathcal{N}$ at these other objects can now be obtained uniquely so long as certain factorizations exist. That is to say, the component of $\mathcal{N}$ at any object ϕ of Z_{Σ_1} (that is, an object corresponding to a partial equivalence relation $\phi(n_1, n_2)$) will also be given by the action of $\ulcorner - \urcorner$, but for this to indeed work to map ϕ into $\mathrm{Hom}_{Z'_{\Sigma_1}}(1, \mathcal{S}(\phi))$, we need to know that $\ulcorner - \urcorner$ when acting on individuals which are related by the partial equivalence relation ϕ produces terms which provably describe individuals related by ϕ.

This is where the Σ_1-ness of ϕ plays a vital role. We can prove that, for any Σ_1 property ϕ, for all x, whenever ϕ holds of x, it furthermore provably holds of x (in the sense that the particular term $\ulcorner x \urcorner$, when substituted into the argument of the particular formula defining ϕ, yields a sentence which is derivable in the formal system ZF-Finite).

Finally, let us observe the naturality of this $\mathcal{N}$. Consider the general form of its naturality squares:

$$
\begin{array}{ccc}
\phi & \xrightarrow{\quad m \quad} & \psi \\
{\scriptstyle \ulcorner\_\urcorner}\downarrow & & \downarrow{\scriptstyle \ulcorner\_\urcorner} \\
\mathrm{Hom}_{\mathrm{Z}'_{\Sigma_1}}(1, \mathcal{S}(\phi)) & \xrightarrow[\mathcal{S}(m)\circ -]{} & \mathrm{Hom}_{\mathrm{Z}'_{\Sigma_1}}(1, \mathcal{S}(\psi))
\end{array}
$$

This says that, for any definable function m, it is provably the case that for every x, we have that applying the function m to x and then constructing the term encoding the result ($\ulcorner m(x) \urcorner$) is a provably equivalent term to taking the term representing x and substituting it into the argument of the formula defining m (what might be called $m(\ulcorner x \urcorner)$ or perhaps $\ulcorner m \urcorner(\ulcorner x \urcorner)$ or at any rate $\mathcal{S}(m)(\ulcorner x \urcorner)$). To be clear, by the provable equivalence of terms here, we do not mean syntactic identity as symbol-strings; rather, we mean that there is a provable equality sentence whose left and right sides are comprised of these terms. That is, whatever the actual result of the function m on the input x is, we must have that this is also provably the same as applying m to the input x. Here, again, the Σ_1-ness of the formula defining m comes to our rescue, telling us that truth entails provability in the appropriate way.

Thus, we obtain an introspective theory $\langle \mathrm{Z}_{\Sigma_1}, \mathrm{Z}'_{\Sigma_1}, \mathcal{S}, \mathcal{N} \rangle$. This concludes our first nontrivial example of an instrospective theory!

However, $\langle \mathrm{Z}_{\Sigma_1}, \mathrm{Z}'_{\Sigma_1}, \mathcal{S}, \mathcal{N} \rangle$ is not actually the most typical introspective theory! It has special properties which we should not expect of a general introspective theory. Its internal Z'_{Σ_1} acts as a perfect mirror image of Z_{Σ_1}, and can thus itself be equipped as an internal introspective theory. The internal Z'_{Σ_1} has in some informal sense no further objects (or morphisms, or equations) beyond the range of $\mathcal{S}$. All of this is not typical for an introspective theory.

Construction 2.10 Let us describe now a more archetypal introspective theory, to guide the reader's intuitions better for how general introspective theories act.

Throughout the construction of Z_{Σ_1}, we have imposed a Σ_1 constraint on formulae (both on the formulae defining objects and on the formulae defining morphisms). If we drop all such Σ_1 constraints and allow arbitrary formulae, we get by the same construction an analogous category Z. Z_{Σ_1} sits inside Z as a subcategory (but not a full subcategory! The inclusion from Z_{Σ_1} into Z is faithful, but not full).

Just as the construction of Z_{Σ_1} could itself be carried out in ZF-Finite to get a Z'_{Σ_1} internal to Z_{Σ_1}, so too can the construction of Z can be carried out in ZF-Finite, to get a Z' internal to Z_{Σ_1}. Yes, this Z' is internal to Z_{Σ_1}, not just internal to Z! Even though Z includes as its objects and morphisms formulae which are not Σ_1, the description of Z (as a category whose objects are symbol-strings for which certain other symbol-strings exist, and whose morphisms are symbol-strings for which certain other symbol-strings exist, and so on) is Σ_1.

Finally, the inclusion of Z_{Σ_1} into Z yields, analogously, an inclusion from Z'_{Σ_1} into Z', internal to Z_{Σ_1}. This means the functor $\mathcal{S}$ from Z_{Σ_1} into the global aspect of Z'_{Σ_1} can just as well be thought of as having Z' for its codomain, and similarly the natural transformation $\mathcal{N}$ can just be well as thought of in this context. (This way of making one introspective theory from another is an instance of the general construction Construction 2.17.)

Summarizing, we get an introspective theory $\langle Z_{\Sigma_1}, Z', \mathcal{S}, \mathcal{N} \rangle$, where Z_{Σ_1} is the lexcategory of Σ_1-definable hereditarily finite sets and Σ_1-definable functions between them up to provable equivalence in ZF-Finite, Z' is the lexcategory internal to Z_{Σ_1} of arbitrary definable sets and arbitrary definable functions between them up to provable equivalence in ZF-Finite, $\mathcal{S}$ assigns to each piece of Z_{Σ_1} the corresponding (globally defined) piece of Z', and $\mathcal{N}$ is the Σ_1-definable function which sends any hereditarily finite set to the canonical term describing it, as well as witnessing the provable entailment from truth to provability for Σ_1 formulae.

Phew! What a long walk it was to get to describing that example! All the better, then, that we have formalized introspective theories so abstractly, and can work with them without having to fuss about such concrete details as in that example. But this is indeed the archetypal example it will be best to keep in mind to guide the reader's intuition throughout all further discussion.

WARNING 2.11 While we have above constructed introspective theories $\langle Z_{\Sigma_1}, Z'_{\Sigma_1} \rangle$ and $\langle Z_{\Sigma_1}, Z' \rangle$, the reader should be cautioned that there is no natural introspective theory $\langle Z, Z' \rangle$. As a check of their understanding, the reader is encouraged to think about why this is. Later on, we will characterize the kind of structure which $\langle Z, Z' \rangle$ has, which is that of a "geminal category" as defined in Geminal categories (Chapter 5).

2.4.2 Examples based on presheaf categories

In this subsection, we will give some other instructive examples of introspective theories based on presheaf categories. These examples are of a very different flavor from those based on logical theories as in the previous section, thus helping to illustrate the generality of the notion of introspective theory.

The examples in the first half of this section are based on the topos of trees and the "later" modality, as used in much work on step-indexing and guarded recursion. This may also be useful to build up intuition as we work towards the more complicated final examples in the latter half of this section.

The examples in the latter half of this section are closely related to the use of Kripke frames to interpret the K4 and GL modal logics.

In both examples in this section, we first construct a locally introspective theory using an unrestricted presheaf category. We then impose some cardinality constraints to cut these down into introspective theories.

All the constructions in this section are unified and vastly generalized in Construction 6.16 and Construction 6.21.

2.4.2.1 Presheaf example related to step-indexing in guarded recursion

We present this example in terms of presheaves over the natural numbers (which comprise the so-called topos of trees), but analogous examples may be constructed for presheaves over arbitrary categories; see the generalization at Models based on presheaf categories (Section 6.5). We focus on presheaves over the natural numbers in this introductory example as it is perhaps the simplest nontrivial presheaf category to consider, and also as the "later" modality on the topos of trees which is at the core of this example is much studied in the literature on guarded recursion (e.g., as in [Bir+11]).

Construction 2.12 Let ω be the poset of natural numbers with their usual ordering, and consider the category of presheaves $\mathrm{Psh}(\omega)$ (often called the topos of trees). We will equip this as a locally introspective theory.

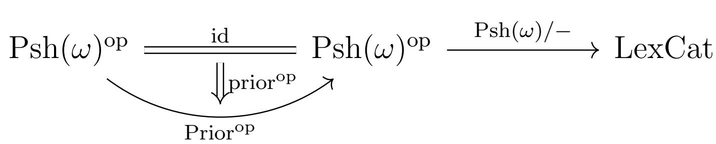
The functor $\mathrm{Successor}$ given by $n : \omega \mapsto n+1 : \omega$ induces correspondingly a functor $\mathrm{Successor}^* : \mathrm{Psh}(\omega) \to \mathrm{Psh}(\omega)$. For convenience, we will use the name Prior rather than $\mathrm{Successor}^*$ to refer to this endofunctor on $\mathrm{Psh}(\omega)$. Thus, $\mathrm{Prior}(P)(n) = P(n+1)$ for $n \in \omega$.

The map $n \leq n+1$ from identity to $\mathrm{Successor}$ as endofunctors on ω induces a corresponding map from Prior to identity as endofunctors on $\mathrm{Psh}(\omega)$ (keeping in mind the contravariance of presheaves). We shall write $\mathrm{prior} : \mathrm{Prior} \to \mathrm{id}$ for this map.

Also, as with any functor between presheaf categories given by composition in this manner, Prior has a right adjoint, given by right Kan extension. [The right adjoint of Prior may be called Later, or is often called $\blacktriangleright$ in guarded recursion literature. It can be described by $\mathrm{Later}(P)(0) = 1$ and $\mathrm{Later}(P)(n+1) = P(n)$ for $n \in \omega$, with the obvious corresponding actions on restriction maps and on morphisms between presheaves. Note that we may pull the map $\mathrm{prior} : \mathrm{Prior} \to \mathrm{id}$ through the adjunction $\mathrm{Prior} \dashv \mathrm{Later}$ to obtain a corresponding map $\mathrm{next} : \mathrm{id} \to \mathrm{Later}$.]

Let C be the $\mathrm{Psh}(\omega)$-indexed lexcategory given by $C(-) = \mathrm{Psh}(\omega)/\mathrm{Prior}(-)$. That is, C is given by applying Prior^* to the self-indexing $\mathrm{Psh}(\omega)/-$. Note that C is locally representable by Theorem 1.20, as Prior has a right adjoint and the self-indexing $\mathrm{Psh}(\omega)/-$ is locally representable (because $\mathrm{Psh}(\omega)$ is locally cartesian closed, as it is a presheaf topos).

What remains to equip $\langle \mathrm{Psh}(\omega), C\rangle$ as a locally introspective theory is to choose a suitable $\mathcal{F}$ from $\mathrm{Psh}(\omega)/-$ to C. We do this via whiskering prior as in the following diagram:



[Pedantically, in this diagram, LexCat must be understood as including lexcategories of comparable size to $\mathrm{Psh}(\omega)$, so that the self-indexing of $\mathrm{Psh}(\omega)$ is valued in LexCat.]

Again, keep in mind the contravariance of indexed structures here, so that $\mathrm{prior} : \mathrm{Prior} \to \mathrm{id}$ does indeed act as a map from any $\mathrm{Psh}(\omega)$-indexed structure X into the corresponding $\mathrm{Prior}^*(X)$.

Thus, we have constructed a locally introspective theory $\langle \mathrm{Psh}(\omega), C, \mathcal{F} \rangle$.

It may be illustrative to alternatively describe this $\mathcal{F}$ in terms of its corresponding $\mathcal{S}$ and $\mathcal{N}$.

For $\mathcal{S}$, let us first observe that $\mathrm{Prior}(1) = 1$; that is, Prior preserves the terminal object[5]. Accordingly, $\mathrm{Glob}(C) = \mathrm{Psh}(\omega)/\mathrm{Prior}(1) = \mathrm{Psh}(\omega)$. And as the component $\mathrm{prior}_1 : \mathrm{Prior}(1) \to 1$ must be the identity on the terminal object, the map $\mathcal{S} : \mathrm{Psh}(\omega) \to \mathrm{Glob}(C)$ corresponding to our choice of $\mathcal{F}$ becomes the identity under this identification.

Furthermore, the map $\mathrm{Hom}_C(1, \mathcal{S}(-)) : \mathrm{Psh}(\omega) \to \mathrm{Psh}(\omega)$ can be seen to be the right adjoint to Prior; thus, it is Later. Finally, as for the $\mathcal{N}$ corresponding to our $\mathcal{F}$, this will be the map $\mathrm{next} : \mathrm{id} \to \mathrm{Later}$ given by pulling $\mathrm{prior} : \mathrm{Prior} \to \mathrm{id}$ through the adjunction $\mathrm{Prior} \dashv \mathrm{Later}$.

This way of equipping $\mathrm{Psh}(\omega)$ as a locally introspective theory is illustrative. Unfortunately, this is not an introspective theory, as our $C = \mathrm{Psh}(\omega)/\mathrm{Prior}(-)$ is merely locally representable, not representable simpliciter.

We do not have that $\mathrm{Ob}(C)$ is itself an object of $\mathrm{Psh}(\omega)$. Essentially, the obstruction is that $C(\mathrm{yoneda}(n)) = \mathrm{Psh}(\omega)/\mathrm{Prior}(\mathrm{yoneda}(n)) = \mathrm{Psh}(\omega)/\mathrm{yoneda}(n-1)$ (for $n \geq 1$) has a proper class of objects, but the presheaves in $\mathrm{Psh}(\omega)$ are set-valued.

We might naively try to ameliorate this problem by replacing $\mathrm{Psh}(\omega) = \mathrm{Set}^\omega$ by $(\mathrm{Set}')^\omega$ where Set' is some full subcategory of Set, such as sets of cardinality below some particular cardinal. But it is soon seen that such a uniform cardinality constraint across all $n \in \omega$ will not be workable for fixing the issue.

Rather, what will fix the issue is to impose a varying cardinality constraint: We shall consider those presheaves whose values at each n come from a particular full sublexcategory Set_n of Set, where these restrictions get looser as n gets larger.

Construction 2.13 Let Set_n for each $n \in \omega$ be a set-sized full sublexcategory of Set. We shall think of each Set_n as a strict category. By $\mathrm{Psh}'(\omega)$, we mean the full sublexcategory of $\mathrm{Psh}(\omega)$ comprising presheaves P such that $P(m) \in \mathrm{Set}_m$ for all $m \in \omega$.

By $\omega_{<n}$, we mean the sub-poset of ω restricted to those naturals which are less than n. By $\mathrm{Psh}'(\omega_{<n})$, we mean the full sublexcategory of $\mathrm{Psh}(\omega_{<n})$ comprising presheaves such that $P(m) \in \mathrm{Set}_m$ for each $m < n$.

Observe that each $\mathrm{Psh}'(\omega_{<n})$ is a set-sized strict lexcategory. Its collection of objects and its collection of morphisms are readily seen to comprise bona fide sets.

We also have obvious restriction maps from $\mathrm{Psh}'(\omega_{<n})$ to $\mathrm{Psh}'(\omega_{<m})$ for $m \leq n$ induced by the inclusion of $\omega_{<m}$ into $\omega_{<n}$, and any composition of such restriction maps yields the appropriate such restriction map.

Thus, we have an ω-indexed set-sized strict lexcategory $C'(n) = \mathrm{Psh}'(\omega_{<n})$. In other words, this C' is a lexcategory internal to $\mathrm{Psh}(\omega)$.

[5] This is closely related to the fact that ω has no maximal element, and would need modification were we carrying out the analogous construction for a poset which had maximal elements.

Also note that once $\mathrm{Set}_0, \mathrm{Set}_1, \ldots, \mathrm{Set}_{n-1}$ are determined, we have already determined what $\mathrm{Psh}'(\omega_{<n})$ is. That is, $\mathrm{Psh}'(\omega_{<n})$ does not depend on the choices of Set_m for $m \geq n$.

Thus, we may inductively choose Set_n for each n such that both $\mathrm{Ob}(\mathrm{Psh}'(\omega_{<n}))$ and $\mathrm{Mor}(\mathrm{Psh}'(\omega_{<n}))$ are among the objects of Set_n.

[For example, we may satisfy this condition by choosing Set_n for each n to be the von Neumann universe $V_{(n+1)\times\omega}$ of sets of rank less than $(n+1) \times \omega$. Many other possibilities are available, this is only one suggestion.]

When we choose Set_n satisfying this inductive condition, we have that C' is not only internal to $\mathrm{Psh}(\omega)$, but indeed is internal to its full subcategory $\mathrm{Psh}'(\omega)$.

We now flesh $\langle \mathrm{Psh}'(\omega), C' \rangle$ out into an introspective theory, by defining an appropriate $\mathcal{S}$ and $\mathcal{N}$. Much like before, $\mathrm{Glob}(C')$ is readily identified with $\mathrm{Psh}'(\omega)$ and we take $\mathcal{S}$ to be this identification. As just as before, we find that under this identification, $\mathrm{Hom}_C(1, -)$ acts as Later, so we may take $\mathcal{N}$ to be $\mathrm{next} : \mathrm{id} \to \mathrm{Later}$.

This completes the description of $\langle \mathrm{Psh}'(\omega), C' \rangle$ as an introspective theory (relative to any suitable choice of the $\{\mathrm{Set}_n\}_{n\in\omega}$).

(Furthermore, the $\{\mathrm{Set}_n\}_{n\in\omega}$ clearly may be chosen so as that each particular Set_n contains any set-sized number of particular desired sets, so that $\mathrm{Psh}'(\omega)$ contains any set-sized number of particular desired objects of $\mathrm{Psh}(\omega)$.)

2.4.2.2 Presheaf examples related to Kripke frames

Here, we consider examples of locally introspective and introspective theories based on Kripke frames. Of note, our first construction of a locally introspective theory works for any transitive Kripke frame (corresponding to the K4 modal logic). When we attempt to make an introspective theory of this by imposing cardinality constraints, we will find we are only able to do this if the transitive Kripke frame is furthermore well-founded (corresponding to the GL modal logic).

CONSTRUCTION 2.14 Let $<$ be a transitive relation on a discrete set $|P|$. The reflexive closure $\leq$ of $<$ equips $|P|$ as a preorder P. Let Q be P augmented with one further element ∞ which is greater than every element from P.

There is an inclusion functor $i : |P| \to Q$, and this induces correspondingly a functor $i^* : \mathrm{Psh}(Q) \to \mathrm{Psh}(|P|)$.

By $|P|_{<q}$ (where q is any value in Q), we mean the discrete subset of $|P|$ comprising those values which, within Q, are less than q. Note that when $p \leq q$, there is a forgetful functor from $\mathrm{Psh}(|P|_{<q})$ to $\mathrm{Psh}(|P|_{<p})$ induced by the inclusion of $|P|_{<p}$ into $|P|_{<q}$. Any composition of such forgetful functors is another forgetful functor of the same form.

Thus, we obtain a Q-indexed lexcategory $C(q) = \mathrm{Psh}(|P|_{<q})$. It is straightforward to observe this is locally representable. We will now equip $\langle \mathrm{Psh}(Q), C \rangle$ as a locally introspective theory by providing a suitable $\mathcal{S}$ and $\mathcal{N}$.

Note that $\mathrm{Glob}(C) = \mathrm{Psh}(|P|_{<\infty}) = \mathrm{Psh}(|P|)$. (Here, the addition of ∞ into Q plays an important role when P contains maximal elements. If we took C to be merely a P-indexed category, then we would find that $\mathrm{Glob}(C)$ ignored any maximal elements in P. This is the only reason for our introduction of ∞.)

Thus, $i^* : \mathrm{Psh}(Q) \to \mathrm{Glob}(C)$. We may refer to this also as $\mathcal{S}$.

Furthermore, note that for $X \in \mathrm{Psh}(Q)$, we have that $\mathrm{Hom}_C(1, \mathcal{S}(X))$ is the presheaf on Q which assigns to $q \in Q$ the product of $X(p)$ over all $p < q$, with restriction maps given by forgetting components as appropriate. .

Thus, we have a map $\mathcal{N} : \mathrm{id}_{\mathrm{Psh}(Q)} \to \mathrm{Hom}_C(1, \mathcal{S}(-))$, such that $\mathcal{N}_X(q) : X(q) \to \mathrm{Hom}_C(1, \mathcal{S}(X))(q)$ is given by the product of all the restriction maps out of $X(q)$ (these restriction maps being part of the structure of the presheaf X itself).

Observation 2.15 The reader is advised to keep in mind that this last construction is very different from Construction 2.12, even if P is chosen to be the poset ω of natural numbers. This distinction is emphasized again in the later discussion at Kripke frame example (Section 3.4.2) and Step-indexing example (Section 3.4.3).

This is an important archetypal example of a locally introspective theory. It corresponds closely to the interpretation of K4 modal logic using a transitive Kripke frame (as we will discuss in Kripke frame example (Section 3.4.2)). However, just as at the beginning of our previous example, we have the issue that this is only a locally introspective theory and not an introspective theory. Once again, the various aspects of C comprise a proper class of objects, too many for $\mathrm{Ob}(C)$ to be given by a set-valued presheaf, preventing C from being representable. And as in our previous example, we will again address this by imposing variable cardinality constraints on our presheaves.

Construction 2.16 We will from hereon out assume that the preorder P is in fact well-founded (and thus so is Q). Suppose given set-sized full sublexcategories Set_q of Set for each $q \in Q$. (It's not actually necessary that we restrict to such a subcategory at $q = \infty$, but for uniformity's sake, we do this for now.). We shall think of each Set_q as a strict category.

We define $\mathrm{Psh}'(Q)$ to be the full sublexcategory of $\mathrm{Psh}(Q)$ comprising presheaves X for which $X(q) \in \mathrm{Set}_q$ for each $q \in Q$. And we analogously define $\mathrm{Psh}'(|P|_{<q})$ to be the full sublexcategory of $\mathrm{Psh}(|P|_{<q})$ comprising presheaves X for which $X(p) \in \mathrm{Set}_p$ for each $p < q$.

There are restriction maps from $\mathrm{Psh}'(|P|_{<q})$ to $\mathrm{Psh}'(|P|_{<p})$ for $p \leq q$ induced by the inclusion of $|P|_{<p}$ into $|P|_{<q}$, and any composition of such restriction maps is such a restriction map. As a result, we have a $\mathrm{Psh}(Q)$-internal lexcategory C' whose component at $q \in Q$ is given by $\mathrm{Psh}'(|P|_{<q})$.

The set-sized category $\mathrm{Psh}'(|P|_{<q})$ only depends on the values of Set_p for $p < q$, and thus we may inductively choose Set_q in such a way that $\mathrm{Ob}(\mathrm{Psh}'(|P|_{<q}))$ as well as $\mathrm{Mor}(\mathrm{Psh}'(|P|_{<q}))$ are both objects of Set_q for each $q \in Q$. When we have done so, it follows that C' is not merely internal to $\mathrm{Psh}(Q)$ but furthermore lives within $\mathrm{Psh}'(Q)$.

We observe that there is a forgetful lexfunctor $\mathcal{S} : \mathrm{Psh}'(Q) \to \mathrm{Glob}(C') = \mathrm{Psh}'(|P|_{<\infty})$, induced by the inclusion of $|P|_{<\infty} = |P|$ into Q.

We observe as above that for $X \in \mathrm{Psh}'(Q)$, we have that $\mathrm{Hom}_{C'}(1, \mathcal{S}(X))$ is the presheaf on Q which assigns to $q \in Q$ the product of $X(p)$ over all $p < q$, with restriction maps given by forgetting components as appropriate.

Finally, we define $\mathcal{N}$ in the same way as above, with $\mathcal{N}_X(q) : X(q) \to \mathrm{Hom}_{C'}(1, \mathcal{S}(X))(q)$ given by the product of all the restriction maps : $X(q) \to X(p)$ for $p < q$.

In this way, we have constructed an introspective theory $\langle \mathrm{Psh}'(Q), C' \rangle$ (relative to any suitable choice of the $\{\mathrm{Set}_q\}_{q \in Q}$).

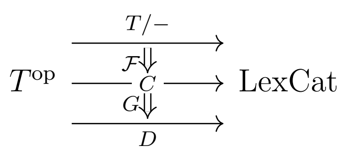
(Furthermore, the $\{\mathrm{Set}_q\}_{q \in Q}$ clearly may be chosen so as that each particular Set_q contains any set-sized number of particular desired sets, so that $\mathrm{Psh}'(Q)$ contains any set-sized number of particular desired objects of $\mathrm{Psh}(Q)$.)

This is an important archetypal example of an introspective theory. It corresponds closely to the interpretation of GL modal logic using a well-founded Kripke frame (as we will discuss in Kripke frame example (Section 3.4.2)).

2.5 Basic constructions

Now let us discuss some general constructions for building new (pre-)introspective theories from old ones or from other data.

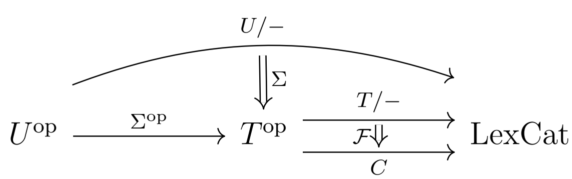
CONSTRUCTION 2.17 If $\langle T, C, \mathcal{F} \rangle$ is a pre-introspective theory, and any lexfunctor $G : C \to D$ is given for some other T-indexed lexcategory D, then $\langle T, D, G \circ \mathcal{F} \rangle$ is itself a pre-introspective theory, like so:



Of course, this yields an introspective or locally introspective theory just in case D is representable or locally representable, respectively.

CONSTRUCTION 2.18 If $\langle T, C, \mathcal{F} \rangle$ is a pre-introspective theory, U is any lexcategory, and $\Sigma : U \to T$ is any functor which preserves pullbacks (we do not require Σ to preserve the terminal object), then $\langle U, \Sigma^* C \rangle$ can naturally be equipped as a pre-introspective theory, like so:

Details.



The 2-cell labelled Σ above indicates the action of Σ when acting as a lexfunctor from U/u to $T/(\Sigma u)$ for each object u in U. (Note that, as finite limits in slice categories are given by pullbacks in the underlying category, and as Σ preserves pullbacks, we do indeed have that this functor from U/u to $T/(\Sigma u)$ preserves finite limits.)

By Theorem 1.10 or Theorem 1.20, if Σ has a right adjoint, we can further observe that if C is representable or locally representable, then so respectively will be $\Sigma^* C$. ■

A particular special case of the above which is often of importance is the following:

CONSTRUCTION 2.19 (SLICE (PRE-)INTROSPECTIVE THEORIES) If $\langle T, C, \mathcal{F} \rangle$ is a pre-introspective theory, and t is any object in T, then the slice category T/t can be equipped in a natural way as a pre-introspective theory as well. If we start from an introspective or locally introspective theory, then so respectively will be the result of this construction.

Details. By the previous construction (Construction 2.18), using the forgetful functor $\Sigma : T/t \to T$, which preserves pullbacks and has a right adjoint (given by pullback).

(Note that in this case, the corresponding 2-cell from $(T/t)/-$ to $T/\Sigma(-)$ is an equivalence, by how iterated slice categories amount to slice categories simpliciter.) ■

When we abuse language and speak of T/t as an introspective theory, the above construction is what we mean.

CONSTRUCTION 2.20 If $\langle T, C, \mathcal{F} \rangle$ is a pre-introspective theory, and S is a full sub-lexcategory of T (thus, with a full and faithful inclusion lexfunctor $i : S \to T$), then $\langle S, i^* C \rangle$ can be equipped in a natural way as a pre-introspective theory as well.

Details. By Construction 2.18 again, taking Σ to be the inclusion functor i. ■

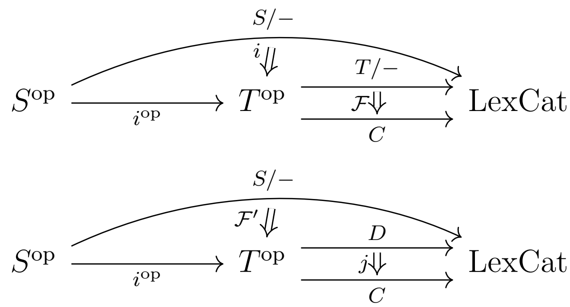
CONSTRUCTION 2.21 If $\langle T, C, \mathcal{F} \rangle$ is a pre-introspective theory, and D is a T-indexed full sub-lexcategory of C containing the range of $\mathcal{F}$ (thus, such that $\mathcal{F} = i \circ \mathcal{F}'$ for a uniquely determined $\mathcal{F}' : T/- \to D$, where $i : D \to C$ is the inclusion), then $\langle T, D, \mathcal{F}' \rangle$ is a pre-introspective theory.

[In this case, conversely, $\langle T, C, \mathcal{F} \rangle$ is obtained from $\langle T, D, \mathcal{F}' \rangle$ and $i : D \to C$ via Construction 2.17.]

The last two constructions are often fruitfully combined: Given a pre-introspective theory $\langle T, C, \mathcal{F} \rangle$, we may first pass from T to a sub-lexcategory S of T and then, after having done so, find that $\mathcal{F}$ when restricted to S factors through a sub-lexcategory D of C. In particular, the following scenario will be of note to us:

CONSTRUCTION 2.22 Let $\langle T, C, \mathcal{F} \rangle$ be a pre-introspective theory, let S be a full sublexcategory of T (with inclusion $i : S \to T$), and let D be a T-indexed full sub-lexcategory of C (with inclusion $j : D \to C$). Suppose furthermore that this D is of the form $i[D']$ for some representable S-indexed lexcategory D'. (It follows that this D' is identified with $i^* D$.)

Finally, suppose also that $\mathcal{F}$ restricted to S has range restricted to D, in that there is a (uniquely determined) $\mathcal{F}'$ making the following composite 2-cells equal:



Then $\langle S, D', \mathcal{F}' \rangle$ is an introspective theory, which we refer to as a **sub-introspection** of $\langle T, C, \mathcal{F} \rangle$.

Observe that our constructions Construction 2.13 and Construction 2.16 were given as sub-introspections of Construction 2.12 and Construction 2.14, respectively.

The concepts of pre-introspective, locally introspective, or introspective theories are all nearly essentially algebraic concepts ("nearly", in that these involve categories up to equivalence rather than strict categories up to isomorphism). Thus these concepts automatically have available all the same properties as for any such nearly essentially algebraic concept. For example, we have free constructions, as are the subject of of our later chapter Geminal categories (Chapter 5). And we have Cartesian products in the straightforward way:

Construction 2.23 Let K be any set, and suppose for each $k \in K$ we are given some pre-introspective theory $\langle T_k, C_k \rangle$. Then we may define the product of these pre-introspective theories in the obvious way. That is, we take the lexcategory $T = \prod_{k \in K} T_k$, and we define also a T-indexed lexcategory C, such that $C(t)$ for any object $t = \{t_k\}_{k \in K}$ in T is the product of $C_k(t_k)$ over each $k \in K$. Similarly, we define the reindexing functors in C componentwise using the reindexing functors the various C_k, and we also define $\mathcal{F} : T/- \to C$ componentwise using the $\mathcal{F}$ for the various pre-introspective theories $\langle T_k, C_k \rangle$. It is readily seen that the result is furthermore locally introspective or introspective if each $\langle T_k, C_k \rangle$ is locally introspective or introspective, respectively.

2.6 The interaction of $\mathcal{S}$ and $\mathcal{N}$

We gather here two small but useful lemmas for reasoning about (pre-)introspective theories, concerning the interaction of $\mathcal{S}$ and $\mathcal{N}$.

Lemma 2.24 ($\mathcal{S}$ With $\mathcal{N}$) Within a pre-introspective theory $\langle T, C \rangle$, let $F : X \to t$ be a morphism of T, and let x be any generalized element of X. We have that $\mathcal{N}_t(F(x)) = \mathcal{S}(F) \circ_C \mathcal{N}_X(x)$.

Proof. This is just the naturality square for $\mathcal{N}$ with respect to F.

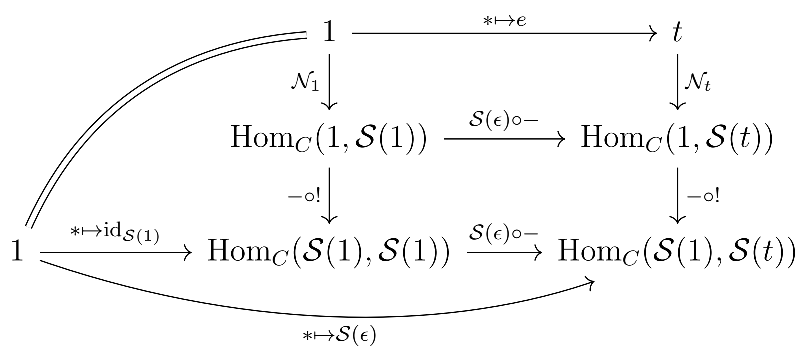
$$\begin{array}{ccc} X & \xrightarrow{\quad F \quad} & t \\ {\scriptstyle \mathcal{N}_X}\big\downarrow & & \big\downarrow{\scriptstyle \mathcal{N}_t} \\ \mathrm{Hom}_C(1,\mathcal{S}(X)) & \xrightarrow[\mathcal{S}(F)\circ_C -]{} & \mathrm{Hom}_C(1,\mathcal{S}(t)) \end{array}$$

∎

LEMMA 2.25 ($\mathcal{S}$ MATCHES $\mathcal{N}$) Within a pre-introspective theory $\langle T, C\rangle$, let t be some object of T and let $\epsilon : 1 \to t$ in T be taken as defining a global element e of t. Then the global element $\mathcal{S}(\epsilon)$ of $\mathrm{Hom}_C(\mathcal{S}(1), \mathcal{S}(t))$ is equal to the the global element $\mathcal{N}_t(e)$ of $\mathrm{Hom}_C(1, \mathcal{S}(t))$ under the canonical isomorphism identifying $\mathrm{Hom}_C(\mathcal{S}(1), \mathcal{S}(t))$ with $\mathrm{Hom}_C(1, \mathcal{S}(t))$.

In short, $\mathcal{S}$ and $\mathcal{N}$ take global elements in T to equal global elements of $C(1)$.

Proof. Consider the following commutative diagram in $\mathrm{Psh}(T)$.



The top arrow is $\epsilon : 1 \to t$, thought of as sending the unique element of 1 to e. The top rectangle is the naturality square for $\mathcal{N}$ with respect to ϵ.

The bottom rectangle is the associativity square for composition in C (specifically, on one side composing with $\mathcal{S}(\epsilon) : \mathcal{S}(1) \to \mathcal{S}(t)$ and on the other side composing with the unique morphism $! : \mathcal{S}(1) \to 1$). Note that the right arrow of this associativity rectangle is the canonical isomorphism given by $\mathcal{S}(1)$ being a terminal object of C.

The bottom wedge is the identity law for composition in C (specifically, composing after the identity on $\mathcal{S}(1)$).

Finally, the left wedge commutes because, as $\mathcal{S}(1)$ is a terminal object of C, we have that $\mathrm{Hom}_C(\mathcal{S}(1), \mathcal{S}(1))$ is a terminal object of $\mathrm{Psh}(T)$; thus, any two parallel maps into it are equal. (Indeed, all arrows in the left wedge are unique isomorphisms between terminal objects.)

Now consider the composites around this commutative diagram along the two outermost paths. Along the bottom, the unique element of 1 is sent to $\mathcal{S}(\epsilon)$. Along the top and right, it is sent to $\mathcal{N}_t(e)$ and then along the canonical isomorphism. This completes the proof.

(We would not ordinarily bother to distinguish between 1 and $\mathcal{S}(1)$ or in general explicitly write out the coherence isomorphisms for a product preserving functor, but in this one example it may be illuminating to see these distinctions and isomorphisms explicitly.) ■

2.7 Recap

We have defined the central notion of our interest, the concept of an introspective theory. We have proven that two different definitions of this concept are equivalent. We have also discussed some slight relaxations of this concept (pre-introspective theories and locally introspective theories). We have seen three archetypal examples of introspective theories (one constructed by considering Σ_1 formulae in familiar theories such as ZF-Finite, another constructed by considering step-indexing in the topos of trees, and another constructed from well-founded transitive Kripke frames). Finally, we have discussed a number of constructions which generate new introspective theories from existing ones.

Chapter 3

Modal logic

3.1 Preview

In this chapter, we will show how to interpret the $\Box$ operator of traditional modal logic in the context of introspective theories (or their generalizations). In particular, after defining the $\Box$ operator in this context, we observe in this chapter how it satisfies the rules of the modal logic K4.

3.2 The box operator

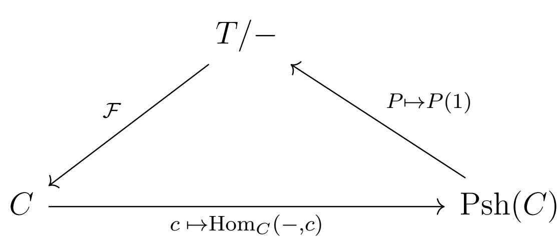
The following notation will be very convenient for us going forward. It is also suggestive of connections with modal logic we will eventually explore:

Let $\langle T, C\rangle$ be a locally introspective theory.

Recall from Presheaves over Indexed Categories (Definition 1.40) the T-indexed category $\mathrm{Psh}(C)$ (the appropriate notion of the category of presheaves on C when C is an indexed category rather than a category simpliciter).

Thus, we have three T-indexed lexcategories of note: T itself (considered as a T-indexed category through the self-indexing $T/-$), C, and $\mathrm{Psh}(C)$.

Between these, we also have a cycle of T-indexed lexfunctors, like so:



Here, the bottom arrow is the Yoneda embedding, sending each object of C to the corresponding representable presheaf. The right arrow takes a presheaf on C to its evaluation at the terminal object of C; that is, to its global elements. The left arrow is the $\mathcal{F}$ which is part of the structure of an introspective theory.

OBSERVATION 3.1 Via Theorem 1.8, we have the t-aspect of the above triangle of indexed lexfunctors is the same as the global aspect of the same triangle of indexed lexfunctors relative to the slice locally introspective theory T/t given by Slice (Pre-)Introspective Theories (Construction 2.19).

DEFINITION 3.2 In general, we will write $\Box$ for a roundtrip around this diagram, starting from any of its three nodes.

Thus, we will write $\Box$ for the T-indexed lexfunctor from T to itself given by $t \mapsto \mathrm{Hom}_C(1, \mathcal{F}(t))$.

We will ALSO write $\Box$ for the T-indexed lexfunctor from C to itself given by $c \mapsto \mathcal{F}(\mathrm{Hom}_C(1, c))$.

And we will ALSO write $\Box$ for the T-indexed lexfunctor from $\mathrm{Psh}(C)$ to $\mathrm{Psh}(C)$, which sends the presheaf P to the presheaf represented by $\mathcal{F}(P(1))$.

When we want to clarify precisely the domain we are operating on, we will write names such as $\Box_{T/-}$, $\Box_C$, or $\Box_{\mathrm{Psh}(C)}$, as appropriate.

As the Yoneda embedding is naturally thought of as the inclusion of a full sub-lexcategory, identifying C with the corresponding representable presheaves within $\mathrm{Psh}(C)$, we may also think of $\Box_{\mathrm{Psh}(C)}$ as a T-indexed lexfunctor from $\mathrm{Psh}(C)$ to C. That is, as the composition of merely the top two arrows above.

The above was all discussed for T, C, and $\mathrm{Psh}(C)$ considered as T-indexed lexcategories, but this all (and the rest of this chapter as well) descends to corresponding structure on their global aspects as well. In particular, we may write $\Box_T$ to denote the global aspect of $\Box_{T/-}$. Keep in mind, the global aspect of $\mathcal{F}$ is $\mathcal{S}$, so wherever in the above we discussed $\mathcal{F}$, this may be rewritten as $\mathcal{S}$ when considering just the global aspect.

REMARK The reason we restricted attention here to T-representable (T, C)-indexed sets (these being the objects of $\mathrm{Psh}(C)$ as we've defined it), rather than arbitrary (T, C)-indexed sets, is so that the map $P \mapsto P(1)$ can indeed be taken as always landing back within T, and not within $\mathrm{Psh}(T)$ more generally.

Similarly, the reason we restricted attention to locally introspective theories (i.e., the case where C is locally T-representable), and not to pre-introspective theories more generally, is so that the Yoneda embedding $c \mapsto \mathrm{Hom}_C(-, c)$ does indeed yield T-representable (T, C)-indexed sets.

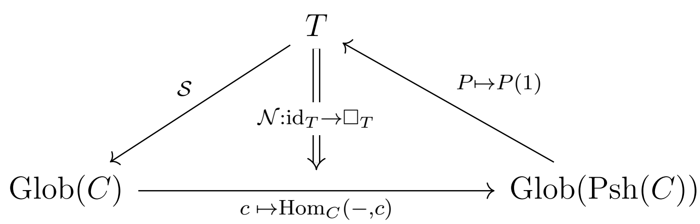
3.3 Modal logic and axiom 4

The choice of this $\Box$ notation for these purposes is meant to convey an analogy with the $\Box$ operator of modal logic, and in particular, with the provability operator of provability logic. We will explore this more in later remarks.

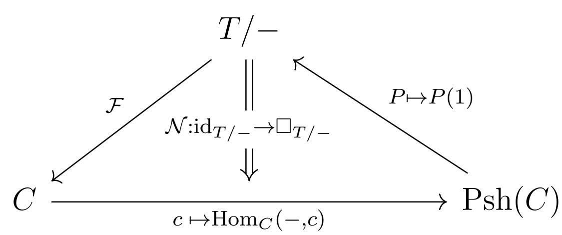
The key point here is that the rules of the $\Box$ operator in a Kripke normal modal logic are essentially the rules of a lex endofunctor on a lexcategory, and any of our $\Box$ operators is certainly lex as a composite of lex functors.

Furthermore, each of our $\Box$ operators comes with a natural transformation from $\Box$ to $\Box\Box$ corresponding to the so-called 4 axiom $\Box A \vdash \Box\Box A$ in modal logic.

For the $\Box$ operator on T this is clear, as the natural transformation $\mathcal{N}$ from identity to $\Box$ encodes the even stronger property $t \vdash \Box t$. The 4 axiom is the special case where t here is of the form $\Box A$.

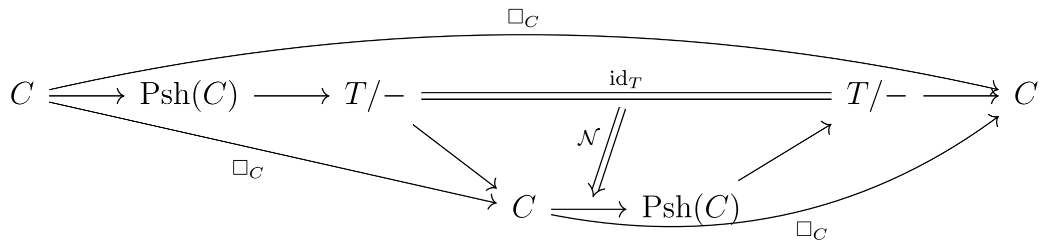


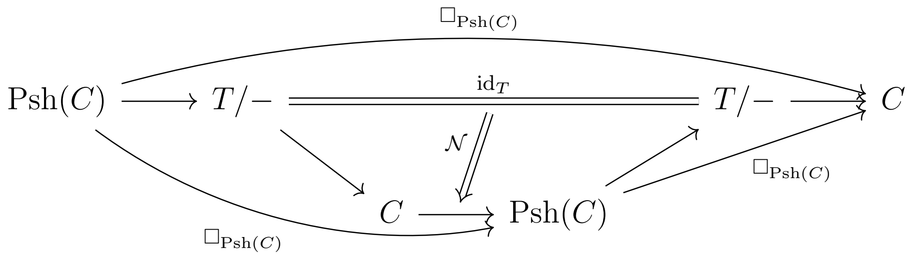
Keeping in mind Observation 3.1, this extends automatically to an indexed endofunctor on $T/-$, which we may call $\Box_{T/-}$, whose aspect at each $t \in T$ is given by the above diagram relative to the slice locally introspective theory T/t. (We abuse notation slightly in the following diagram and write $\mathcal{N}$ for the map from identity to $\Box_{T/-}$ whose each aspect is given by the appropriate instance of the previous diagram.)



We do not have such a strong natural transformation from identity to $\Box$ as acting on the other corners of the triangle (C or $\mathrm{Psh}(C)$). However, by taking the natural transformation from t to $\Box_T t$ and whiskering it on both sides along the trips from any other corner of the triangle into and out of T, we get a natural transformation from $\Box$ to $\Box\Box$ at each other corner of the triangle as well.

(The general principle here is that, given morphisms f and g between the same pair of objects in opposite directions, composing with f on one side and g on the other turns $(gf)^n$ into $(fg)^{n+1}$, and thus whiskering in this way turns a natural transformation : $(gf)^n \to (gf)^m$ into a natural transformation : $(fg)^{n+1} \to (fg)^{m+1}$)





(In the above two diagrams, all parallel paths commute except where an explicit natural transformation is noted, and all unmarked edges are the corresponding of the three edges along our triangle. In the diagram illustrating axiom 4 for $\square_{\mathrm{Psh}(C)}$, we note that we can think of the codomain of $\square_{\mathrm{Psh}(C)}$ as either $\mathrm{Psh}(C)$ or more narrowly its full subcategory C.)

Thus, our $\square$ follows all the rules of the modal logic K4, in each of these contexts. Shortly, we shall see that the general logic followed by $\square_C$ in a locally introspective theory is conversely no stronger than K4, while in an introspective theory, it is the modal logic GL. Indeed, in the very next chapter we will see how in an introspective theory we get the last ingredient for the modal logic GL, Löb's theorem.

3.4 As applied to our archetypal examples

3.4.1 ZF-Finite examples

Recall from Construction 2.10 the introspective theory $\langle \mathrm{Z}_{\Sigma_1}, \mathrm{Z}', \mathcal{S}, \mathcal{N} \rangle$, where Z_{Σ_1} is the lexcategory of Σ_1-definable hereditarily finite sets and Σ_1-definable functions between them up to provable equivalence in ZF-Finite, Z' is the lexcategory internal to Z_{Σ_1} of arbitrary definable sets and arbitrary definable functions between them up to provable equivalence in ZF-Finite, $\mathcal{S}$ assigns to each piece of Z_{Σ_1} the corresponding (globally defined) piece of Z', and $\mathcal{N}$ is the Σ_1-definable function which sends any hereditarily finite set to the canonical term describing it, as well as witnessing the provable entailment from truth to provability for Σ_1 formulae.

Let us now consider the behavior of our $\square$ operators here. For convenience of discussion, we will discuss just their global aspects.

First, observe that we have $\square_{\mathrm{Z}_{\Sigma_1}} : \mathrm{Z}_{\Sigma_1} \to \mathrm{Z}_{\Sigma_1}$, defined by $\mathrm{Hom}_{\mathrm{Z}'}(1, \mathcal{S}(-))$. This takes any object in Z_{Σ_1} and sends it to the object of global sections of the corresponding object in

Z_{Σ_1}'s internal view of itself via Z'. This is like sending "the object of Xes" to "the object of definitions of Xes within ZF-Finite". In the particular case where the object in question is subterminal (thus representing a proposition), this is like sending the proposition X to the proposition "There is a proof in ZF-Finite of X". Thus, this operator acts just as the traditional provability operator in this context.

Next, we turn our attention to the Z_{Σ_1}-internal endolexfunctor $\Box : Z' \to Z'$. Again, we shall just look at its global aspect. Presuming the soundness of ZF-Finite in our metatheory for further convenience, we have that $\mathrm{Glob}(Z')$ is identified with Z, the lexcategory of arbitrary definable sets and arbitrary definable functions between them up to provable equivalence in ZF-Finite. Thus, we are considering $\Box : Z \to Z$, defined by $\mathcal{S}(\mathrm{Hom}_Z(1, -))$. This again takes any object in Z and sends it to the object of global sections of the corresponding object in Z' (the internal view Z has of itself). So again, in the same way, this is like sending "the object of Xes" to "the object of Xes which are definable within ZF-Finite", and in the particular case where the object in question is subterminal and thus represents a proposition, this is like sending the proposition X to the proposition "There is a proof in ZF-Finite of X". So again, this operator acts just as the traditional provability operator in this context.

3.4.2 Kripke frame example

Recall the locally introspective theory $\langle \mathrm{Psh}(Q), C\rangle$ from Construction 2.14, constructed from a transitive relation $<$ on a set P, with Q being P augmented with a new maximum element ∞ and construed as a preorder category using the $<$ relation. Our definition of C was such that $\mathrm{Glob}(C) = \mathrm{Psh}(|P|) = \mathrm{Set}^{|P|}$. Our definition of $\mathcal{S} : \mathrm{Psh}(Q) \to \mathrm{Psh}(|P|)$ was the forgetful functor which ignores the $<$ relation and ∞ element in Q. Our definition of C also ensures that $\mathrm{Hom}_C(1, -) : \mathrm{Psh}(|P|) \to \mathrm{Psh}(Q)$ sent $f \in \mathrm{Psh}(|P|)$ to the presheaf F on Q such that $F(q)$ is the product of $f(p)$ over all $p < q$, with restriction maps given by projection.

Thus, $\Box_{\mathrm{Psh}(|P|)}$ acts like so: For $F \in \mathrm{Psh}(|P|)$ and $x \in P$, we have that $\Box_{\mathrm{Psh}(|P|)}(F)(x) = \mathcal{S}(\mathrm{Hom}_C(1, F))$ is the product of $F(y)$ over all $y < x$.

Similarly, $\Box_{\mathrm{Psh}(Q)}$ acts like so: For $F \in \mathrm{Psh}(Q)$ and $x \in Q$, we have that $\Box_{\mathrm{Psh}(Q)}(F)(x)$ is the product of $F(y)$ over all $y < x$. (This $\Box_{\mathrm{Psh}(Q)}(F)$ furthermore is equipped with restriction maps given by projection.)

In particular, we may consider the special case of subterminal presheaves, amounting to propositions which may be true in some places while not in others. Subterminal presheaves in $\mathrm{Psh}(|P|)$ are simply subsets of P. Given such a subset $F \subseteq P$, the above tells us that $\Box_{\mathrm{Psh}(|P|)}(F)$ is identified with the (downwards closed) subset of P containing precisely those x such that all $y < x$ are contained in F.

Similarly, subterminal presheaves in $\mathrm{Psh}(Q)$ are simply downwards closed subsets of Q. Given such a downwards closed subset $F \subseteq Q$, the above tells us that $\Box_{\mathrm{Psh}(Q)}(F)$ is identified with the downwards closed subset of Q containing precisely those x such that all $y < x$ are contained in F.

Thus, our notion of $\Box$ acts on these subterminal presheaves in precisely the manner of the traditional interpretation of the $\Box$ operator in Kripke frames.

In this way, we may note that any identity which holds of the $\Box_C$ operator in all locally introspective theories $\langle T, C\rangle$ must be an identity which holds of the $\Box$ operator on propositions in Kripke frames constructed from a transitive $<$ relation. This in turn is well known to entail holding of the $\Box$ operator in the modal logic K4 . Thus, in this sense (which we refrain from stating formally for now, though it clearly could be turned into various formal statements), the logic of the $\Box$ operator in locally introspective theories is no stronger than the modal logic K4. We've also already seen in Modal logic and axiom 4 (Section 3.3) that it is at least as strong as the modal logic K4, and thus the logic of the $\Box$ operator in general locally introspective theories can be identified with K4.

In the above, we discussed the locally introspective theory constructed by Construction 2.14 from a transitive relation $<$. Recall that, in the particular case where $<$ is furthermore well-founded, we may impose cardinality constraints in a suitable fashion to turn this into an introspective theory (this being Construction 2.16). The nature of the $\Box$ operators on the introspective theory Construction 2.16 will be essentially the same as for Construction 2.14, as the former is a sub-introspection of the latter. Thus, again, these $\Box$ operators will match the traditional interpretation of the box operator in Kripke frames. In this way, we find that the logic of the $\Box_C$ operators in introspective theories $\langle T, C\rangle$ is no stronger than that of the $\Box$ operator in well-founded transitive Kripke frames, which is well known to correspond to the modal logic GL (given by K4 + Löb's theorem). In the next chapter, we will prove the remarkable fact that Löb's theorem (and thus all of GL) is validated in all introspective theories, so that the logic of the $\Box_C$ operator in introspective theories can be said to be precisely that of the modal logic GL.

3.4.3 Step-indexing example

Recall the locally introspective theory $\langle \mathrm{Psh}(\omega), C\rangle$ from Construction 2.12, in which $\mathcal{S} : \mathrm{Psh}(\omega) \to \mathrm{Glob}(C)$ acts as an equivalence. It was also observed in that construction that $\mathrm{Hom}_C(1, \mathcal{S}(-)) : \mathrm{Psh}(\omega) \to \mathrm{Psh}(\omega)$ acts as the operator Later, such that $\mathrm{Later}(P)(0) = 1$ and $\mathrm{Later}(P)(n+1) = P(n)$, for $P \in \mathrm{Psh}(\omega)$ and $n \in \omega$.

Thus, the operator $\Box_{\mathrm{Psh}(\omega)}$ here is precisely this operator Later.

Note that this is quite distinct from the $\Box$ operator obtained for Construction 2.14 as discussed above, even when considering the poset ω. The distinction is that $\mathrm{Later}(P)(n)$, for positive n, is not the product of $P(m)$ at all $m < n$ but rather simply their limit, $P(n-1)$.

However, this distinction is not visible when considering only subterminal presheaves in $\mathrm{Psh}(\omega)$. These amount to downwards closed subsets of ω, and for such a downwards closed subset $F \subseteq \omega$, we will have that $\Box_{\mathrm{Psh}(\omega)}$ is the downwards closed subset of ω which includes n just in case either $n = 0$ or $n - 1 \in F$, which in turn happens just in case all $m < n$ are contained in F. This again is the same behavior as the traditional $\Box$ operator on a Kripke frame corresponding to ω. Note however that in this locally introspective theory, we do not have the ability to discuss propositions which are not downwards closed.

The $\Box$ operators on the introspective theory Construction 2.13 are again essentially the same as those on the locally introspective theory Construction 2.12, as the former is a sub-introspection of the latter.

3.5 Recap

We have defined the $\Box$ operator in the context of locally introspective theories $\langle T, C\rangle$, acting on each of T, C, and $\mathrm{Psh}(C)$.

We have furthermore shown how each of these $\Box$ operators satisfies the rules of K4 modal logic.

We have also seen that the traditional interpretation of K4 modal logic in transitive Kripke frames is a special case of our $\Box$ operator for locally introspective theories (thus letting us conclude a tight correspondence between locally introspective theories and K4 modal logic).

The traditional provability operator on ZF-Finite (or just as well, on Peano Arithmetic) was also seen as a special case of our $\Box$ operator on an introspective theory. And the step-indexing operator considered in guarded recursion on the topos of trees was also seen as a special case of our $\Box$ operator on locally introspective or introspective theories.

In the next chapter, we will show that in any introspective theory, the $\Box$ furthermore satisfies the Löb property of the modal logic GL (aka, "provability logic"). Thus, the correspondence between locally introspective theories and the modal logic K4 is complemented by a correspondence between introspective theories and the modal logic GL.

Chapter 4

Löb's theorem

4.1 Preview

In this chapter, we prove our most important theorem, justifying the significance of the simple concept of introspective theories. We show how every introspective theory automatically satisfies a general version of Löb's theorem, acting as the construction of general fixed points. We will also see how Löb's theorem is in full generality a phenomenon linked to presheaves, and not only constrained to representable presheaves.

The key results of this chapter are those covered in Presheaf diagonalization for pre-introspective theories (Section 4.4) and Bootstrapping to Löb's theorem for introspective theories (Section 4.5), culminating in Löb's Theorem for Introspective Theories (Theorem 4.19), our most important theorem. All material in those sections is original to this work.

The material on the Löb property in general categories in The Löb property in abstract (Section 4.2) includes some observations which can also be found (either explicitly or implicitly) in existing literature. We give our own exposition of this material, which felt important to include in a clean and complete exposition of the significance of our key results. In particular, we confirm how these general Löb property results continue to be applicable in our particular introspective theory context, even without common presumptions such as cartesian closure, and even with care taken to distinguish the roles of T, C, and $\mathrm{Psh}(C)$ in a general introspective theory $\langle T, C \rangle$.

The discussion in Lawvere's fixed point theorem (Section 4.3) concerns Lawvere's fixed point theorem, which of course is not original to us, but we also include some reframing and generalization of this which is due to us rather than Lawvere. The discussion in Relating variations on Lawvere's fixed point theorem (Section 4.8) compares our reframing to some other reframings of Lawvere's fixed point theorem in the existing literature.

The sole theorem in The self-indexing cannot be representable, except trivially (Section 4.6) is the same theorem as proven in [PT89]. We re-note it here simply to observe that it follows as a special case of our more general Löb's Theorem for Introspective Theories

(Theorem 4.19).

4.2 The Löb property in abstract

Definition 4.1 Let D be any category with a terminal object and let $\Box : D \to D$ be a terminal-object-preserving endofunctor on D. We say $\Box$ has the **Löb property** if, for every object Ω of D and every morphism $g : \Box\Omega \to \Omega$, there exists a morphism $\omega : 1 \to \Omega$ making the following square commute:

$$\begin{array}{ccc} \Box 1 & \xrightarrow{\Box\omega} & \Box\Omega \\ \| & & \downarrow g \\ 1 & \xrightarrow[\omega]{} & \Omega \end{array}$$

In other words, for every $\Omega \in D$ and $g : \Box\Omega \to \Omega$, there is a fixed point of $\omega \mapsto g \circ (\Box\omega) : \mathrm{Hom}_D(1, \Omega) \to \mathrm{Hom}_D(1, \Omega)$.

If such fixed points are furthermore always unique, we say $\Box$ has the **Löb property with uniqueness**. (Note that the Löb property with uniqueness is the same as saying that the unique map $: \Box 1 \to 1$ is an initial $\Box$-algebra.)

Observation 4.2 For example, the identity endofunctor on the category of complete lattices and monotonic maps between them has the Löb property (this amounts to the Knaster-Tarski fixed point theorem). However, this does not have the Löb property with uniqueness.

In this chapter, we will establish that for every introspective theory $\langle T, C\rangle$, each aspect of each of $\Box_{T/-}$, $\Box_C$, and $\Box_{\mathrm{Psh}(C)}$ has the Löb property with uniqueness. That such a strong result follows from such a minimal and simple categorical structure motivates much of our interest in the concept of introspective theories.

But before we establish this version of Löb's theorem for introspective theories in particular, we will develop the theory of the Löb property and its consequences a little further in abstract.

Theorem 4.3 Let D and E be categories with terminal objects, and let $M : D \to E$ and $N : E \to D$ be functors preserving terminal objects. Suppose $NM : D \to D$ has the Löb property. Then so does $MN : E \to E$. Furthermore, if NM has the Löb property with uniqueness, then so does MN.

Proof. This is by the general theorem that fixed points of a composition of functions are in bijection with fixed points of any cyclic rearrangement of that composition (as f and g themselves restrict to inverse maps between fixed points of gf and fg). In particular, letting $comp_x(y) = x \circ y$, the fixed points of $\omega_E \mapsto comp_g(M(N(\omega_E)))$ are in bijection with the fixed points of $\omega_D \mapsto N(comp_g(M(\omega_D)))$, which is to say, of $\omega_D \mapsto comp_{N(g)}(N(M(\omega_D)))$.

The latter fixed points must exist (or exist uniquely) if $NM : D \to D$ has the Löb property (or the Löb property with uniqueness, respectively), and thus in such cases so do the former fixed points, establishing the corresponding property for $MN : E \to E$. ■

THEOREM 4.4 Let D be any lexcategory, and let $\Box : D \to D$ be a terminal-object-preserving endofunctor on D. If $\Box$ has the Löb property, then it furthermore has the Löb property with uniqueness.

Proof. We must show that, given any two commutative squares as below (with the same g on the right hand side of each), the morphisms ω and ψ are equal:

$$\begin{array}{ccc} \Box 1 & \xrightarrow{\Box\omega} & \Box\Omega \\ \| & & \downarrow{\scriptstyle g} \\ 1 & \xrightarrow[\omega]{} & \Omega \end{array} \qquad\qquad \begin{array}{ccc} \Box 1 & \xrightarrow{\Box\psi} & \Box\Omega \\ \| & & \downarrow{\scriptstyle g} \\ 1 & \xrightarrow[\psi]{} & \Omega \end{array}$$

Let $h : H \to 1$ be the equalizer of ω and ψ. We will have that $\omega = \psi$ just in case h is an isomorphism. As this h is monic, making H a subobject of 1, we will have that h is an isomorphism just in case there is any map from 1 to H.

Thanks to the Löb property, this in turn occurs just in case there is some map from $\Box H$ to H. And by the definition of H as an equalizer, this occurs just in case there is some map from $\Box H$ to 1 which gives equal results when composed with ω and with ψ.

But the map $\Box h : \Box H \to 1$ does indeed have this property, as seen in the following commutative diagram (where the top left square commutes because $h; \omega = h; \psi$):

$$\begin{array}{ccccc} \Box H & \xrightarrow{\Box h} & \Box 1 & = & 1 \\ {\scriptstyle \Box h}\downarrow & & {\scriptstyle \Box\omega}\downarrow & & \downarrow{\scriptstyle \omega} \\ \Box 1 & \xrightarrow{\Box\psi} & \Box\Omega & \xrightarrow{g} & \Omega \\ \| & & \downarrow{\scriptstyle g} & & \\ 1 & \xrightarrow[\psi]{} & \Omega & & \end{array}$$

This completes the proof. ■

REMARK Note that Theorem 4.4 makes essential use of the structure available in a lexcategory. We can see this by considering the example from Observation 4.2, which has the Löb property but not the Löb property with uniqueness. This is possible as the category of complete lattices and arbitrary monotonic maps lacks equalizers.

The application of these abstract results to locally introspective theories in particular is like so:

THEOREM 4.5 If $\langle T, C \rangle$ is a locally introspective theory and t is an object of T such that at least one of $\Box_{T/-}$, $\Box_C$, or $\Box_{\mathrm{Psh}(C)}$ has the Löb property (without presumed uniqueness) at its t-aspect, then all three have the Löb property with uniqueness at their t-aspect.

If this happens for every $t \in T$, we say this locally introspective theory itself has the **Löb property**.

Proof. By Theorem 4.3, when considering the definitions of the various $\Box$ operators given via the triangle at Definition 3.2, we find that if any of these $\Box$ operators have the Löb property at their t-aspect, then all three do. By Theorem 4.4, we can furthermore conclude the Löb property with uniqueness. ∎

Theorem 4.4 is only a special case of a much broader and important theorem which we now discuss.

Theorem 4.6 Let D be any lexcategory and let $\Box : D \to D$ be any terminal-object-preserving functor. Let E be any representable D-indexed category. (Note that $\Box$ acting on E induces also another representable D-indexed category $\Box E$, as well as a functor from each d-defined aspect of E to the $(\Box d)$-defined aspect of $\Box E$, for $d \in D$. In particular, as $\Box$ is terminal-object-preserving, $\Box$ acts as a functor from the global aspect of E to the global aspect of $\Box E$.)

Suppose also given a D-indexed functor $f : \Box E \to E$, and let the endofunctor F on the global aspect of E be given by first applying $\Box$ to arrive in the global aspect of $\Box E$, then applying f to arrive back in the global aspect of E.

If $\Box$ has the Löb property, then there is an F-hylomorphism (as in Definition 1.82) between any F-coalgebra $W : w \to F(w)$ and any F-algebra $M : F(m) \to m$ in the global aspect of E. And if $\Box$ furthermore has the Löb property with uniqueness, then this hylomorphism is unique.

Proof. A hylomorphism from W to M is a fixed point of $x \mapsto M \circ F(x) \circ W : \mathrm{Hom}_E(w, m) \to \mathrm{Hom}_E(w, m)$. But as $F(x) = f(\Box x)$, this is the same as a fixed point for $x \mapsto g(\Box x)$ where $g(-)$ is defined by $M \circ f(-) \circ W : \Box\, \mathrm{Hom}_E(w, m) \to \mathrm{Hom}_E(w, m)$.

The hylomorphisms from W to M are thus the same as the fixed points given by the Löb property with respect to this g. This completes the proof. ∎

We now demonstrate how Theorem 4.4 can be seen as a special case of Theorem 4.6:

Corollary 4.7 Let D be any lexcategory, and let $\Box : D \to D$ be a terminal-object-preserving endofunctor on D. If $\Box$ has the Löb property, then it furthermore has the Löb property with uniqueness.

Proof. Let E be an arbitrary object of D (thus, a representable D-indexed set) and let us construe this also as a representable D-indexed discrete category. Let $f : \Box E \to E$ be an arbitrary map in D, and as above, let us take $F : \mathrm{Hom}_D(1, E) \to \mathrm{Hom}_D(1, E)$ to be given as the composition of $\Box : \mathrm{Hom}_D(1, E) \to \mathrm{Hom}_D(1, \Box E)$ with $f \circ - : \mathrm{Hom}_D(1, \Box E) \to \mathrm{Hom}_D(1, E)$.

As E is a discrete category, observe that any F-coalgebra or F-algebra in the global aspect of E amounts to a fixed point of $f \circ \Box(-) : \mathrm{Hom}_D(1, E) \to \mathrm{Hom}_D(1, E)$. The Löb property tells us such fixed points exist, while Theorem 4.6 tells us there is a hylomorphism between any such fixed points. But as E is a discrete category, such a hylomorphism amounts to just an equality between the two elements of $\mathrm{Hom}_D(1, E)$. Thus, any two such fixed points are equal, which is to say, we have the Löb property with uniqueness. ■

(It is perhaps easy to miss how the presumption of equalizers in D has been used in the argument for Corollary 4.7. At one point within its invocation of Theorem 4.6, the argument considers the object $\Box\,\mathrm{Hom}_E(w, m)$ for parallel $w, m \in \mathrm{Hom}_D(1, E)$. As such, it depends upon the fact that $\mathrm{Hom}_E(w, m)$ is a representable D-indexed set. This object $\mathrm{Hom}_E(w, m)$ of D is given by an equalizer between parallel maps from 1 to E in D; this is where the fact that D is a lexcategory is essential.)

Corollary 4.8 Consider the same setup as of Theorem 4.6, and presume $\Box$ has the Löb property with uniqueness (as we now know follows automatically from the Löb property on a lexcategory). Then any fixed point of F (in the sense of an object e of the global aspect of E along with an isomorphism between e and $F(e)$) is simultaneously an initial F-algebra and a terminal F-coalgebra. In particular, any two such fixed points are isomorphic, via a unique F-algebra isomorphism.

Proof. In that context, Theorem 4.6 says that every F-coalgebra has a *unique* hylomorphism into every F-algebra. In the particular case that the coalgebra is invertible, this can be read as a morphism between algebras, and establishes that the coalgebra's inverse is an initial algebra. Dually, for any invertible algebra, this establishes its inverse as a terminal coalgebra. ■

Remark The argument we have given for Theorem 4.6 and thus for Corollary 4.8 is essentially the same as that given for Lemma 7.6 in [Bir+11]. For convenience for our purposes, we have framed this in terms of internal categories, though in [Bir+11] it is more properly framed as about enriched categories more generally. On the other hand, this argument is given in [Bir+11] in a context where the uniqueness of the Löb property has already been presumed, whereas we have noted that this argument can also be given in a context where only the weaker Löb property without uniqueness has been presumed, and then this argument can be used to in fact derive said uniqueness in a lexcategory.

Arguments establishing that the weaker Löb property entails the Löb property with uniqueness in contexts with identity types have been noted in the literature on guarded recursion. For example, as Theorem V.8 in [BM13] and as Theorem 9.5 in [Gra+21]. However, we are unaware of any prior observation in the literature that this uniqueness can also be understood as a special case of the existence of coalgebra-to-algebra hylomorphisms, unifying those arguments.

Theorem 4.9 The identity endofunctor on a cartesian closed category has the Löb property with uniqueness just in case the category is the trivial terminal category.

Proof. Taking $\Box$ to be this identity endofunctor and applying the Löb property with uniqueness to the morphism $\mathrm{id}_{B^A} : \Box(B^A) \to (B^A)$, for arbitrary objects A and B, we find that each B^A has a unique global element, which is to say, there is a unique map between any pair of objects. Thus all objects become isomorphic to the terminal object. ∎

COROLLARY 4.10 The identity endofunctor on a cartesian closed category with equalizers has the Löb property just in case the category is the trivial terminal category.

Proof. By combining Theorem 4.9 and Theorem 4.4. ∎

OBSERVATION 4.11 The example from Observation 4.2 shows that it is possible for the identity endofunctor on a nontrivial cartesian closed category to have the Löb property, so long as neither uniqueness nor equalizers are presumed.

4.3 Lawvere's fixed point theorem

Let us refresh the reader on Lawvere's fixed point theorem [Law69], which captures the general structure of many diagonalization arguments and their relationship to cartesian closed structure. We shall first review a proof of Lawvere's fixed point theorem close in spirit to Lawvere's framing of his result.

Then we will note a slight generalization for which essentially the same argument applies. Then in the next section we will turn this generalization into a result in the context of general pre-introspective theories. Then we will specialize further down to introspective theories, and observe a wonderful "bootstrapping" phenomenon which arises there, which shall ultimately provide us with the Löb property in that context, which is our main result.

THEOREM 4.12 (LAWVERE'S FIXED POINT THEOREM) Let T be an arbitrary category. Let X be an object of T and let Ω be any T-indexed set. Suppose also given some map $App' : X \to \Omega^X$ (equivalent to the data of a map $App : X \times X \to \Omega$).

Let $\star$ be any object of T. By a "point" of a T-indexed set, we shall mean an element of its aspect at $\star$ (equivalent to the data of a map into it from $\star$).

Suppose App has the surjectivity-like property that, for every map $F : X \to \Omega$, there is a point f of X, such that for every point x of X, we have that $App(f, x) = F(x)$.

Then for any map $g : \Omega \to \Omega$, there exists a point ω of Ω such that $\omega = g(\omega)$. That is to say, g has a fixed point.

Proof. Let $F : X \to \Omega$ be the following composition:

$$X \xrightarrow{\langle \mathrm{id}_X, \mathrm{id}_X \rangle} X \times X \xrightarrow{App} \Omega \xrightarrow{g} \Omega$$

That is, for any generalized element x of X, we have that $F(x) = g(App(x, x))$.

We know there exists a point f of X which corresponds with F in the manner of our surjectivity-like supposition on App. Now consider the instance of this surjectivity-like supposition where $x = f$. This tells us that $App(f, f) = F(f)$. But $F(f) = g(App(f, f))$.

Thus, taking $\omega = App(f, f)$, we have that $\omega = g(\omega)$ as desired. ∎

Let us make a few remarks on the scope of generality of this theorem.

Lawvere originally states this theorem specifically for the case where T is a cartesian closed category, but later in [Law69] notes that this implies the theorem just as well for the case where T is merely a category with finite products, as any category can be embedded as a full subcategory of a cartesian closed category in a way which preserves any products or exponentials already present (via the Yoneda embedding). [Law69] does not explicitly consider examples where the original category of interest T lacks finite products, such that $X \times X$ is not an object of T, nor consider taking Ω to be merely a T-indexed set rather than an object of T, but of course these are covered in the same way by the same insight that we can work in $\mathrm{Psh}(T)$ instead of T.

Having observed that we can just as well frame the theorem with any of its objects drawn from $\mathrm{Psh}(T)$ rather than T, the reader might then well wonder why in our framing we have allowed some objects to be in $\mathrm{Psh}(T)$ but still constrained others (such as X) to come from T. We chose this particular framing partly as this is closest to the applications we have in mind, and also partly for what amount to stylistic reasons. In particular, having stated the theorem in this form, interpreting the surjectivity condition on App only requires quantification over the set of morphisms from object X to presheaf Ω (i.e., the set $\Omega(X)$), instead of requiring quantification over the class of natural transformations from a presheaf X to another presheaf Ω (which is potentially a proper class, if T is proper-class-sized). But this is not really of much importance, and again the more general form of the theorem follows readily from the ostensibly less general one.

[Law69] also only states this theorem in the particular case where $\star$ is a terminal object. In general, we can always pass from T to a slice category $T/\star$, and in so doing we will turn what was $\star$-defined data in T into globally defined data in $T/\star$ (a la Theorem 1.8). So constraining $\star$ to be a terminal object does not constrain the theorem excessively. However, it does constrain the theorem slightly, in that interpreting the surjectivity precondition in $T/\star$ in this way results in a stronger (that is, less often satisfied) surjectivity precondition than in the more flexible framing of the theorem we have given: The surjectivity condition in $T/\star$ would amount to requiring that for every $F : \star \times X \to \Omega$ in $\mathrm{Psh}(T)$, we could find a corresponding f. However, we have only required surjectivity with respect to the more constrained set of $F : X \to \Omega$ in $\mathrm{Psh}(T)$.

We do not actually need this extra flexibility for proving our main result. For our purposes, just like Lawvere's, it would suffice to always take $\star$ to be a terminal object. But we note the availability of this flexibility all the same (if only for the purpose of comparison at the end of this chapter to other variants on Lawvere's fixed point theorem recently noted in the literature, such as Magmoidal Fixed Point Theorem (Theorem 4.25)).

Even this loosened surjectivity presumption is still far overkill as far as the needs of the argument go. All that really matters is for one specific definable value to be in the range of App'. But in general practice and for our particular purposes, this is always established because of some such surjectivity condition anyway, so that seems the most useful framing in which to give the theorem.

Having said all that about the wide applicability of Lawvere's Fixed Point Theorem

(Theorem 4.12), we actually will need to generalize it slightly further for our purposes. Having given the above discussion of the traditional theorem to prime the reader's intuitions through familiarity, we now put forward the following simple generalization:

Theorem 4.13 (Self-Related Point Theorem) Let T be an arbitrary category. Let $\star$ and X be objects of T and let Ω be any T-indexed set. Suppose also given some map $App' : X \to \Omega^X$ (equivalent to the data of a map $App : X \times X \to \Omega$).

As before, we shall use "point of" as shorthand for "element of the $\star$-aspect of".

Suppose also given a binary relation R on the points of Ω. (We needn't presume R to be symmetric or transitive or any such thing.). And suppose App has the surjectivity-like property that, for every morphism $F : X \to \Omega$, there is a point f of X, such that for every point x of X, we have $R(App(f, x), F(x))$.

Then there exists a point ω of Ω such that $R(\omega, \omega)$. That is to say, R has a self-related point.

Proof. Let $F : X \to \Omega$ be the following composition:

$$X \xrightarrow{\langle \mathrm{id}_X, \mathrm{id}_X \rangle} X \times X \xrightarrow{App} \Omega$$

That is, for any generalized element x of X, we have that $F(x) = App(x, x)$.

We know there exists a point f of X in accordance with our surjectivity-like supposition on App'. Now consider the instance of the surjectivity-like supposition where $x = f$. This tells us that $R(App(f, f), F(f))$. But $F(f) = App(f, f)$.

Thus, we have found a point of Ω which is related to itself by R, as desired. ■

It may not be obvious that this generalizes Lawvere's Fixed Point Theorem (Theorem 4.12). The following shows how this is so:

Corollary 4.14 (Relatedly-Fixed Point Theorem) Consider the same setup as of Self-Related Point Theorem (Theorem 4.13), and furthermore, suppose given $g : \Omega \to \Omega$.

Then there exists a point ω of Ω such that $R(\omega, g(\omega))$. We might describe this as "ω is an R-fixed point of g".

Proof. Consider the binary relation R_g on points of Ω given by $R_g(\omega_1, \omega_2) = R(\omega_1, g(\omega_2))$.

We have been given the supposition that, for every morphism $F : X \to \Omega$, there is a point f of X, such that for every point x of X, we have $R(App(f, x), F(x))$.

As this holds for arbitrary $F : X \to \Omega$, this also holds when an arbitrary F is replaced by $g \circ F : X \to \Omega$. That is to say, for every $F : X \to \Omega$, there is a point f of X, such that for every point x of X, we have $R(App(f, x), (g \circ F)(x))$, which is to say, $R_g(App(f, x), F(x))$.

But this is precisely the surjectivity supposition we need in order to invoke Self-Related Point Theorem (Theorem 4.13) with R_g in place of R. Doing so, we obtain a point ω of Ω such that $R_g(\omega, \omega)$, which is to say $R(\omega, g(\omega))$, as desired. ■

Now we can see that Lawvere's Fixed Point Theorem (Theorem 4.12) is of course the instance of Relatedly-Fixed Point Theorem (Corollary 4.14) where the relation R is taken to be equality. But Relatedly-Fixed Point Theorem (Corollary 4.14) is strictly more general in allowing the use of an arbitrary relation.

(As for the relation between Relatedly-Fixed Point Theorem (Corollary 4.14) and Self-Related Point Theorem (Theorem 4.13), each is an instance of the other. We above obtained Relatedly-Fixed Point Theorem (Corollary 4.14) as a corollary of Self-Related Point Theorem (Theorem 4.13). But also conversely, Self-Related Point Theorem (Theorem 4.13) is the special case of Relatedly-Fixed Point Theorem (Corollary 4.14) where g is taken to be id_Ω.)

At any rate, we shall find the added flexibility of allowing a relation in place of equality to be valuable in the next sections, as we begin to specialize towards our application in introspective theories.

4.4 Presheaf diagonalization for pre-introspective theories

Theorem 4.15 (Pre-introspective Diagonalization) Let $\langle T, C, \mathcal{S}, \mathcal{N} \rangle$ be a pre-introspective theory. Let $\star_T$ be the terminal object of T and let $\star_C$ be the terminal object of C.[1] Throughout the following, we use "point of" as shorthand for "element of the $\star_T$-aspect of".

Furthermore, let P be a (T, C)-indexed set, in the sense of Double or multiple indexing (Section 1.6). We will write in the following $P(c)$ to mean the T-indexed set $t \mapsto P(t, c)$, for globally defined objects c of C.

Suppose also given some object $\Omega \in T$ with a map $\mathcal{Q} : \Omega \to P(\star_C)$ such that the induced function $\mathcal{Q} \circ - : \mathrm{Hom}(X \times X, \Omega) \to \mathrm{Hom}(X \times X, P(\star_C))$ is surjective.

Suppose also given some object $X \in T$ and map $\alpha : X \to P(\mathcal{S}(X))$. We also make a surjectivity-like assumption on α. Specifically, we suppose that for every global element p of $P(\mathcal{S}(X))$, there is a point x of X such that $\alpha(x) = p$, as points of $P(\mathcal{S}(X))$.

Finally, let g be a globally defined element of $P(\mathcal{S}(\Omega))$.

Then we obtain a point ω of Ω, such that $\mathcal{Q}(\omega) = \mathcal{N}_\Omega(\omega)^* g$.

Proof. We shall show how this is an instance of Self-Related Point Theorem (Theorem 4.13).

We define $App : X \times X \to \Omega$ like so: Consider the two projection maps $\pi_1, \pi_2 : X \times X \to X$, as the two generic $(X \times X)$-defined elements of X. We thus obtain also $(X \times X)$-defined elements $\alpha(\pi_1)$ of $P(\mathcal{S}(X))$ and $\mathcal{N}_X(\pi_2)$ of $\mathrm{Hom}_C(\star_C, \mathcal{S}(X))$. Combining these via the presheaf action of P, we get $(\mathcal{N}_X(\pi_2))^*(\alpha(\pi_1))$ as an $(X \times X)$-defined element of $P(\star_C)$. By the surjectivity presumption on $\mathcal{Q}$, we find a preimage of this under the action of $\mathcal{Q} : \Omega \to P(\star_C)$. We take this preimage to be our $App : X \times X \to \Omega$.

[1]We use this $\star$ notation rather than 1 notation so that we can make the observation that this theorem's proof actually applies more generally, not depending on any limit structure. It would suffice to let T be any category, let C be a T-indexed category, let $\mathcal{S}$ be a functor from T to the global aspect of C, let $\mathcal{N}$ be a map from t to $\mathrm{Hom}_C(\star_C, \mathcal{S}(t))$, natural in $t \in T$, let $\star_T$ be any object of T, and let $\star_C$ be any globally defined object of C. Knowing that the proof makes no use of limit structure may make it easier to follow.

Thus, for any generalized elements x_1, x_2 of X with the same domain, we have that $\mathcal{Q}(App(x_1, x_2)) = (\mathcal{N}_X(x_2))^*(\alpha(x_1))$.

We must now establish an appropriate surjectivity supposition on App for invoking Self-Related Point Theorem (Theorem 4.13).

Let an arbitrary $F : X \to \Omega$ be given. We then have that $\mathcal{S}(F) : \mathcal{S}(X) \to \mathcal{S}(\Omega)$ in the global aspect of C. We can apply the action of P along this morphism to g (a global element of $P(\mathcal{S}(\Omega))$), thus obtaining a global element $\mathcal{S}(F)^*g$ of $P(\mathcal{S}(X))$. By the surjectivity-like assumption on α we made, we now have a corresponding point f of X, such that $\alpha(f) = \mathcal{S}(F)^*g$ (the right side here having been reinterpreted from a global element into a point).

It follows that for every point x of X, we have that $\mathcal{N}_X(x)^*\alpha(f) = \mathcal{N}_X(x)^*\mathcal{S}(F)^*g$.

Note that by the definition of App, we have that $\mathcal{Q}(App(f, x)) = (\mathcal{N}_X(x))^*\alpha(f)$.

Also note that by $\mathcal{S}$ With $\mathcal{N}$ (Lemma 2.24), we have that $\mathcal{S}(F) \circ_C \mathcal{N}_X(x) = \mathcal{N}_\Omega(F(x))$. Thus, by the functoriality of P, we have that $\mathcal{N}_X(x)^*\mathcal{S}(F)^*g = \mathcal{N}_\Omega(F(x))^*g$.

Combining these last three paragraphs, we have that $\mathcal{Q}(App(f, x)) = \mathcal{N}_\Omega(F(x))^*g$.

If we define the relation $R(\omega_1, \omega_2)$ as the equation $\mathcal{Q}(\omega_1) = \mathcal{N}_\Omega(\omega_2)^*g$ accordingly, we have now established the surjectivity supposition required in order to invoke Self-Related Point Theorem (Theorem 4.13). From this invocation, we get a point of Ω which is related by R to itself, which is just what we desired, completing the proof. ■

COROLLARY 4.16 In many cases we are interested in (though not all!), we furthermore take $P(\star_C)$ to be T-representable and take Ω to be $P(\star_C)$, with $\mathcal{Q} : \Omega \to P(\star_C)$ as the identity map between these. We then automatically have that the aspect of $\mathcal{Q}$ at $X \times X$ is surjective as required.

THEOREM 4.17 Suppose given a locally introspective theory $\langle T, C, \mathcal{S}, \mathcal{N} \rangle$ and an object P in the global aspect of $\mathrm{Psh}(C)$.

If there is any object X of T with an isomorphism from X to $P(\mathcal{S}(X))$, then, within the global aspect of $\mathrm{Psh}(C)$, for every $g : \Box P \to P$, we obtain an $\omega : 1 \to P$, such that $\omega = g \circ \omega'$, where $\omega' = \Box_{\mathrm{Psh}(C)}(\omega) : 1 \to \Box P$. In other words, we obtain the instance of the Löb property constrained specifically to P.

We get the same result also if, within $\mathrm{Glob}(C)$, there is any object Y along with an isomorphism from Y to $\mathcal{S}(P(Y))$.

Proof. Any isomorphism $\alpha : X \to P(\mathcal{S}(X))$ (or even just a retraction) will automatically satisfy the surjectivity-like precondition allowing us to invoke Pre-introspective Diagonalization (Theorem 4.15) via Corollary 4.16, which takes Ω as $P(\star_C)$ and $\mathcal{Q}$ as identity. Everything follows immediately from this, but has just been Yoneda-ized in its phrasing.

Specifically, keep in mind via the Yoneda lemma that the data of a map from $c \in C$ to $P \in \mathrm{Psh}(C)$ is the same as an element of $P(c)$. In this way, our $g : \Box P \to P$ can be seen as indeed an element of $P(\Box P) = P(\mathcal{S}(P(\star_C))) = P(\mathcal{S}(\Omega))$, as required.

The invocation of Pre-introspective Diagonalization (Theorem 4.15) via Corollary 4.16 will give us a global element ω of $\Omega = P(\star_C)$ such that $\omega = \omega'^* g$, where $\omega' = \mathcal{N}_\Omega(\omega)$ is a global element of $\Box P(\star_C)$. Again, by the Yoneda lemma, such an ω corresponds to a map from 1 to P in the global aspect of $\mathrm{Psh}(C)$, such an ω' corresponds to a map from 1 to $\Box P$ (specifically, $\omega' = \Box\omega$, by $\mathcal{S}$ Matches $\mathcal{N}$ (Lemma 2.25)), and our equation relating ω and ω' is that that $\omega = g \circ \omega'$.

For the last remark about starting from a fixed point for $\mathcal{S}(P(-))$ rather than a fixed point for $P(\mathcal{S}(-))$, observe that if we have a Y isomorphic to $\mathcal{S}(P(Y))$, then by taking X to be $P(Y)$, we obtain an X isomorphic to $P(\mathcal{S}(X))$.[2]

■

4.5 Bootstrapping to Löb's theorem for introspective theories

This last theorem gives us an instance of the Löb property, but comes with the precondition of a certain isomorphism.

Incredibly, we can bootstrap away this isomorphism precondition, in the context of an introspective theory. That is, in the context of an introspective theory, we can use one particular instance of Pre-introspective Diagonalization (Theorem 4.15) itself to provide the very isomorphisms necessary in order to then re-invoke Pre-introspective Diagonalization (Theorem 4.15) via Theorem 4.17.

Our plan is to consider the (T, C)-indexed set P such that $P(t, c)$ is the set of isomorphism classes of $C(t)/c$, with the action of P on morphisms of C being given by pullback (while the action of P on morphisms of T is given by the reindexing action of the T-indexed category C).[3]

In more detail, for any fixed t and any morphism $m : c_1 \to c_2$ of $C(t)$, the action $P(t, m) : P(t, c_2) \to P(t, c_1)$ is given by pullback in the lexcategory $C(t)$ along m; that is, this is given by $m^* : C(t)/c_2 \to C(t)/c_1$ considered as taking isomorphism classes of objects to isomorphism classes of objects. Note that this reindexing along morphisms in C is indeed strictly functorial, because we are working with isomorphism classes of objects rather than with objects simpliciter.

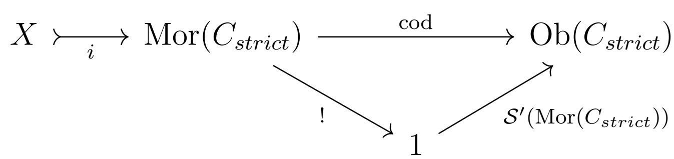
We now choose any internal category C_{strict} in T which presents C (by definition, such an internal category exists in an introspective theory; there may be multiple non-isomorphic such internal categories presenting C, but any will do for our purposes) and we take Ω to be $\mathrm{Ob}(C_{strict})$, with $\mathcal{Q} : \Omega \to P(\star_C)$ sending each object of each aspect of C_{strict} to its isomorphism class within the corresponding aspect of C. Note that every component of $\mathcal{Q}$

[2] This is a special case of the bijective correspondence between fixed points of cyclic rearrangements of compositions, which we also observed within the proof of Theorem 4.3.

[3] Note that this (T, C)-indexed set P is NOT presumed to be T-representable! Indeed, we cannot generally hope for this, as we do not presume T to have any regularity or exactness properties such that we could carry out internal to T such quotienting constructions as would yield the object of isomorphism classes of C.

as a natural transformation between presheaves on T is surjective (because C is presented by C_{strict}, the isomorphism classes of C and of C_{strict} are the same, and there is clearly a surjection from the objects of C_{strict} (at any aspect) onto the isomorphism classes of C_{strict} (at the same aspect)). Thus in particular the component of $\mathcal{Q}$ at the object $X \times X$ of T is surjective. .

We take X to be the subobject of $\mathrm{Mor}(C_{strict})$ comprising those morphisms whose codomain is $\mathcal{S}(\mathrm{Mor}(C_{strict}))$. That is, the object given by the following equalizer diagram.

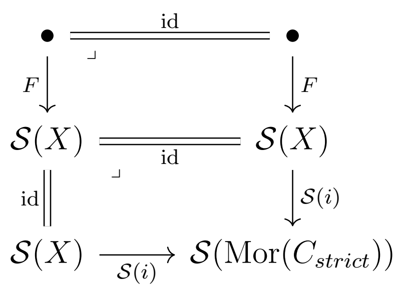


In the above diagram, we have labelled an arrow with the name $\mathcal{S}'(\mathrm{Mor}(C_{strict}))$. By this we mean some arbitrary globally defined object of C_{strict} which presents the globally defined object $\mathcal{S}(\mathrm{Mor}(C_{strict}))$ of C. We pedantically caution that there may actually be multiple non-equal global elements of $\mathrm{Ob}(C_{strict})$ which present objects isomorphic to $\mathcal{S}(\mathrm{Mor}(C_{strict}))$. But any arbitrary choice of some such element will be fine to use as the arrow in this diagram for our purposes.[4]

Note that, by virtue of being an equalizer, the inclusion map $i : X \to \mathrm{Mor}(C_{strict})$ in T is monic, and thus (as $\mathcal{S}$ is a lexfunctor) so also is $\mathcal{S}(i) : \mathcal{S}(X) \to \mathcal{S}(\mathrm{Mor}(C_{strict}))$ in C. From this, we can define our $\alpha : X \to P(\mathcal{S}(X))$ and establish its surjectivity condition. Specifically, observe that pullback along $\mathcal{S}(i)$ gives us a functor $\mathcal{S}(i)^* : C/\mathcal{S}(\mathrm{Mor}(C_{strict})) \to C/\mathcal{S}(X)$. If we focus on the action of $\mathcal{S}(i)^*$ on objects, consider its input object as presented by an object of $C_{strict}/\mathcal{S}'(\mathrm{Mor}(C_{strict}))$ (whose objects comprise X), and consider its output object modulo isomorphism, this yields $\mathcal{S}(i)^* : X \to P(\mathcal{S}(X))$, which we take as our definition of α.

As for the surjectivity condition, let F be an arbitrary global element of $P(\mathcal{S}(X))$; that is, an arbitrary isomorphism class of objects of $C/\mathcal{S}(X)$. The pushforward (i.e., composition) action of $\mathcal{S}(i)$ gives us a functor from $C/\mathcal{S}(X) \to C/\mathcal{S}(\mathrm{Mor}(C_{strict}))$, taking F to $\mathcal{S}(i) \circ F$, an isomorphism class of objects of the global aspect of $C/\mathcal{S}(\mathrm{Mor}(C_{strict}))$. This will be presented by at least one globally defined element f of X (keeping in mind the definition of X); there may be multiple non-equal such f but any will do. Observe that $\alpha(f)$ is the isomorphism class of $C/\mathcal{S}(X)$ corresponding to $\mathcal{S}(i) \circ F$ pulled back along $\mathcal{S}(i)$. This isomorphism class is the same as that of F itself, because of the monicity of $\mathcal{S}(i)$, like so:

[4] Indeed, it is readily seen that even two non-equal such choices will still lead to isomorphic Xes. Or more precisely, isomorphic results as an object of T, though not isomorphic as a subobject of $\mathrm{Mor}(C_{strict})$, as the specific choice of inclusion map $i : X \to \mathrm{Mor}(C_{strict})$ will vary. But again, any so-arising choice will be fine for our purposes.



Thus, $\alpha(f) = F$ as an element of $P(\mathcal{S}(X))$, establishing the required surjectivity condition on α.

Thus, all presumptions are satisfied for us to be able to apply Pre-introspective Diagonalization (Theorem 4.15) with these definitions, for an arbitrary globally defined element g of $P(\mathcal{S}(\Omega))$.

In particular, let G be an arbitrary globally defined object of $\mathrm{Psh}(C)$. (In fact, it suffices for G merely to have reindexing along isomorphisms rather than arbitrary morphisms of C; that is, for G to be an object of $\mathrm{Psh}(\mathrm{core}(C))$, where $\mathrm{core}(C)$ is the subcategory of C containing just its invertible morphisms.)

This will be presented by an object of $T/\operatorname{Ob}(C_{strict})$ (the map into $\operatorname{Ob}(C_{strict})$ whose fiber at any object c_{strict} of C_{strict} is the set $G(c)$, where c is the object of C presented by c_{strict}). By applying $\mathcal{S}$ to this, we get a globally defined object of $C/\mathcal{S}(\operatorname{Ob}(C_{strict}))$, which is to say, a global element of $P(\mathcal{S}(\Omega))$. Take this to be our g.

Invoking Pre-introspective Diagonalization (Theorem 4.15) (on the introspective theory $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$, with all other inputs (P, Ω, $\mathcal{Q}$, X, α, and g) as described with the same name above), we now get a globally defined element ω of $\Omega = \operatorname{Ob}(C_{strict})$ such that $\mathcal{Q}(\omega) = \mathcal{N}_\Omega(\omega)^* g$. This equation is saying precisely that ω presents an object Y of C such that Y is isomorphic to $\mathcal{S}(G(Y))$.

Thus, we have proven the following:

Theorem 4.18 For any introspective theory $\langle T, C\rangle$, and any globally defined object G of $\mathrm{Psh}(C)$, or even of $\mathrm{Psh}(\mathrm{core}(C))$, there is some object $Y \in \mathrm{Glob}(C)$ along with an isomorphism from Y to $\mathcal{S}(G(Y))$.

Combining this with Theorem 4.17 to eliminate the latter's isomorphism precondition, we now reach the following conclusion:

Theorem 4.19 (Löb's Theorem for Introspective Theories) Suppose given an introspective theory $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$.

Then, within $\mathrm{Glob}(\mathrm{Psh}(C))$, for every object P and morphism $g : \Box P \to P$, we obtain an $\omega : 1 \to P$, such that $g \circ (\Box\omega) = \omega$.

In other words, the global aspect of $\Box_{\mathrm{Psh}(C)}$ has the Löb property. Keeping in mind the equivalences of Theorem 4.5, we may conclude that the global aspects of $\Box_{T/-}$, $\Box_C$, and $\Box_{\mathrm{Psh}(C)}$ all have the Löb property with uniqueness.

OBSERVATION 4.20 We can consider the particular case where P is C-representable, just as $\Box P$ is. In other words, where $P(-) = \mathrm{Hom}_C(-, c)$ is the representable presheaf on C represented by some object c of C. All traditional accounts of Löb's theorem are along these lines. But note that we can also just as well consider this Löb's Theorem for Introspective Theories (Theorem 4.19) for non-representable presheaves P, a significant generalization of the traditional viewpoint.

COROLLARY 4.21 For any introspective theory $\langle T, C\rangle$, every aspect of $\Box_T$, $\Box_C$, and $\Box_{\mathrm{Psh}(C)}$ has the Löb property with uniqueness.

In other words, every introspective theory has the Löb property, in the terminology of Theorem 4.5.

Proof. By Observation 3.1, each aspect of any of these $\Box$ functors is the global aspect of the corresponding $\Box$ functor on the corresponding slice introspective theory. Thus, we simply invoke Löb's Theorem for Introspective Theories (Theorem 4.19) on this slice introspective theory. ■

The above is our key result. The fact that the simple definition of introspective theories is enough to lead to their satisfying the Löb property with uniqueness motivates much of our interest in the concept of introspective theories.

OBSERVATION 4.22 The fixed points produced by Theorem 4.18 are furthermore unique up to canonical isomorphism, by combining Löb's Theorem for Introspective Theories (Theorem 4.19) with Corollary 4.8.

4.6 The self-indexing cannot be representable, except trivially

We note an important corollary of the above:

THEOREM 4.23 Let T be any lexcategory, and equip it as an introspective theory $\langle T, C, \mathcal{F}\rangle = \langle T, T/-, \mathrm{id}\rangle$ by taking C to be T's self-indexing and $\mathcal{F}$ to be the identity (a la Example 2.3). Recall from Theorem 1.32 that this will be locally introspective (that is, the self-indexing will be locally representable) precisely when T is locally cartesian closed.

This will furthermore be introspective (that is, the self-indexing will be representable) only when T is the trivial terminal category.

Proof. For a lexcategory T equipped as a pre-introspective theory in this way, the operation $\Box_T$ acts as the identity.

And by Löb's Theorem for Introspective Theories (Theorem 4.19), if T is an introspective theory, then $\Box_T$ will have the Löb property with uniqueness.

But by Theorem 4.9, the identity endofunctor on a cartesian closed lexcategory has the Löb property with uniqueness only when the category is the trivial terminal category. ■

This "no-go" result was demonstrated in [PT89] by an essentially identical argument to the argument we have given, when the abstractions in our argument are unwound to this special case.

But by generalizing to introspective theories, we are able to expand from this negative result (there are no nontrivial lexcategories whose self-indexing is representable) to a positive result (there are many nontrivial examples of introspective theories, which all end up satisfying the Löb property with uniqueness and all the further corollaries of this noted in The Löb property in abstract (Section 4.2)).

Observation 4.24 From the above, we see that, though the Löb property holds for all introspective theories automatically, it does not hold automatically for merely locally introspective theories (as there are many locally cartesian closed categories which are nontrivial. Counterexamples could also be constructed from non-well-founded transitive relations using Construction 2.14.). However, we have also seen there are some natural examples of locally introspective but not fully introspective theories with the property that arbitrarily loose sub-introspections of them can be made into introspective theories, as in the relationship between our archetypal examples Construction 2.14 and Construction 2.16, or the relationship between our archetypal examples Construction 2.12 and Construction 2.13. Such locally introspective theories will thus inherit the Löb property from their sub-introspections.

4.7 As applied to our archetypal examples

Here we discuss the application of Löb's Theorem for Introspective Theories (Theorem 4.19) to our archetypal examples of introspective theories:

4.7.1 ZF-Finite examples

Recall from Construction 2.10 that we have a natural introspective theory $\langle Z_{\Sigma_1}, Z' \rangle$, where Z_{Σ_1} is the lexcategory of Σ_1-definable hereditarily finite sets and Σ_1-definable functions between them up to provable equivalence in ZF-Finite, and Z' is the lexcategory internal to Z_{Σ_1} of arbitrary definable sets and arbitrary definable functions between them up to provable equivalence in ZF-Finite.

Recall from the discussion at ZF-Finite examples (Section 3.4.1) that the global aspect of Z' can be identified with Z (the actual category of arbitrary definable sets and functions between them in ZF-Finite) and that the $\Box$ operator acts on this by sending each "the object of Xes" to "the object of definitions of Xes within ZF-Finite". In the particular case where the object in question is subterminal (thus representing a proposition), this amounts to the traditional provability operator sending the proposition X to the proposition "There is a proof in ZF-Finite of X".

Thus, as applied to these subterminal objects, our Löb property with uniqueness for this introspective theory is indeed the namesake Löb property of traditional logic: It tells us that if there is a proof that the provability of X entails X, then there is in fact an unconditional proof of X. Gödel's second incompleteness theorem follows as the special case of this where X is a manifest falsehood, and Gödel's first incompleteness theorem then readily follows from the second incompleteness theorem.

But we may consider non-subterminal objects as well, and here our Löb property with uniqueness gives us a form of guarded recursion in the context of such logical theories as ZF-Finite. Specifically, for any definable function from definitions of Xes to actual Xes, there is a unique (up to provable equivalence) definition of an X which is provably equivalent to the given function applied to its own definition.[5]

We are not aware of guarded recursion having been strongly investigated in this context before. We aspire to explore working with this form of guarded recursion further in future work. For now we simply observe it as a vast generalization of the traditional purely propositional interpretation of Löb's theorem in logic.

4.7.2 Kripke frame example

Recall the introspective theory $\langle \mathrm{Psh}'(Q), C' \rangle$ from Construction 2.16, constructed from a well-founded transitive relation $<$ on a set P, with Q being P augmented with a new maximum element ∞ and construed as a preorder category using the $<$ relation. The $\mathrm{Psh}'(Q)$ here is a full sublexcategory of $\mathrm{Psh}(Q)$, defined by certain cardinality constraints, but these cardinality constraints can be taken to be arbitrarily loose such that any set-sized number of particular desired objects of $\mathrm{Psh}(Q)$ can be found within $\mathrm{Psh}'(Q)$.

The global aspect of C' here is a certain full sublexcategory of $\mathrm{Set}^{|P|}$ (again defined by cardinality constraints, which may again be taken to be arbitrarily loose such that any set-sized number of particular desired objects can be found within this). Recall from the discussion at Kripke frame example (Section 3.4.2) that the $\Box$ operator acts on this such that $\Box F(x)$ is the product of $F(y)$ over all $y < x$, where $F \in \mathrm{Set}^{|P|}$ and $x, y \in P$. For subterminal F acting as propositions, this corresponds to the traditional interpretation of the $\Box$ operator in a Kripke frame, such that $\Box F$ is true at a world just in case F is true at all lower worlds.

In this context, the Löb property with uniqueness which we are given by Löb's Theorem for Introspective Theories (Theorem 4.19) tells us that we may define functions by transfinite recursion: Given at each $x \in P$ a function g from $\prod_{y<x} F(y)$ to $F(x)$, we obtain a uniquely determined function G whose domain is P such that each $G(x)$ is given by g applied to the values of G at $y < x$.

In the particular case where F is subterminal representing an proposition (that is, an arbitrary subset of $|P|$), this amounts to the principle of transfinite induction or "strong induction": It tells us that a proposition holds of all of P so long as it holds of any particular $x \in P$ once it holds of all $y < x$.

[5]Here, all mentions of definability and provability are with respect to the particular theory ZF-Finite, though analogous constructions of introspective theories can be carried out for other logical theories as well, such as any computably enumerable extension of ZF-Finite, as we later discuss at Observation 6.15.

Of course, these principles of transfinite recursion/induction over well-founded transitive relations are well-known and easy to establish directly, without all the machinery of introspective theories. (The induction principle here is after all the very defining characteristic of well-foundedness.) But it is remarkable to observe how these phenomena are in this way unified with the phenomena of Löb's theorem in traditional logic (as discussed at ZF-Finite examples (Section 4.7.1)), not just in the form of the Löb property result but in the particular derivation of it as well.

4.7.3 Step-indexing example

The application of our Löb's theorem with uniqueness results to the introspective theory Construction 2.13 corresponding to step-indexing in the topos of trees is similar to the one just discussed. Recall from the discussion at Step-indexing example (Section 3.4.3) that we here have a $\Box$ operator on (an arbitrarily loose full sublexcategory of) $\mathrm{Psh}(\omega)$, where ω is the poset of natural numbers, such that $\Box F(0) = 1$ and $\Box F(n+1) = n$, for $n \in \omega$ and $F \in \mathrm{Psh}(\omega)$.

Our Löb property with uniqueness thus tells us that we may define functions on the natural numbers by the most familiar kind of recursion: Given any specified value at 0, and any specified way to transform a value at n into a value at $n+1$ for each $n \in \omega$, there is a unique function on the natural numbers taking on the specified value at 0 and whose value at each $n+1$ is derived from its value at n in the specified way.

In the particular special case where we are dealing with subterminal objects of $\mathrm{Psh}(\omega)$, these amount to downwards closed subsets of ω, and the above specializes to the principle of ordinary induction for these: Given a downwards closed subset of ω, if it contains 0 and is closed under successor, then it contains all of ω.

Again, all of this is quite familiar and easy to demonstrate directly without any invocation of the machinery of introspective theories (these amount to the characteristic properties of the natural numbers as a natural numbers object within Set). But again, it is remarkable that we can in this way see these as strongly unified with the analogous properties and the derivation of those properties for our other archetypal examples, including the case of ZF-Finite examples (Section 4.7.1) which has no direct relationship to presheaves over a well-founded structure.

4.8 Relating variations on Lawvere's fixed point theorem

Although not important for our main narrative, we note here some further comments on the relation of Lawvere's fixed point theorem to generalizations of ours or others.

First, we observe that Lawvere's Fixed Point Theorem (Theorem 4.12) can be straightforwardly re-obtained as a special case of our Pre-introspective Diagonalization (Theorem 4.15).

Proof. First, we handle the special case of Lawvere's Fixed Point Theorem (Theorem 4.12) where T has finite limits and Ω is an object of T.

This is a special case of Pre-introspective Diagonalization (Theorem 4.15) where we take the pre-introspective theory $\langle T, C, \mathcal{F}\rangle$ to be the trivial one where C is the self-indexing $T/-$ and $\mathcal{F}$ is the identity.

Furthermore, P is taken to be the (T, C)-indexed set represented by Ω; that is, such that $P(t, c) = \mathrm{Hom}_T(t \times c, \Omega)$. Note that $P(\mathcal{S}(t))$ for objects t of T is therefore the T-indexed set Ω^t. In particular, $P(1)$ is thus isomorphic to Ω. As in Corollary 4.16, we can take $\mathcal{Q}$ to be this isomorphism (one can think of it as an identity if one likes), and this will then automatically be surjective on its $X \times X$ aspect.

We take $\alpha : X \to P(\mathcal{S}(X)) = \Omega^X$ to be given by the map $App' : X \to \Omega^X$ presumed in Lawvere's Fixed Point Theorem (Theorem 4.12). The surjectivity presumption from Lawvere's Fixed Point Theorem (Theorem 4.12) then becomes the surjectivity presumption of Pre-introspective Diagonalization (Theorem 4.15).

And to give a g in the global aspect of $P(\mathcal{S}(\Omega)) = \Omega^\Omega$ is precisely the data presumed by the name g in Lawvere's Fixed Point Theorem (Theorem 4.12).

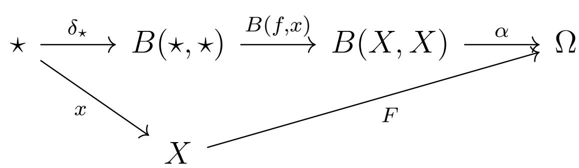
This matches all the presumptions of Pre-introspective Diagonalization (Theorem 4.15) up with corresponding presumptions from Lawvere's Fixed Point Theorem (Theorem 4.12), and the conclusion we then obtain from Pre-introspective Diagonalization (Theorem 4.15) is readily seen to be the same as the conclusion from Lawvere's Fixed Point Theorem (Theorem 4.12).

The above shows how to obtain Lawvere's Fixed Point Theorem (Theorem 4.12) as an instance of Pre-introspective Diagonalization (Theorem 4.15) when T is a lexcategory and Ω is an object of T. We then obtain Lawvere's Fixed Point Theorem (Theorem 4.12) in full (that is, for arbitrary categories T and T-indexed sets Ω) from this special case, by first replacing T with $\mathrm{Psh}(T)$, as noted in our discussion following our presentation of Lawvere's Fixed Point Theorem (Theorem 4.12). ∎

We also note in passing that another interesting generalization of Lawvere's Fixed Point Theorem (Theorem 4.12) was recently remarked upon in [Rob21]. The following (or rather, its contrapositive) was given as Theorem 11 there. We shall present our own proof.

Theorem 4.25 (Magmoidal Fixed Point Theorem) Let T be an arbitrary category with objects $\star$ and Ω, and let $B : T \times T \to T$ be a bifunctor on T such that we have a transformation $\delta_t : t \to B(t, t)$ natural in t from T. As ever, use "point of" to mean "element of the $\star$-aspect of".

Suppose given an object X of T and an $\alpha : B(X, X) \to \Omega$ with the pointwise surjectivity property that for every $F : X \to \Omega$, there is a point f of X, such that for every point x of X, we have that the following diagram commutes:

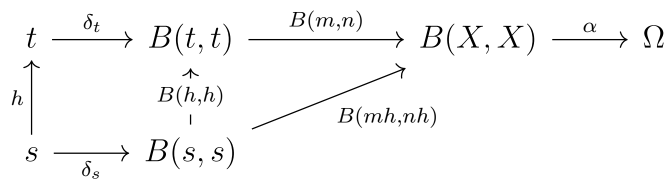


Then for every $g : \Omega \to \Omega$, there is a point ω of Ω such that $\omega = g(\omega)$. That is to say, a fixed point of g.

Proof. Take $App : X \times X \to \Omega$ to be defined like so: For each object t of T, we define App_t : $\mathrm{Hom}(t, X) \times \mathrm{Hom}(t, X) \to \mathrm{Hom}(t, \Omega)$ by giving $App_t(m, n)$ as the following composition:

$$t \xrightarrow{\delta_t} B(t,t) \xrightarrow{B(m,n)} B(X,X) \xrightarrow{\alpha} \Omega$$

That this definition of App_t is natural in t follows from the naturality of δ and the functoriality of B. Specifically, naturality with respect to $h : s \to t$ is seen as follows:



The desired result now follows by Lawvere's Fixed Point Theorem (Theorem 4.12). ■

Lawvere's Fixed Point Theorem (Theorem 4.12) is of course the special case of Magmoidal Fixed Point Theorem (Theorem 4.25) where B is the familiar cartesian product and δ is the familiar diagonal transformation. Thus, in [Rob21], Magmoidal Fixed Point Theorem (Theorem 4.25) is considered as a generalization of Lawvere's fixed point theorem. But as we've just seen, Magmoidal Fixed Point Theorem (Theorem 4.25) is also a special case of Lawvere's fixed point theorem, appropriately construed (as in our formulation of Lawvere's Fixed Point Theorem (Theorem 4.12) which removes the $\star = 1$ constraint), despite the seeming mismatch between general bifunctors and specifically cartesian products. As noted before, there is no need for $X \times X$ to be T-representable, and if such closure of our underlying category is insisted upon, we can just as well always pass to $\mathrm{Psh}(T)$ first.

Chapter 5

Geminal categories

5.1 Preview

In this chapter, we build the machinery to give an explicit yet tractably compact description of the initial introspective theory (which we call the theory of "geminal categories"). This is the key result of this chapter.

We also show the remarkable result that any strict introspective theory can itself be equipped in a natural way as a model of this initial introspective theory; that is, any strict introspective theory can be seen as a geminal category.

(This last statement is easy to misinterpret, so let me be a bit more clear as to what I mean by this. I do not mean the trivial statement that every introspective theory extends the initial introspective theory. Rather, I mean that the theory of strict introspective theories extends the initial introspective theory (even though the theory of strict introspective theories is not itself an introspective theory).)

We will also discuss a partial converse of sorts, a way to extract an introspective theory from a geminal category, with the extracted introspective theory having a certain terminality property (that is, we construct a sort of co-free introspective theory induced by the given geminal category).

This chapter requires some preliminary concepts to be established in Multiply internal structures (Section 5.2) and Strict introspective theories (Section 5.3). The basic definitions concerning geminal categories are then given in Defining geminal categories (Section 5.4) through Compactly defined geminal categories (Section 5.6). After all this machinery has been built, the key result that the theory of geminal categories is in fact the initial introspective theory is ultimately demonstrated in The free introspective theory (Section 5.8). We then discuss co-free constructions in Co-free introspective theories and geminal categories (Section 5.12).

5.2 Multiply internal structures

Before we get to the main material of this chapter, it will be helpful to introduce the concept of "multiply internal" structures, which are used heavily throughout this chapter.

First, a small remark on notation: Recall that if we have a lexfunctor $F : C \to D$ and a structure S internal to C, then we obtain a structure $F(S)$ of the same sort internal to D. Often, we shall write $F[S]$ for this instead of $F(S)$, to emphasize this particular operation as visually distinct from all the other ways in which parentheses can be used.

Definition 5.1 Let C_0 be a lexcategory, and let C_1 be the global aspect of a lexcategory internal to C_0. Now suppose given some structure S internal to C_1. We may say that this structure S is **doubly internal** to C_0.

We may iterate this process. Suppose now that C_2 is the global aspect of some lexcategory internal to C_1, which in turn remains the global aspect of some lexcategory internal to C_0. We can now speak of structures internal to C_2 as being **triply internal** to C_0.

And in general, given a sequence $C_0, C_1, C_2, \dots, C_n$ where each C_{i+1} is the global aspect of a lexcategory internal to C_i, we may speak of structures internal to C_n as being **$(n+1)$-tuply internal** to C_0 (and in the same way n-tuply internal to C_1, $(n-1)$-tuply internal to C_2, and so on). That is, we recursively define an $(n+1)$-tuply internal structure as a structure internal to the global aspect of an n-tuply internal lexcategory, with the base case being that the only 0-tuply internal lexcategory of some C is C itself.

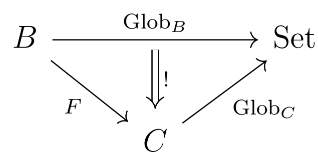
(Multiply internal structures can equivalently be thought of as multiply indexed structures (in the sense of Double or multiple indexing (Section 1.6)) satisfying suitable representability conditions, but they are probably more easily understood in the presentation just given.)

Definition 5.2 Observe that whenever C is a lexcategory and D is a C-indexed locally representable lexcategory, the global sections functor $\mathrm{Hom}_D(1, -)$ can be seen as an indexed lexfunctor from D to the self-indexing $C/-$; in particular, the global aspect of this lets us see $\mathrm{Hom}_D(1, -)$ as a lexfunctor from the global aspect of D to C itself. Let us write $\Gamma_D : \mathrm{Glob}(D) \to C$ to refer to this last lexfunctor, or drop the subscript and write simply Γ where there is no need to disambiguate which D we are referencing. (In particular, when writing $\Gamma[S]$ with no subscript on the Γ, we always mean $\Gamma_X[S]$ where S is singly internal to X, though X may in turn be internal or multiply internal to some other category.)

Thus, if S is some structure internal to the global aspect of D, we find that $\Gamma_D[S]$ is a structure of the same sort internal to C. In this way, any doubly-internal structure S yields a singly-internal structure $\Gamma[S]$, and more generally, any $(n+1)$-tuply internal structure S yields an n-tuply internal structure $\Gamma[S]$.

Note that any lexcategory C can also be thought of as a lexcategory internal to Set, and thus Γ_C in this instance is the same as $\mathrm{Glob}(-) : C \to \mathrm{Set}$. In this case, we may write Glob_C for this map, to emphasize that we are specifically dealing with a global sections lexfunctor whose domain is C and whose codomain is Set.

Definition 5.3 Recall Lemma 1.58, which tells us that, for any lexcategory B, the global sections functor Glob_B is initial among all lexfunctors from B to Set. Thus, for any lexfunctor $F : B \to C$, we obtain a unique natural transformation as in the following diagram:

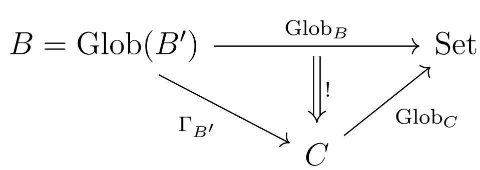


In this way, for any B-internal structure S, we obtain a homomorphism from $\mathrm{Glob}_B(S)$ to $\mathrm{Glob}_C(F[S])$. We refer to this homomorphism as $\mathrm{Induced}(F, S)$.

More explicitly, $\mathrm{Induced}(F, S)$ is the action of the functor F taking each $x \in \mathrm{Hom}_B(1, s)$ to $F(x) \in \mathrm{Hom}_C(F(1), F(s)) = \mathrm{Hom}_C(1, F(s))$, where s is any object of the diagram in B corresponding to the structure S.

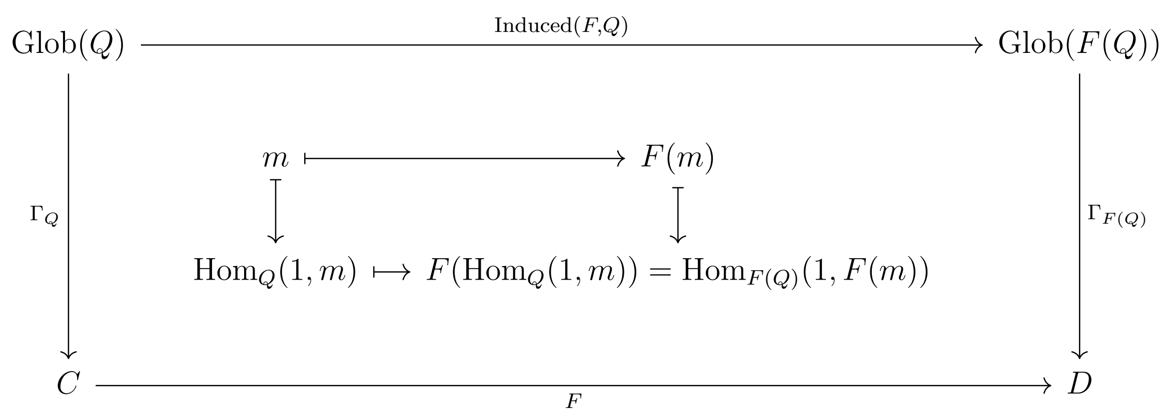
This process can be carried out in the internal logic of a lexcategory as well. That is, if $F : B \to C$ is an internal lexfunctor between V-internal lexcategories, and S is some structure internal to the global aspect of B (thus doubly internal to V), we get a V-internal homomorphism $\mathrm{Induced}(F, S) : \Gamma_B[S] \to \Gamma_C[F[S]]$ in the same way. (Note that $F[S]$ here, the application of an internal lexfunctor $F : B \to C$ to a structure in the global aspect of B, is the same as what could also be called $\mathrm{Glob}(F)[S]$ where $\mathrm{Glob}(F) : \mathrm{Glob}(B) \to \mathrm{Glob}(C)$.)

Observation 5.4 If C is a lexcategory and B is the global aspect of some C-indexed locally representable lexcategory B', then $\mathrm{Glob}_B(-) = \mathrm{Hom}_{\mathrm{Glob}_C(B')}(1, -) = \mathrm{Hom}_C(1, \mathrm{Hom}_{B'}(1, -))$ $= \mathrm{Glob}_C(\Gamma_{B'}(-))$. Thus, Glob_B and $\mathrm{Glob}_C \circ \Gamma_{B'}$ are isomorphic. As the former is initial among lexfunctors from B to Set, so is the latter, and thus in this case the natural transformation described in Definition 5.3 becomes an isomorphism:



That is to say, $\mathrm{Induced}(\Gamma_{B'}, S) : \mathrm{Glob}(S) \to \mathrm{Glob}(\Gamma_{B'}[S])$ is always an isomorphism.

Lemma 5.5 If $F : C \to D$ is a strict lexfunctor, and Q is a C-internal lexcategory, then $F \circ \Gamma_Q = \Gamma_{F(Q)} \circ \mathrm{Induced}(F, Q)$. That is to say, the following outer diagram commutes, as evidenced by the inner chase of an arbitrary datum m in $\mathrm{Glob}(Q)$:



Definition 5.6 Note that any structure S which is n-tuply internal to a lexcategory C (for $n \geq 1$) is ultimately described by some kind of diagram within C, and thus taken by a lexfunctor $F : C \to D$ to a structure of the same sort n-tuply internal to D as well. It is natural to refer to this as $F[S]$ in the same way as for singly internal S.

This operation $F[S]$ for multiply internal S can be inductively understood like so: The base case is when S is singly internal to the domain of F, in which case $F[S]$ is just the ordinary application of F to yield a structure singly internal to the codomain of F. On the other hand, if S is n-tuply internal to $\mathrm{dom}(F)$ for $n \geq 2$, then there is some $\mathrm{dom}(F)$-internal lexcategory B' such that S is $(n-1)$-tuply internal to $\mathrm{Glob}(B')$. In this case, we have also the lexfunctor $\mathrm{Induced}(F, B') : \mathrm{Glob}(B') \to \mathrm{Glob}(F[B'])$, and thus we can understand $F[S]$ as meaning $\mathrm{Induced}(F, B')[S]$, reducing us from the n-tuply internal case to the $(n-1)$-tuply internal case.

Note 5.7 Our notation for dealing with switching internality levels can sometimes cause expressions simultaneously involving structures at different levels of internality to get pretty cluttered. We recommend that readers first read such expressions and diagrams treating all instances of $\Gamma[X]$ or $\mathrm{Glob}(X)$ as simply saying X, and treating all instances of $\mathrm{Induced}(F, S)$ as simplying saying F, to understand the gists of these expressions. A formal account of how to rigorously reason using this less verbose shorthand can be given, but we save such an account of terser notation for future work.

It may be helpful to keep in mind that when $F : A \to B$ is a map, then $\Gamma[F] : \Gamma[A] \to \Gamma[B]$, $\mathrm{Glob}(F) : \mathrm{Glob}(A) \to \mathrm{Glob}(B)$, or $\mathrm{Induced}(F, S) : \mathrm{Glob}(S) \to \mathrm{Glob}(F[S])$ do not change the action of the map F, per se, but rather merely restrict its domain and then restrict its codomain accordingly ($\Gamma[F]$ or $\mathrm{Glob}(F)$ restrict F to just its action on global elements rather than elements in arbitrary aspects of its domain, while $\mathrm{Induced}(F, S)$ restricts a lexfunctor F to just its action on global elements of the objects used in S, rather than the action of F on elements in arbitrary aspects of arbitrary objects in its domain).

5.3 Strict introspective theories

It will be technically convenient for us to work in this chapter with a slightly less "presentation-free" variant of our notion of introspective theories.

Definition 5.8 A **strict introspective theory** is a strict lexcategory T, a lexcategory C internal to T, a strict lexfunctor $\mathcal{S}$ from T to the global aspect of C, and a natural transformation $\mathcal{N}$ from id_T to $\mathrm{Hom}_C(1, \mathcal{S}(-))$.

As usual, to name a strict introspective theory, we can enumerate the entire ordered tuple $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$, or sometimes we just note $\langle T, C\rangle$ or T explicitly and leave the rest implicit.

The definition of a strict introspective theory differs from the definition of an ordinary introspective theory (Definition 2.7) in the following ways: T is made strict (thus, its internal structures can be considered up to equality instead of mere isomorphism), we demand the selection of C as a particular T-internal lexcategory up to equality (instead of simply up to presenting equivalent indexed categories), and we take $\mathcal{S}$ as a strict lexfunctor (thus, $\mathcal{S}$ preserves chosen basis limits on-the-nose).

Clearly, any strict introspective theory presents some introspective theory. Conversely, we have the following:

Theorem 5.9 Any introspective theory $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$ is presented by some strict introspective theory.

Proof. Suppose given an introspective theory $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$. By definition of the representability of C, we can choose some lexcategory C_{int} internal to T which presents the T-indexed category C. (That is, even though C itself is only specified up to equivalence of indexed categories, we can choose a specific presentation of it by a representable indexed strict category C_{int} which is specified more fine-grainedly up to isomorphism of indexed strict categories.)

Now using Theorem 1.27, let T_{strict} be some strict lexcategory which presents T and which has the freeness property that any lexfunctor from T to a strict lexcategory L is presented by some strict lexfunctor from T_{strict} to L. Because T_{strict} presents T, we can choose some specific internal lexcategory C_{strict} in T_{strict} (this C_{strict} being specified up to equality!) which presents C_{int}. Because C_{strict} presents C_{int} which in turn presents C, $\mathcal{S}$ can be viewed as a (non-strict) lexfunctor from T to the global aspect of C_{strict}. Now using the freeness property of T_{strict}, we obtain a strict lexfunctor $\mathcal{S}_{strict}$ from T_{strict} to the global aspect of C_{strict}, such that $\mathcal{S}_{strict}$ presents $\mathcal{S}$.

Finally, we deal with $\mathcal{N}$. Natural transformations are essentially unaffected by strictness considerations. That is, given parallel strict functors A_{strict} and B_{strict}, natural transformations between these are in bijection with natural transformations between the non-strict functors these present. So our original $\mathcal{N}$ corresponds to a unique natural transformation between the identity on T_{strict} and $\mathrm{Hom}_{C_{strict}}(1, \mathcal{S}_{strict}(-))$.

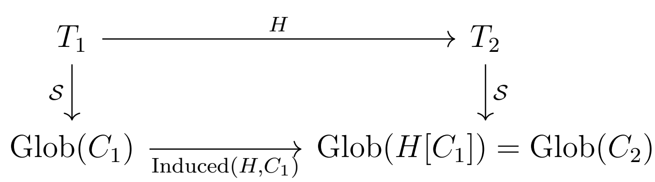
Thus, we have obtained a strict introspective theory $\langle T_{strict}, C_{strict}, \mathcal{S}_{strict}, \mathcal{N}\rangle$ presenting the introspective theory $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$. ■

Strict introspective theories are slightly more convenient than introspective theories for phrasing the results of this chapter, because strict introspective theories are themselves an essentially algebraic notion. That is, there is an essentially algebraic theory such that the models of this theory are the strict introspective theories. (This is in precisely the same way that the theory of strict categories is essentially algebraic, while the theory of categories construed up to equivalence is not quite essentially algebraic.)

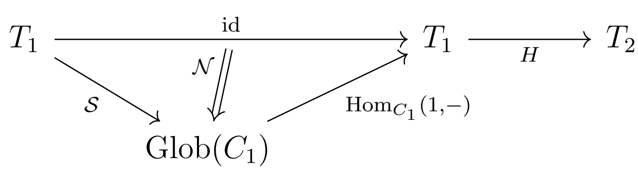
As with any essentially algebraic theory, we get automatically a corresponding notion of homomorphism.

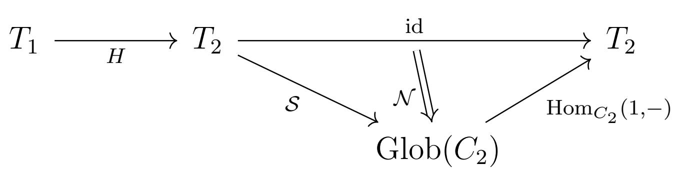
Definition 5.10 A **homomorphism** between strict introspective theories $\langle T_1, C_1, \mathcal{S}, \mathcal{N}\rangle$ and $\langle T_2, C_2, \mathcal{S}, \mathcal{N}\rangle$ is a strict lexfunctor $H : T_1 \to T_2$ such that $H[C_1] = C_2$, and $\mathrm{Induced}(H, C_1) \circ \mathcal{S} = \mathcal{S} \circ H$, and $H[\mathcal{N}_t] = \mathcal{N}_{H(t)}$ for each object t of T_1.

The condition relating H to $\mathcal{S}$ is illustrated like so:



The condition relating H to $\mathcal{N}$ is that the following two natural transformations are equal:





That the codomains of these two natural transformations are equal follows from the previous conditions.

Such homomorphisms are closed under composition and thus we obtain the category of strict introspective theories.

As the category of models of an essentially algebraic theory, this category must have an initial object. That is, there is a strict introspective theory with a unique homomorphism into any other strict introspective theory. In this chapter, we will find a tractable explicit description of this initial strict introspective theory.

5.4 Defining geminal categories

We will give two different presentations of the definition of "geminal categories". First, in this section, we give a definition using several infinite sequences of data and of equations. These infinite sequences will be highly redundant in that their first few entries suffice to derive all their later entries, but the advantage of this verbose definition is that it is made manifest how the definition contains a nested copy of itself.[1] Later, at Geminal category, compact presentation (Definition 5.15), we will see a much more compact definition eliminating these redundancies.

Definition 5.11 (Geminal category) A **geminal category**[2] internal to lexcategory C_0 consists of several ingredients:

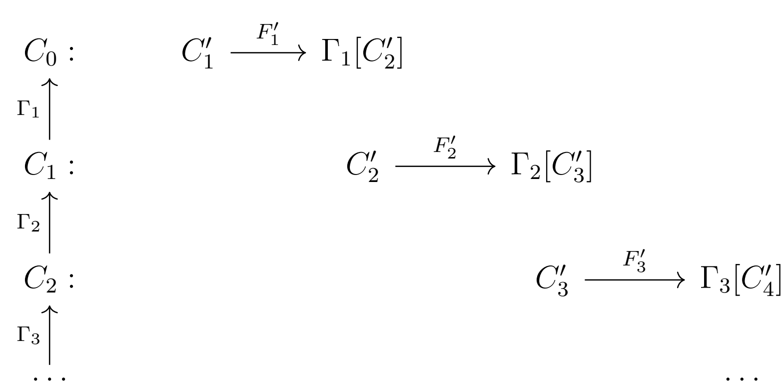
- The first ingredient is an infinite sequence $C_1, C_2, C_3, \ldots$, in which each C_i (for $i \geq 1$) is the global aspect of a lexcategory C'_i internal to C_{i-1}.

Thus, each C'_{i+n} is n-tuply internal to C_i.

(Throughout the following, it will be useful to keep in mind that we are using these general naming habits: Primed names are used for internal structures, while unprimed names certain corresponding global structures. Furthermore, names subscripted with index i arise from structure internal to C_{i-1}.)

- The second ingredient comprising a geminal category is an infinite sequence of internal lexfunctors $F'_1, F'_2, F'_3, \ldots$, where each $F'_i : C'_i \to \Gamma[C'_{i+1}]$ is internal to C_{i-1} (for $i \geq 1$).

Pictorially, this can be envisioned like so:

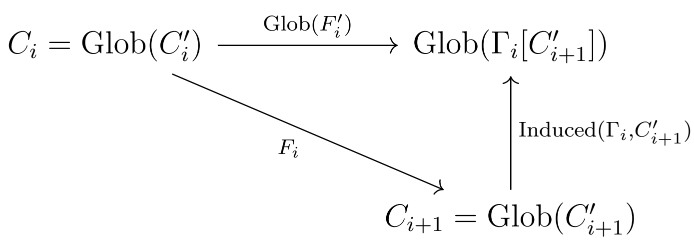


[1] In the manner which is sometimes outside of mathematics called "the Droste effect".

[2] Another evocative name for this concept might be "nesting doll category".

Here, the first row is structure internal to C_0, the second row is structure internal to C_1 (thus, doubly internal to the ambient C_0), the third row is structure internal to C_2 (thus, triply internal to the ambient C_0), and so on. We also for convenience use the abbreviation Γ_i for $\Gamma_{C'_i} : C_i \to C_{i-1}$ for $i \geq 1$, illustrating these in the vertical line on the left of the picture.

From the internal lexfunctor $F'_i : C'_i \to \Gamma_i[C'_{i+1}]$, we shall also define a lexfunctor $F_i : C_i \to C_{i+1}$ like so: As Observation 5.4 tells us that $\text{Induced}(\Gamma_i, C'_{i+1}) : \text{Glob}(C'_{i+1}) \to \text{Glob}(\Gamma_i[C'_{i+1}])$ is an isomorphism, we take $F_i : C_i \to C_{i+1}$ to be the unique map making the following diagram commute:



These F_i are convenient as they line up straightforwardly:

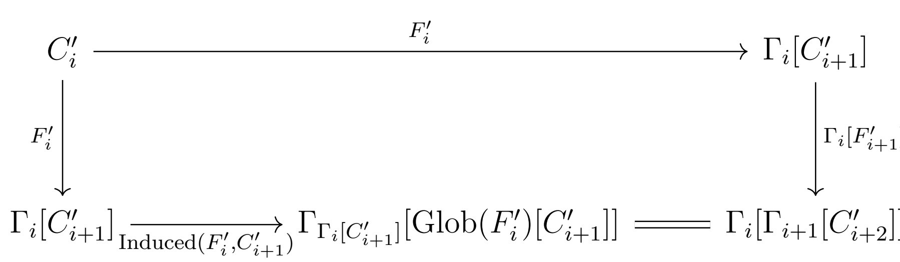
$$C_1 \xrightarrow{F_1} C_2 \xrightarrow{F_2} C_3 \xrightarrow{F_3} \dots$$

Finally, the last ingredients we require are some equations:

- We require that $F_i[C'_j] = C'_{j+1}$ and $F_i[F'_j] = F'_{j+1}$ for $j > i \geq 1$.

  (We are using Definition 5.6 here to apply F_i to structures multiply internal to its domain C_i.)

- Furthermore, we require that the following diagram of lexfunctors internal to C_{i-1} commutes, for each $i \geq 1$. We call this equation E_i.



  That is, we require that $\text{Induced}(F'_i, C'_{i+1}) \circ F'_i = \Gamma_i[F'_{i+1}] \circ F'_i$. This could be glossed as "$F'_i \circ F'_i = F'_{i+1} \circ F'_i$", in abuse of notation a la Note 5.7.

  (To derive the identity in the bottom-right of the above diagram, first note that $\text{Glob}(F'_i)[C'_{i+1}] = \text{Induced}(\Gamma_i, C'_{i+1})[F_i[C'_{i+1}]] = \text{Induced}(\Gamma_i, C'_{i+1})[C'_{i+2}]$.

  Thus, $\Gamma_{\Gamma_i[C'_{i+1}]}[F'_i[C'_{i+1}]] = \Gamma_{\Gamma_i[C'_{i+1}]}[\text{Induced}(\Gamma_i, C'_{i+1})[C'_{i+2}]] = \Gamma_i[\Gamma_{i+1}[C'_{i+2}]]$, where the last step is by Lemma 5.5.)

This concludes the definition of a geminal category internal to C_0.

By a **geminal category** simpliciter, we mean of course the case where $C_0 = \mathrm{Set}$. (Note that in this case, C'_1 can be identified with its global aspect C_1, in the same way that any structure internal to Set can be identified with its global aspect, as the global elements functor from Set to Set is the identity.). We wrote out here the definition for general C_0, instead of specifically for $C_0 = \mathrm{Set}$, in order to emphasize certain symmetries in this definition.

When being fully explicit, we reference a geminal category by enumerating its components $\langle C'_1, C'_2, C'_3, \ldots; F'_1, F'_2, F'_3, \ldots \rangle$. Given such a geminal category K, we may write $|K|$ to refer to its underlying lexcategory C'_1.

All aforementioned structure apart from C_0 itself has been given as i-tuply internal to C_0 for some $i > 0$. Thus, all of this structure is indeed given by diagrams within C_0.

Indeed, this definition of geminal category is manifestly essentially algebraic. That is, there is an essentially algebraic theory such that models of that theory internal to C_0 are the same thing as geminal categories internal to C_0.

Our ultimate goal will be to show that this theory of geminal categories is the initial introspective theory. This is the whole motivation for our study of geminal categories. But to show this result, we must develop some other machinery first.

5.5 Geminal category homomorphisms

As geminal categories are defined by an essentially algebraic theory, we automatically get a notion of homomorphism between geminal categories. It amounts to the following:

Definition 5.12 Given two geminal categories $\langle C'_1, C'_2, C'_3, \ldots; F'_1, F'_2, F'_3, \ldots \rangle$ and $\langle D'_1, D'_2, D'_3, \ldots; \phi'_1, \phi'_2, \phi'_3, \ldots \rangle$, a **homomorphism** from the former to the latter consists of a strict lexfunctor $H : C'_1 \to D'_1$ such that $H[C'_i] = D'_i$ and $H[F'_i] = \phi'_i$ for each $i > 1$, while also the following diagram commutes:

$$\begin{array}{ccc} C'_1 & \xrightarrow{\quad H \quad} & D'_1 \\ \Big\downarrow{\scriptstyle F'_1} & & \Big\downarrow{\scriptstyle \phi'_1} \\ \mathrm{Glob}(C'_2) & \xrightarrow[\mathrm{Induced}(H, C'_2)]{} & \mathrm{Glob}(H[C'_2]) = \mathrm{Glob}(D'_2) \end{array}$$

(In the above, $H[C'_i]$ and $H[F'_i]$ make use of Definition 5.6 to denote the application of H to multiply internal structures.)

Theorem 5.13 Given any geminal category $K = \langle C'_1, C'_2, C'_3, \ldots; F'_1, F'_2, F'_3, \ldots \rangle$, we have also that $\langle C'_2, C'_3, C'_4, \ldots; F'_2, F'_3, F'_4 \ldots \rangle$ comprises a geminal category internal to $|K| = C'_1$. We refer to this internal geminal category as $\mathrm{InteriorGeminal}(K)$.

We furthermore have that F'_1 acts as a geminal category homomorphism from K to the global aspect of $\mathrm{InteriorGeminal}(K)$. We refer to this homomorphism as $\mathrm{IntoSelf}(K) : K \to \Gamma[\mathrm{InteriorGeminal}(K)]$.

Proof. This is all direct by definition.

For the first part, each condition imposed upon each C'_i or F'_i in the definition of a geminal category comes with an analogous condition imposed upon C'_{i+1} or F'_{i+1}. Thus, it is immediate that the given $\mathrm{InteriorGeminal}(K)$ satisfies the conditions to be a geminal category internal to $|K|$.

For the second part, the definition of a geminal category directly imposes upon F'_1 precisely the conditions which are necessary for F'_1 to comprise a geminal category homomorphism from K to the global aspect of $\mathrm{InteriorGeminal}(K)$. In particular, equation E_1 from Geminal category (Definition 5.11) is identical to the commutative diagram from Definition 5.12, in this context. ■

Via the yoga of functorial semantics, Theorem 5.13 states how the theory of geminal categories can be equipped as an introspective theory. In detail, this is given like so:

Construction 5.14 Let $\mathrm{Th}(\mathrm{GC})$ be the free strict lexcategory with an internal geminal category (that is, in the terminology of Quasi-equational theories (Section 1.10), we take $\mathrm{Th}(\mathrm{GC})$ to be the classifying strict lexcategory $\mathcal{C}_{\mathbb{T}}$, where $\mathbb{T}$ is the theory of geminal categories).

Thus, strict lexfunctors from $\mathrm{Th}(\mathrm{GC})$ to any other strict lexcategory D correspond to geminal categories internal to D, while natural transformations between such lexfunctors correspond to homomorphisms between these D-internal geminal categories.

Let K denote the $\mathrm{Th}(\mathrm{GC})$-internal geminal category corresponding to the identity functor on $\mathrm{Th}(\mathrm{GC})$.

By Theorem 5.13 in the internal logic of $\mathrm{Th}(\mathrm{GC})$, we obtain also a geminal category $\mathrm{InteriorGeminal}(K)$ internal to $|K|$, as well as a homomorphism $\mathrm{IntoSelf}(K) : K \to \Gamma[\mathrm{InteriorGeminal}(K)]$.

Thus, there is some strict lexfunctor $\mathcal{S}$ from $\mathrm{Th}(\mathrm{GC})$ to the global aspect of $|K|$, corresponding to $\mathrm{InteriorGeminal}(K)$. Furthermore, there is some natural transformation $\mathcal{N}$ from the identity functor on $\mathrm{Th}(\mathrm{GC})$ to $\mathrm{Hom}_{|K|}(1, \mathcal{S}(-))$, corresponding to $\mathrm{IntoSelf}(K)$.

Putting this together, we have a strict introspective theory $\langle \mathrm{Th}(\mathrm{GC}), |K|, \mathcal{S}, \mathcal{N} \rangle$.

5.6 Compactly defined geminal categories

The above all amounts to an infinitary presentation of the theory of geminal categories. For this reason, we call it the “verbose presentation” of geminal categories. However, it turns out this same theory can be finitely axiomatized as well.

DEFINITION 5.15 (GEMINAL CATEGORY, COMPACT PRESENTATION) A **compactly presented geminal category** internal to lexcategory C_0 consists of the structure C'_i, F'_i, and equations E_i of the verbose presentation, but only for $i \in \{1, 2\}$.

(Here, in interpreting the codomain of F'_2, we take C'_3 to be $F_1[C'_2]$, and in interpreting the equation E_2, we take F'_3 to be $F_1[F'_2]$, where F_1 is defined from F'_1 just as in Geminal category (Definition 5.11).)

That is, a compactly presented geminal category internal to C_0 consists of the following six pieces of data:

- A lexcategory C'_1 internal to C_0. We refer to its global aspect as C_1, and we refer to $\Gamma_{C'_1} : C_1 \to C_0$ as Γ_1.
- A lexcategory C'_2 internal to C_1. We refer to its global aspect as C_2, and we refer to $\Gamma_{C'_2} : C_2 \to C_1$ as Γ_2.
- A lexfunctor $F'_1 : C'_1 \to \Gamma_1[C'_2]$, internal to C_0. We define $F_1 : C_1 \to C_2$ from this just as in Geminal category (Definition 5.11).
- A lexfunctor $F'_2 : C'_2 \to \Gamma_2[C'_3]$, internal to C_1.

  (Here, C'_3 is defined as $F_1[C'_2]$.)
- The equation $\mathrm{Induced}(F'_1, C'_2) \circ F'_1 = \Gamma_1[F'_2] \circ F'_1$, internal to C_0. We call this equation E_1.
- The equation $\mathrm{Induced}(F'_2, C'_3) \circ F'_2 = \Gamma_2[F'_3] \circ F'_2$, internal to C_1. We call this equation E_2.

  (Here, F'_3 is defined as $F_1[F'_2]$.)

As usual, we reference a compactly presented geminal category by enumerating the ordered tuple $\langle C'_1, C'_2; F'_1, F'_2 \rangle$.

Clearly, the structure defining a compactly presented geminal category is part of the structure in our verbose definition of a geminal category. But in fact, these are equivalent definitions.

THEOREM 5.16 The structure of a compactly presented geminal category uniquely determines the further structure of a geminal category (as originally defined in Geminal category (Definition 5.11))).

Proof. Throughout the following, as before, we define each F_i from the corresponding F'_i just as in Geminal category (Definition 5.11).

By definition, in a geminal category, we must have that $C'_j = F_1[C'_{j-1}]$ and $F'_j = F_1[F'_{j-1}]$ for each $j > 2$.

Accordingly, if we are given the structure in Geminal category, compact presentation (Definition 5.15), and we are to extend it to all the further structure in Geminal category

(Definition 5.11), we may use the above particular recurrences to inductively define C'_j and F'_j for each $j > 2$, ultimately in terms of the base cases of $j \in \{1, 2\}$ which we have been given. Adopt these definitions throughout the following accordingly.

The equations given to us directly in the compact presentation are the equations E_1 and E_2 of the verbose presentation. Furthermore, we again obtain the equation E_i for each $i > 2$ inductively by applying F_1 to E_{i-1}.

What remains is only to see that each F_i takes C'_j to C'_{j+1} and takes F'_j to F'_{j+1}, for $j > i \geq 1$.

We prove this by induction on i. For the base case of $i = 1$, we have ensured this by construction. As for the inductive step, suppose we know this already holds for i. Then for $j > i + 1$ we have $F_{i+1}[C'_j] = F_{i+1}[F_i[C'_{j-1}]] = \mathrm{Induced}(F_i, C'_{i+1})[F_i[C'_{j-1}]] = F_i[F_i[C'_{j-1}]] = F_i[C'_j] = C'_{j+1}$, where the second step is by the global aspect of E_i (along with some applications of Observation 5.4), the third step is by Definition 5.6, and the other steps are by our induction hypothesis. And similarly with F' in place of C' throughout as well. ∎

COROLLARY 5.17 In Definition 5.12, the conditions $H[C'_i] = D'_i$ and $H[F'_i] = \phi'_i$ automatically follow for all $i > 2$ once they hold for $i = 2$.

Thus, we can go back and forth between thinking of geminal categories in either the verbose or compact presentation as we please, whichever is most convenient at any moment.

5.7 Geminal categories from introspective theories

CONSTRUCTION 5.18 From a strict introspective theory $\langle T, C, \mathcal{S}, \mathcal{N} \rangle$, we obtain a geminal category $\langle T, C; \mathcal{S}, \mathcal{N}_C \rangle$, whose underlying lexcategory is T. This is the canonical way to view an introspective theory as a geminal category.

Proof. It is immediate in the definition of a strict introspective theory that C is a lexcategory internal to T, and $\mathcal{S}$ is a lexfunctor from T to $\mathrm{Glob}(C)$. This gives us the first three out of the six ingredients of Geminal category, compact presentation (Definition 5.15).

As for $\mathcal{N}_C$ (meaning the components of the natural transformation $\mathcal{N}$ at the objects of the diagram within T which defines the T-internal lexcategory C), this gives us a T-internal lexfunctor from C to $\mathrm{Hom}_C(1, \mathcal{S}[C]) = \Gamma[\mathcal{S}[C]]$. This is the fourth ingredient of Geminal category, compact presentation (Definition 5.15).

What remains are to verify equations E_1 and E_2. In this context, E_1 is a special case of $\mathcal{S}$ Matches $\mathcal{N}$ (Lemma 2.25), while E_2 is given by the naturality of $\mathcal{N}$ with respect to the components of $\mathcal{N}_C$ themselves.

This completes the construction. We observe furthermore that strict introspective theory homomorphisms are automatically geminal category homomorphisms between the geminal categories obtained by this construction. ∎

There is another closely related construction which is of even more importance:

CONSTRUCTION 5.19 From a strict introspective theory $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$, we obtain a T-internal geminal category $\langle C, \mathcal{S}[C]; \mathcal{N}_C, \mathcal{S}[\mathcal{N}_C]\rangle$, whose underlying lexcategory is C.

Proof. This is the result of first obtaining the geminal category $\gamma = \langle T, C, \mathcal{S}, \mathcal{N}_C\rangle$ from Construction 5.18, and then forming $\mathrm{InteriorGeminal}(\gamma)$. ■

5.8 The free introspective theory

We now are ready to prove our main result about geminal categories.

THEOREM 5.20 The strict introspective theory given in Construction 5.14 is the initial strict introspective theory.

Proof. We must show there is a unique homomorphism from the strict introspective theory $\langle \mathrm{Th(GC)}, K\rangle$ of Construction 5.14 to any other strict introspective theory $\langle T, D\rangle$.

Such a homomorphism is comprised of a strict lexfunctor $H : \mathrm{Th(GC)} \to T$ satisfying certain conditions. By the nature of $\mathrm{Th(GC)}$, this amounts to a geminal category $\langle D'_1, D'_2, D'_3, \ldots; F'_1, F'_2, F'_3, \ldots\rangle$ internal to T satisfying certain conditions.

One particular geminal category internal to T is the one that is given by $\gamma = \langle D, \mathcal{S}[D]; \mathcal{N}_D, \mathcal{S}[\mathcal{N}_D]\rangle$, as noted at Construction 5.19. In verbose terms, this geminal category is $\langle D, \mathcal{S}[D], \mathcal{S}[\mathcal{S}[D]], \ldots; \mathcal{N}_D, \mathcal{S}[\mathcal{N}_D], \mathcal{S}[\mathcal{S}[\mathcal{N}_D]], \ldots\rangle$, with each successive component being $\mathcal{S}$ applied to the previous component.

What remains is to show that the lexfunctor $H : \mathrm{Th(GC)} \to T$ corresponding to this γ uniquely satisfies the conditions of Definition 5.10.

The condition "$H[C_1] = C_2$" in Definition 5.10 says in this context that we must use a geminal category whose underlying lexcategory is D.

The condition concerning $\mathcal{N}$ in Definition 5.10, along with the definition of $\mathcal{N}$ in Construction 5.14, says that we must use a geminal category whose first lexfunctor component is $\mathcal{N}_D$.

Finally, the commutative diagram concerning $\mathcal{S}$ in Definition 5.10, along with the definition of $\mathcal{S}$ in Construction 5.14, says we must use a geminal category such that each successive component of this geminal category is $\mathcal{S}$ applied to the previous component.

The conjunction of these conditions clearly is uniquely satisfied by γ. This completes the proof. ■

OBSERVATION 5.21 Given the result of Theorem 5.20, we can rephrase Construction 5.18 as telling us that every strict introspective theory is a model of the initial introspective theory, so to speak. In other words, there is a lexfunctor interpreting the initial introspective theory into the theory of strict introspective theories. This is quite remarkable!

5.9 Geminal gadgets

We have now successfully described the initial introspective theory. But we can also take our free construction results a bit further than this.

Specifically, every introspective theory is, among other things, an essentially algebraic theory extending the theory of strict lexcategories. That is, we have a functor from the category of strict introspective theories to the category of strict lexcategories with a designated internal lexcategory (this functor takes $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$ to $\langle T, C\rangle$). This functor has a left adjoint.

Put in other words, for any essentially algebraic theory Th such that models of Th come with an underlying strict lexcategory, there is a free strict introspective theory $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$ with a designated T-internal model of Th with underlying lexcategory C.

For simplicity as a first introduction, everything done previously was the special case where Th was simply the theory of strict lexcategories itself. But now we describe the more general results, which follow by almost exactly the same reasoning as used before:

Specifically, let models of Th be called "gadgets", and maps between them called "gadget homomorphisms". Then the free introspective theory extending Th is the theory of "geminal gadgets", with the definition of a "geminal gadget" being exactly as in either definition of a "geminal category", but with all instances of lexcategories and lexfunctors replaced by gadgets and gadget homomorphisms.

This is by exactly the same arguments as we have just given. All the results and arguments given earlier in this chapter apply just as well mutatis mutandis when lexcategories and lexfunctors are replaced by gadgets and gadget homomorphisms, except for Construction 5.18 (it will not be the case that an arbitrary strict introspective theory can be viewed as a geminal gadget). However, the analogue of the construction Construction 5.19 still holds (i.e., given an introspective theory $\langle T, C\rangle$ such that C is the underlying lexcategory of a gadget, then the geminal category structure which C is equipped with by Construction 5.19 furthermore underlies geminal gadget structure).

5.10 Archetypal examples of geminal categories

Given any introspective theory $\langle T, C\rangle$ and any lexfunctor $f : T \to \mathrm{Set}$, we obtain automatically a geminal category $f[\mathrm{InteriorGeminal}(\langle T, C\rangle)] = f[\langle C, \mathcal{S}[C]; \mathcal{N}_C, \mathcal{S}[\mathcal{N}_C]\rangle]$.

Thus we have archetypal geminal categories corresponding to each of our archetypal introspective theories:

5.10.1 ZF-Finite example

Recall the introspective theory $\langle Z_{\Sigma_1}, Z'\rangle$ of Construction 2.10, in which Z_{Σ_1} was the category of Σ_1-definable sets and functions modulo provable equality in ZF-Finite, while Z'

was the internal construction of the category Z of Σ_1-definable sets and functions modulo provable equality in ZF-Finite

There is a lexfunctor $f : Z_{\Sigma_1} \to \mathrm{Set}$ which sends each definition in ZF-Finite to the set or function it actually defines (in particular, $f[Z'] = Z$). We thus obtain a geminal category $f[\mathrm{InteriorGeminal}(\langle Z_{\Sigma_1}, Z' \rangle)] = \langle Z, Z'; F_1, F_2' \rangle$ in which Z is the category of arbitrary definable sets and functions modulo provable equality in ZF-Finite, while Z' is the analogous construction internal to Z. $F_1 : Z \to \mathrm{Glob}(Z')$ straightforwardly sends each definable set or function in Z to the corresponding construction in $\mathrm{Glob}(Z')$, and F_2' is the Z-internal lexfunctor constructed exactly analogously to F_1.

Notably, this example of a geminal category does not require us to incorporate Σ_1 constraints anywhere. In this sense, it is a more familiar object for study than Z_{Σ_1} itself was. We had noted in Warning 2.11 that this structure $\langle Z, Z' \rangle$ is not an introspective theory, but we see here that the natural structure it has is as a geminal category instead.

5.10.2 Kripke frame example

In Construction 2.16, we constructed from any well-founded pre-order P (and a suitable choice of set-sized full sublexcategories Set_q of Set for each $q \in P$), an introspective theory T. As the global aspect of $\mathrm{InteriorGeminal}(T)$ for this introspective theory, we get a geminal category $\langle C_1, C_2'; F_1, F_2' \rangle$ where C_1 is the full subcategory of $\mathrm{Set}^{|P|}$ comprising those presheaves X for which $X(p) \in \mathrm{Set}_p$ for each $p \in P$. The C-internal lexcategory C_2' is the $|P|$-indexed category such that for each $p \in P$, we have $C_2'(p) = \prod_{q<p} \mathrm{Set}_q$. Thus, $\mathrm{Glob}(C_2') = \prod_{p \in P} \prod_{q<p} \mathrm{Set}_q$. The functor $F_1 : C \to \mathrm{Glob}(C_2')$ is then defined by $F_1(X)(p)(q) = X(q)$, and then $\Gamma_{C_2'} \circ F_1 : C \to C$ is the map $X \mapsto p \mapsto \prod_{q<p} X(q)$. Accordingly, $\Gamma_{C_2'}[F_1[C_2']]$ is the $|P|$-indexed category given by $p \mapsto \prod_{q<p} \prod_{r<q} \mathrm{Set}_r$. And the map $F_2' : C_2' \to \Gamma_{C_2'}[F_1[C_2']]$ is given by the obvious projections (that is, its aspect at p maps $\prod_{r<p} \mathrm{Set}_r$ to $\prod_{q<p} \prod_{r<q} \mathrm{Set}_r$ in the obvious way, taking advantage of the transitivity of $<$).

5.10.3 Step-indexing example

Our last archetypal example of a geminal category would be $\mathrm{Glob}(\mathrm{InteriorGeminal}(T))$, where T is our archetypal example of an introspective theory constructed in Construction 2.13. However, with this particular introspective theory, we have the property that the geminal category homomorphism $\mathrm{IntoSelf}(T) : T \to \mathrm{Glob}(\mathrm{InteriorGeminal}(T))$ is an isomorphism[3]. Thus, this $\mathrm{Glob}(\mathrm{InteriorGeminal}(T))$ is not very illustrative of the distinctive nature of geminal categories as differentiated from introspective theories in general. But it is useful as a reminder that every example of an introspective theory is also an example of a geminal category (via Construction 5.18)!

[3]Pedantically, we should say that there is a choice of strict introspective theory presenting T for which this is an isomorphism

5.11 Modal logic in geminal categories

Recall from The box operator (Section 3.2) and Modal logic and axiom 4 (Section 3.3) that every introspective theory $\langle T, C\rangle$ comes with an internal endolexfunctor $\Box_C = \mathcal{F}(\mathrm{Hom}_C(1, -)) : C \to C$ which interprets the modal logic GL, in that we have a canonical natural transformation $4 : \Box_C \to \Box_C\Box_C$ induced by the natural transformation $\mathcal{N}$.

As the theory of geminal categories is itself an introspective theory, we thus obtain on the underlying lexcategory of any geminal category an endolexfunctor with the same properties. However, it may be tricky to see what this box operator directly amounts to for geminal categories (as a geminal category does not contain such structure as $\mathcal{F}$, which we used when defining the box operators of an introspective theory). The following lemma will help us see how the box operator for geminal categories can be more directly defined in terms of geminal category structure.

LEMMA 5.22 Let $\langle T, C\rangle$ be an introspective theory. Then the T-internal endolexfunctor $\Box_C : C \to C$ as described in Definition 3.2 matches the composition of the T-internal lexfunctors $\mathcal{N}_C : C \to \Box_T C$ and $\mathrm{Hom}_{\mathcal{S}[C]}(1, -) : \Box_T C \to C$.

In other words, the following diagram of T-indexed lexcategories and lexfunctors commutes:

$$\begin{array}{ccc} C & \xrightarrow{\mathcal{N}_C} & \Box_T C = \Gamma_C[\mathcal{S}[C]] \\ {\scriptstyle \mathrm{Hom}_C(1,-)}\downarrow & & \downarrow{\scriptstyle \mathrm{Hom}_{\mathcal{S}[C]}(1,-)} \\ T/- & \xrightarrow[\mathcal{F}]{} & C \end{array}$$

Proof. Via Slice (Pre-)Introspective Theories (Construction 2.19) and Theorem 1.8, it suffices to show that the global aspects of such diagrams commute, as arbitrary aspects can be seen as global aspects of slice introspective theories.

The global aspect of the above diagram amounts to the following (keeping in mind $\mathcal{S}$ Matches $\mathcal{N}$ (Lemma 2.25) for the top arrow):

$$\begin{array}{ccc} \mathrm{Glob}(C) & \xrightarrow{\mathrm{Induced}(\mathcal{S},C)} & \mathrm{Glob}(\mathcal{S}[C]) \\ {\scriptstyle \Gamma_C}\downarrow & & \downarrow{\scriptstyle \Gamma_{\mathcal{S}[C]}} \\ T & \xrightarrow[\mathcal{S}]{} & \mathrm{Glob}(C) \end{array}$$

Finally, we observe that this last diagram commutes as an instance of Lemma 5.5. ■

COROLLARY 5.23 Interpreting Lemma 5.22 in the particular context of the introspective theory $\langle \mathrm{Th}(\mathrm{GC}), |K|\rangle$ of Construction 5.14, we find that any geminal category $C_1' = \langle C_1', C_2'; F_1', F_2'\rangle$

comes with an endolexfunctor $\Box_{C'_1} = \Gamma_{C'_2} \circ F'_1 : C'_1 \to C'_1$, along with a natural transformation $4 : \Box_{C'_1} \to \Box_{C'_1}\Box_{C'_1}$ corresponding to the action of F'_2.

Observation 5.24 For any geminal category $\langle C'_1, C'_2; F'_1, F'_2 \rangle$, the operator $\Box_{C'_1} = \Gamma_{C'_2} \circ F'_1 : C'_1 \to C'_1$ defined above has the Löb property with uniqueness (as defined at Definition 4.1). This follows from the fact that, by Corollary 4.21, every aspect of every box operator on every introspective theory has the Löb property with uniqueness, and thus in particular, in the internal logic of the introspective theory $\langle \mathrm{Th}(\mathrm{GC}), |K| \rangle$ of Construction 5.14, we have that $\Box_K$ has the Löb property with uniqueness (in more detail, the Löb property is satisfied with respect to the generic morphism of $|K|$ in its $\mathrm{Mor}(|K|)$-aspect, thus establishing that it holds for all geminal categories with respect to all of their morphisms; uniqueness follows by a similar argument, or just by invoking Theorem 4.4).

5.12 Co-free introspective theories and geminal categories

We have above discussed how to create free introspective theories, which can be thought of as produced by a certain left adjoint functor[4]. In this section, we discuss some right adjoint constructions, which can be thought of as "co-free".

Construction 5.25 Construing strict introspective theories as geminal categories via Construction 5.18 gives us a functor from the category of strict introspective theories to the category of geminal categories (or more generally, a functor from the category of V-internal strict introspective theories to the category of V-internal geminal categories, for any fixed lexcategory V). This functor has a right adjoint.

As this works for arbitrary lexcategories V, this right adjoint admits an explicit description, as a purely lex construction.

For linguistic convenience, we shall in the following take V as Set, but it will be clear that the same explicit construction works for any ambient lexcategory V.

We are tasked with showing that, for any geminal category C'_1, there is a suitably terminal strict introspective theory with a geminal category homomorphism to C'_1.

We will first give the details of the construction, and then give the proof that it has the terminality property.

Construction details. Let $C'_1 = \langle C'_1, C'_2; F'_1, F'_2 \rangle$ be an arbitrary geminal category. Via Corollary 5.23, this comes with an endolexfunctor $\Box_{C'_1} = \Gamma_{C'_2} \circ F'_1 : C'_1 \to C'_1$, along with a natural transformation $4 : \Box_{C'_1} \to \Box_{C'_1}\Box_{C'_1}$ corresponding to the action of F'_2.

We will in the following write $\Box$ with no subscript to mean this $\Box_{C'_1}$.

Via Coalgebras As Strict Lexcategory (Lemma 1.83), the category of $\Box$-coalgebras is a strict lexcategory. Among these $\Box$-coalgebras, there are some coalgebras $m : c \to \Box c$ with the property that the following diagram commutes:

[4] Specifically, left adjoint to the forgetful functor from strict introspective theories to strict lexcategories with a designated internal lexcategory.

$$\begin{array}{ccc} c & \xrightarrow{m} & \Box c \\ {\scriptstyle m}\big\downarrow & & \big\downarrow{\scriptstyle \Box m} \\ \Box c & \xrightarrow[4_c]{} & \Box\Box c \end{array}$$

Let Z be the full subcategory of those $\Box$-coalgebras with the specified property. It is readily seen that this Z is closed under the finite limits of the category of $\Box$-coalgebras, and thus is itself a strict lexcategory. (Indeed, Z can be defined as the equalizer of two strict lexfunctors from the $\Box$-coalgebras to the $\Box\Box$-coalgebras.)

Note that we have the following commutative diagram of internal lexfunctors in C'_1:

$$\begin{array}{ccc} C'_2 & \xrightarrow{F'_2} & \Box C'_2 \\ {\scriptstyle F'_2}\big\downarrow & & \big\downarrow{\scriptstyle \Box F'_2 = \Gamma_{C'_2}[F_1[F'_2]] = \Gamma_{C'_2}[F'_3]} \\ \Box C'_2 & \xrightarrow[4_{C'_2} = \mathrm{Induced}(F'_2, C'_2)]{} & \Box\Box C'_2 \end{array}$$

This diagram commutes by equation E_2. But this is also the commutative diagram which establishes that the internal lexcategory $F'_2 : C'_2 \to \Box C'_2$ within the category of $\Box$-coalgebras is furthermore within its subcategory Z. When thinking of F'_2 as an internal lexcategory within Z in this way, let us call it Z_2.

Note that the strict lexfunctor $F'_1 : C'_1 \to \mathrm{Glob}(C'_2)$ is such that for any object or morphism x of C'_1, if we interpret $F'_1(x)$ as a morphism from 1 to $\mathrm{Ob}(C'_2)$ or $\mathrm{Mor}(C'_2)$, we obtain a commutative diagram of the following form:

$$\begin{array}{ccc} 1 & \xrightarrow{F'_1(x)} & \mathrm{Ob}(C'_2) \text{ or } \mathrm{Mor}(C'_2) \\ {\scriptstyle !}\big\downarrow & & \big\downarrow{\scriptstyle F'_2} \\ \Box 1 = 1 & \xrightarrow[\Box\left(F'_1(x)\right) = \Gamma_{C'_2}[F'_1[F'_1(x)]]]{} & \Box\left(\mathrm{Ob}(C'_2) \text{ or } \mathrm{Mor}(C'_2)\right) \end{array}$$

That this diagram commutes is by equation E_1 of C'_1 being a geminal category. Thus $F'_1(x)$ amounts to a global element of $\mathrm{Ob}(Z_2)$ or $\mathrm{Mor}(Z_2)$, and thus F'_1 acts as a strict lexfunctor from C'_1 to $\mathrm{Glob}(Z_2)$. By composing this with the projection functor $\pi : Z \to C'_1$, we get a strict lexfunctor $\mathcal{S} : Z \to \mathrm{Glob}(Z_2)$. (It may be surprising that this $\mathcal{S}$ will discard all information lost in the projection from Z to C'_1, but this will indeed be the correct one for our purposes!)

Finally, for $\mathcal{N}$, we observe by unwinding definitions that $\mathrm{Hom}_{Z_2}(1, \mathcal{S}(-)) : Z \to Z$ is the functor which takes a coalgebra on carrier object c to the coalgebra $4_c : \Box c \to \Box\Box c$ and which takes coalgebra morphisms to the corresponding naturality square for the natural transformation 4 (thus, $\pi \circ \mathrm{Hom}_{Z_2}(1, \mathcal{S}(-)) = \Box \circ \pi : Z \to C_1'$). Thus, by the defining condition of Z, we get for each $m \in \mathrm{Ob}(Z)$ a coalgebra morphism from m to $\mathrm{Hom}_{Z_2}(1, \mathcal{S}(m))$ whose underlying morphism in C_1' is m itself, as described by the following commutative diagram in C_1':

$$\begin{array}{ccc} c & \xrightarrow{\quad m \quad} & \Box c \\ {\scriptstyle m}\downarrow & & \downarrow{\scriptstyle \Box m} \\ \Box c & \xrightarrow[\mathrm{Hom}_{Z_2}(1,\mathcal{S}(m)) = 4_c]{} & \Box\Box c \end{array}$$

In the above commutative diagram within C_1', the top arrow is the coalgebra m, while the bottom arrow is the coalgebra $\mathrm{Hom}_{Z_2}(1, \mathcal{S}(m))$.

These maps from each m to $\mathrm{Hom}_{Z_2}(1, \mathcal{S}(m))$ comprise a natural transformation $\mathcal{N}$ between id_Z and $\mathrm{Hom}_{Z_2}(1, \mathcal{S}(-))$ whose naturality is demonstrated like so: Consider any two coalgebras m_1 and m_2 in Z and a coalgebra map $h : m_1 \to m_2$. The condition for h to be a coalgebra map is the very same as the naturality square for this $\mathcal{N}$, amounting to the following commutative diagram in C_1':

$$\begin{array}{ccc} c & \xrightarrow{\quad \pi(h) \quad} & d \\ {\scriptstyle m_1}\downarrow & & \downarrow{\scriptstyle m_2} \\ \Box c & \xrightarrow[\Box(\pi(h)) = \pi(\mathrm{Hom}_{Z_2}(1,\mathcal{S}(h)))]{} & \Box d \end{array}$$

Thus we have a strict introspective theory $Z = \langle Z, Z_2, \mathcal{S}, \mathcal{N} \rangle$.

By construction, the projection functor $\pi : Z \to C_1'$, is a strict lexfunctor which takes Z_2 to C_2' and takes $\mathcal{N}_{Z_2}$ to F_2', while furthermore the following diagram commutes:

$$\begin{array}{ccc} Z & \xrightarrow{\quad \pi \quad} & C_1' \\ {\scriptstyle \mathcal{S}}\downarrow & & \downarrow{\scriptstyle F_1'} \\ \mathrm{Glob}(Z_2) & \xrightarrow[\mathrm{Induced}(\pi, Z_2)]{} & \mathrm{Glob}(\pi[Z_2]) = \mathrm{Glob}(C_2') \end{array}$$

Thus, we have π as a geminal category homomorphism from Z to C_1'.

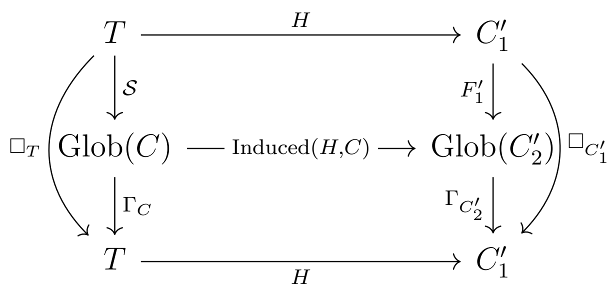
Having described the strict introspective theory Z and its geminal category homomorphism π to C_1', we now prove their terminality among all strict introspective theories with a designated geminal category homomorphism to C_1':

Proof. Let $T = \langle T, C \rangle$ be an arbitrary strict introspective theory, and let $H : T \to C_1'$ be a geminal category homomorphism. We will show that there is a unique strict introspective theory homomorphism $\beta : T \to Z$ such that $\pi \circ \beta = H$.

The condition $\pi \circ \beta = H$ tells us right away what the carriers of the coalgebras and coalgebra morphisms produced by β must be. Furthermore, our construction of $\mathcal{N}_z$ for objects z of Z was such that $\pi(\mathcal{N}_z)$ in C_1' is the very same as the coalgebra z itself. Thus, for any object $t \in T$, the specific coalgebra $\beta(t)$ will be the one given by $\pi(\mathcal{N}_{\beta(t)})$, which by virtue of β being a strict introspective theory homomorphism must be the same as $\pi(\beta(\mathcal{N}_t)) = H(\mathcal{N}_t)$.

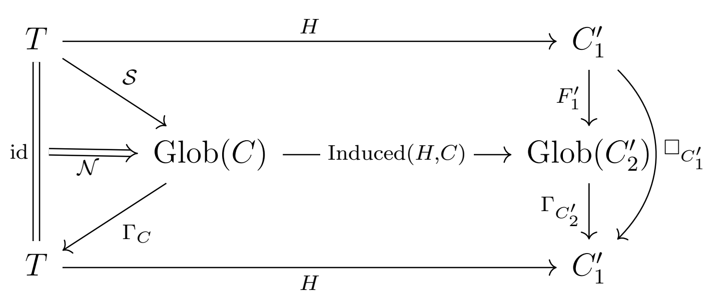
Thus, the uniqueness of β is assured and what remains is only to see that a β so-constructed is indeed a strict introspective theory homomorphism. First, let us see that we in fact do have such a $\beta : T \to Z$ as a strict lexfunctor:

By virtue of H being a geminal category homomorphism, we have that $H \circ \Box_T = \Box_{C_1'} \circ H$. In detail, this is seen via the following commutative diagram:



In the above diagram, the left side is the definition of $\Box_T$ and the right side is the definition of $\Box_{C_1'}$. The top rectangle is one of the conditions in Definition 5.12 and the bottom rectangle is by Lemma 5.5.

Thus, the whiskering of $\mathcal{N} : \mathrm{id}_T \to \Box_T$ along H yields a natural transformation from H to $H \circ \Box_T = \Box_{C_1'} \circ H$. Illustrated like so:



This natural transformation from H to $\square_{C_1'} \circ H$ acts as a functor β from T to the category of $\square_{C_1'}$-coalgebras, such that $\pi \circ \beta = H$. As H is a strict lexfunctor and π creates basic limits (a la Coalgebras As Strict Lexcategory (Lemma 1.83)), this β is also a strict lexfunctor.

Not only that, but for each $t \in T$, the $\square_{C_1'}$-coalgebra $\beta(t) = H(\mathcal{N}_t)$ has the property that the following diagram commutes, as this diagram is H applied to the naturality square in T for $\mathcal{N} : \mathrm{id}_T \to \square_T$ with respect to the morphism $\mathcal{N}_t$:

$$\begin{array}{ccc} H(t) & \xrightarrow{H(\mathcal{N}_t)} & H(\square_T t) = \square_{C_1'} H(t) \\ \Big\downarrow{\scriptstyle H(\mathcal{N}_t)} & & \Big\downarrow{\scriptstyle H(\square_T \mathcal{N}_t) = \square_{C_1'} H(\mathcal{N}_t)} \\ H(\square_T t) = \square_{C_1'} H(t) & \xrightarrow[H(\mathcal{N}_{\square_T t}) = 4_{H(t)}]{} & H(\square_T \square_T t) = \square_{C_1'} H(\square_T t) = \square_{C_1'} \square_{C_1'} H(t) \end{array}$$

The commuting of the above diagram is the condition for $\beta(t)$ to be within the category Z.

Thus, β is indeed a strict lexfunctor from T to Z, such that $\pi \circ \beta = H$.

We have three more conditions to show to demonstrate that β is a strict introspective theory homomorphism. We must show it interacts in the appropriate way with C, $\mathcal{S}$, and $\mathcal{N}$.

The required condition on β with respect to C is that $\beta[C]$ should equal Z_2. Note that $\beta(C)$ is the lexcategory in Z corresponding to the internal lexfunctor $H(\mathcal{N}_C)$ in C_1'. By virtue of $H : T \to C_1'$ being a geminal category homomorphism, this $H(\mathcal{N}_C)$ is F_2', which is indeed our definition of the lexcategory Z_2 in Z, as required.

The required condition on β with respect to $\mathcal{S}$ is that the following diagram should commute:

$$\begin{array}{ccc} T & \xrightarrow{\beta} & Z \\ {\scriptstyle \mathcal{S}}\Big\downarrow & & \Big\downarrow{\scriptstyle \mathcal{S}} \\ \mathrm{Glob}(C) & \xrightarrow[\mathrm{Induced}(\beta, C)]{} & \mathrm{Glob}(\beta[C]) = \mathrm{Glob}(Z_2) \end{array}$$

By unwinding definitions and using the fact that H is a geminal category homomorphism, we find that both paths above yield the same result when applied to any object $t \in T$; specifically, this will be the global element of $\mathrm{Ob}(Z_2)$ whose underlying global element of $\mathrm{Ob}(C_2')$ is given by F_1' applied to $H(t)$. For the above diagram to furthermore commute as applied to any morphism in T, it suffices to know that following both paths with the projection $P : \mathrm{Glob}(Z_2) \to \mathrm{Glob}(C_2')$ commutes. By unwinding definitions, $P \circ \mathcal{S} \circ \beta = F_1' \circ H$ while $P \circ \mathrm{Induced}(\beta, C) \circ \mathcal{S} = \mathrm{Induced}(H, C) \circ \mathcal{S}$. As H is a geminal category homomorphism from T to C_1', these are indeed the same.

Finally, the required condition on β with respect to $\mathcal{N}$ is that $\beta(\mathcal{N}_t) = \mathcal{N}_{\beta(t)}$ for each $t \in T$. Note that $\beta(\mathcal{N}_t)$ is the coalgebra morphism in Z given by applying H to the naturality square in T for $\mathcal{N} : \mathrm{id}_T \to \Box_T$ with respect to the morphism $\mathcal{N}_t$. On the other hand, $\mathcal{N}_{\beta(t)}$ is the coalgebra morphism in Z from $\beta(t) = H(\mathcal{N}_t)$ to $4_{H(t)} = F'_2 = H(\mathcal{N}_{\Box_T t})$ whose underlying morphism in C'_1 is itself $\beta(t)$. Thus, $\beta(\mathcal{N}_t)$ and $\mathcal{N}_{\beta(t)}$ are both the same, both being the following commuting diagram (which we had already considered above):

$$\begin{array}{ccc}
H(t) & \xrightarrow{\quad H(\mathcal{N}_t) \quad} & H(\Box_T t) = \Box_{C'_1} H(t) \\
\downarrow{\scriptstyle H(\mathcal{N}_t)} & & \downarrow{\scriptstyle H(\Box_T \mathcal{N}_t) = \Box_{C'_1} H(\mathcal{N}_t)} \\
H(\Box_T t) = \Box_{C'_1} H(t) & \xrightarrow[H(\mathcal{N}_{\Box_T t}) = 4_{H(t)}]{} & H(\Box_T \Box_T t) = \Box_{C'_1} H(\Box_T t) = \Box_{C'_1} \Box_{C'_1} H(t)
\end{array}$$

This completes the demonstration that β is a strict introspective theory homomorphism, and thus completes the proof of the desired terminality property. ■

COROLLARY 5.26 As left adjoints preserve initial objects, the above tells us that the initial strict introspective theory (which we described in Theorem 5.20) is also the initial geminal category. Thus, the underlying lexcategory of the initial geminal category is the initial lexcategory with an internal geminal category!

CONSTRUCTION 5.27 Consider the functor from the category of geminal categories to the category of strict lexcategories with a designated internal geminal category, given by sending any geminal category G to the strict lexcategory $|G|$ with internal geminal category $\mathrm{InteriorGeminal}(G)$. This functor has a right adjoint.

In other words, for any strict lexcategory C_0 with an internal geminal category γ, there is a geminal category G equipped with a strict lexfunctor $H : |G| \to C_0$ satisfying $H[\mathrm{InteriorGeminal}(G)] = \gamma$, which is terminal among all so equipped geminal categories (in the sense that for any other such geminal category K with strict lexfunctor $J : |K| \to C_0$ satisfying $J[\mathrm{InteriorGeminal}(K)] = \gamma$, there is a unique geminal category homomorphism $M : K \to G$ such that $H \circ M = J$).

This co-free G admits an explicit description. Indeed, just as before, this explicit construction can be carried out internal to an arbitrary ambient lexcategory V (with C_0 as a V-internal lexcategory and γ as a $\mathrm{Glob}(C_0)$-internal geminal category, and constructing from these a terminal V-internal geminal category G with a V-internal lexfunctor $H : |G| \to C_0$ such that $H[\mathrm{InteriorGeminal}(G)] = \gamma$).

Construction details. For convenience, we will speak as though C_0 is internal to Set, but the following construction would work just as well were C_0 internal to any ambient lexcategory.

Let us use the names $\langle C_1, C_2, C_3, \ldots; F_1, F_2, F_3, \ldots \rangle$ to refer to the components of the C_0-internal geminal category γ.

Let $G_0 = C_0 \times \mathrm{Glob}(C_1)$. We have that $\gamma \times \mathrm{InteriorGeminal}(\gamma)$ is a geminal category $\langle G_1, G_2, G_3, \ldots; \phi_1, \phi_2, \phi_3, \ldots \rangle$ internal to G_0, with each $G_n = C_n \times C_{n+1}$ and $\phi_n = F_n \times F_{n+1}$, for $n \geq 1$. The terminal geminal category G which we construct will have $|G| = G_0$ and $\mathrm{InteriorGeminal}(G) = \gamma \times \mathrm{InteriorGeminal}(\gamma)$.

Specifically, let strict lexfunctor $\phi_0 : G_0 \to \mathrm{Glob}(G_1)$ be defined by $\phi_0(c_0, c_1) = (c_1, F_1(c_1))$. It's straightforward to then verify that $G = \langle G_0, G_1, G_2, \ldots; \phi_0, \phi_1, \phi_2, \ldots \rangle$ is a geminal category. The only nontrivial condition to verify is the equation $\mathrm{Induced}(\phi_0, G_1) \circ \phi_0 = \mathrm{Glob}(\phi_1) \circ \phi_0 : G_0 \to \mathrm{Glob}(\Gamma_{G_1}[G_2])$. Unwinding the definitions of ϕ_0 and ϕ_1, we find that the first component of this equation amounts to the tautology $F_1 = F_1$, while the second component of this equation amounts to the equation $\mathrm{Induced}(F_1, C_2) \circ F_1 = \Gamma_{C_1}[F_2] \circ F_1$ of the geminal category γ.

We also clearly have a projection strict lexfunctor H from $|G| = C_0 \times C_1$ to C_0, which satisfies $H[\mathrm{InteriorGeminal}(G)] = \gamma$. Having described the construction's details, we now prove that this construction has the stated terminality property:

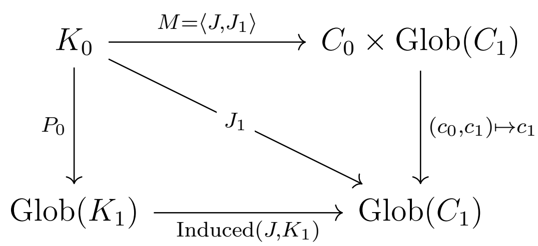
Proof. Suppose given any geminal category $K = \langle K_0, K_1, K_2, \ldots; P_0, P_1, P_2, \ldots \rangle$ and strict lexfunctor $J : K_0 \to C_0$ such that $J[\mathrm{InteriorGeminal}(K)] = \gamma$.

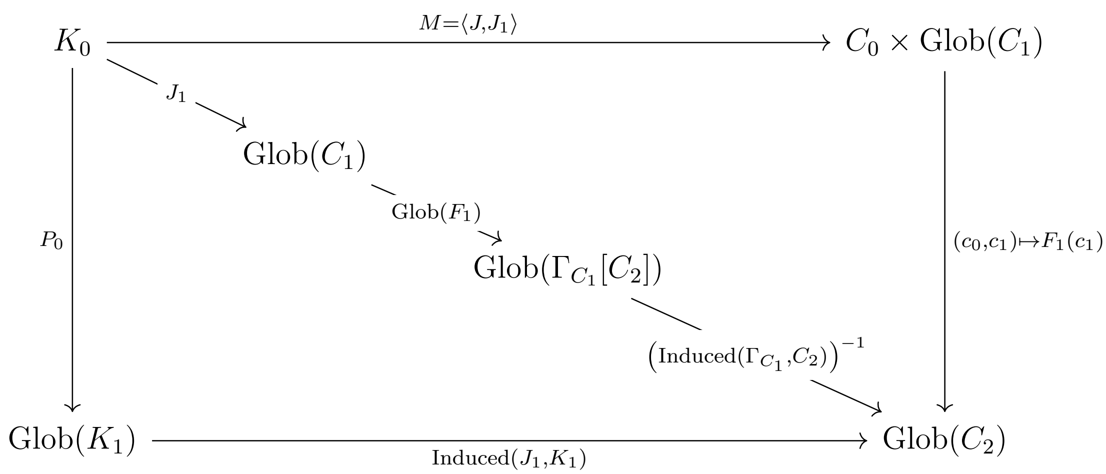
A strict lexfunctor M from $|K| = K_0$ to $|G| = C_0 \times \mathrm{Glob}(C_1)$ is given by a pair of strict lexfunctors $J_0 : K_0 \to C_0$ and $J_1 : K_0 \to \mathrm{Glob}(C_1)$. Since H is simply projection of the C_0 component, we will have that $H \circ M = J$ precisely when $J_0 = J$. Thus, specifying such M is determined by specifying J_1 alone. We must show that there is a unique J_1 making this M into a geminal category homomorphism from K to G.

Keeping in mind Definition 5.12, we see the conditions for such M to be a geminal category homomorphism. First of all, we must have that $M[\mathrm{InteriorGeminal}(K)] = \mathrm{InteriorGeminal}(G)$, which is to say, $J[\mathrm{InteriorGeminal}(K)] = \gamma$ (which has already been presumed) and $J_1[\mathrm{InteriorGeminal}(K)] = \mathrm{InteriorGeminal}(\gamma)$. On top of this, the final condition for M to be a geminal category homomorphism is that the following diagram commutes:

$$
\begin{array}{ccc}
K_0 & \xrightarrow{M = \langle J, J_1 \rangle} & C_0 \times \mathrm{Glob}(C_1) \\
\Big\downarrow{\scriptstyle P_0} & & \Big\downarrow{\scriptstyle \phi_0 \text{ [which is } (c_0, c_1) \mapsto (c_1, F_1(c_1))]} \\
\mathrm{Glob}(K_1) & \xrightarrow[\mathrm{Induced}(M, K_1)]{} & \mathrm{Glob}(C_1) \times \mathrm{Glob}(C_2)
\end{array}
$$

This diagram commutes just in case both of the following diagrams commute, which separately consider its $\mathrm{Glob}(C_1)$ and $\mathrm{Glob}(C_2)$ components:





In each of the above diagrams, the top-right triangle trivially commutes, so the commutativity condition for the overall square is equivalent to the commutativity of the bottom-left triangle.

From the diagram for the $\mathrm{Glob}(C_1)$ component, we see that J_1 is uniquely determined as $\mathrm{Induced}(J, K_1) \circ P_0$. All that remains is to verify that this choice of J_1 does indeed satisfy the condition $J_1[\mathrm{InteriorGeminal}(K)] = \mathrm{InteriorGeminal}(\gamma)$, as well as the condition of the commutative diagram for the $\mathrm{Glob}(C_2)$ component.

For the former condition, we have the chain of equations

$$
\begin{aligned}
&J_1[\mathrm{InteriorGeminal}(K)]\\
&= \mathrm{Induced}(J, K_1)[P_0[\mathrm{InteriorGeminal}(K)]]\\
&= \mathrm{Induced}(J, K_1)[\mathrm{InteriorGeminal}(\mathrm{InteriorGeminal}(K))]\\
&= \mathrm{InteriorGeminal}(J[\mathrm{InteriorGeminal}(K)])\\
&= \mathrm{InteriorGeminal}(\gamma)
\end{aligned}
$$

And as for the final commutativity condition, this follows like so:

$$\begin{tikzcd} K_0 \arrow[r, "P_0"'] \arrow[d, "P_0"'] \arrow[rr, bend left, "J_1"] & \mathrm{Glob}(K_1) \arrow[r, "\mathrm{Induced}(J, K_1)"'] \arrow[d, "\mathrm{Glob}(P_1)"] & \mathrm{Glob}(C_1) \arrow[d, "\mathrm{Glob}(F_1)"] \\ \mathrm{Glob}(K_1) \arrow[r, "\mathrm{Induced}(P_0, K_1)"] \arrow[rr, bend right, "\mathrm{Induced}(J_1, K_1)"'] & \mathrm{Glob}(\Gamma_{K_1}[K_2]) \arrow[r, "\mathrm{Induced}(J, \Gamma_{K_1}[K_2])"] & \mathrm{Glob}(\Gamma_{C_1}[C_2]) \end{tikzcd}$$

In the above commutative diagram, the top equation is our definition of J_1 as $\mathrm{Induced}(J, K_1) \circ P_0$, and the bottom equation follows from this definition as well. The left square commutes as part of the definition of K being a geminal category, and the right square commutes because $J[P_1] = F_1$ (which was part of our presumption that $J[\mathrm{InteriorGeminal}(K)] = \gamma$).

This completes the proof of the terminality property of G. ■

Chapter 6

Examples in the wild

6.1 Preview

In previous chapters, we have defined introspective theories and geminal categories. That is, we have axiomatized the theory of introspective theories and the theory of geminal categories. Now we look at some notable models of these axiomatic theories, which is to say, at some notable specific examples of introspective theories and of geminal categories. These examples are of a sort which might be considered to have been found "in the wild", instead of being freely syntactically constructed as the examples of the last chapter were.

There are two broad classes of models/examples of note in this chapter:

Firstly, there are those which are similar in flavor to the traditional instances of Gödelian phenomena studied in logic. These are based on logical theories which have some internal ability to discuss themselves, such as Peano Arithmetic, or higher-order intuitionistic logic, or the like. Here, it has long been recognized that Gödelian phenomena arise at the propositional level, but the full phenomenon of guarded recursion for types and terms which we proved for introspective theories in Theorem 4.18 and Löb's Theorem for Introspective Theories (Theorem 4.19) has not been noted in these contexts before. We also give an example of a model of this sort which goes well beyond computability or even countability, thus beyond many traditional approaches to presenting the Gödelian phenomena in logic.

The second class of models/examples we consider are based on presheaves over categories with a suitably well-founded subset of morphisms. Here, the existence of guarded recursion is straightforward, but it is the unification with our general theory which is of note. Among these models are examples like step-indexing in the topos of trees, the canonical model discussed in the literature on guarded recursion. Distinct from this are also models which capture the traditional interpretation of GL modal logic in well-founded transitive Kripke frames. We stress that we are able with these latter models to faithfully interpret traditional Boolean GL modal logic, unlike step-indexing in the topos of trees, whose non-Boolean logic validates such sentences as $\neg\neg\Box 0$ which are not theorems of GL.

6.2 The main initiality-based construction

CONSTRUCTION 6.1 Let $Special$ be a left comma-stable sub-2-category of LexCat, in the sense of Definition 1.51. Furthermore, suppose $Special$ has an initial object T, and that this T has an internal lexcategory C such that $\mathrm{Glob}(C)$ is itself an object of $Special$.

Then we obtain a unique lexfunctor $\mathcal{S} \in Special(T, \mathrm{Glob}(C))$, by the initiality of T.

Furthermore, by Theorem 1.56, we have that id_T is initial within $\mathrm{LexCat}(T, T)$. Thus, in particular, there is a unique natural transformation $\mathcal{N} : \mathrm{id}_T \to \mathrm{Hom}_C(1, \mathcal{S}(-))$. In this way, we obtain an introspective theory $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$.

THEOREM 6.2 Let $Special$ and $\langle T, C, \mathcal{S}, \mathcal{N}\rangle$ be given as in Construction 6.1 above.

Consider also any other introspective theory $\langle T', C', \mathcal{S}', \mathcal{N}'\rangle$ such that $\mathcal{S}' : T' \to \mathrm{Glob}(C')$ lives in $Special$. By the initiality of T within $Special$, we get a unique special lexfunctor $H : T \to T'$. If this H is such that $H[C] = C'$, then this H is also an introspective theory homomorphism (in the sense of Definition 5.10, suitably modified for the non-strict context). The condition that H interacts appropriately with $\mathcal{S}$ and $\mathcal{S}'$ is automatic by the initiality of T within $Special$. Furthermore, the condition that H interacts appropriately with $\mathcal{N}$ and $\mathcal{N}'$ is automatic by the fact that H is initial within $\mathrm{LexCat}(T, T')$, thanks to Theorem 1.56.

6.3 Self-initializing and super-initializing theories

6.3.1 The initial model as a geminal category

CONSTRUCTION 6.3 Suppose given some lexcategory Th (the theory of "gadgets"), along with a lexcategory C internal to Th (the underlying lexcategory of a gadget).

Furthermore, suppose given an initial gadget G_1 with an initial internal gadget G_2. That is, suppose given some lexcategory V such that $\mathrm{LexCat}(Th, V)$ has an initial object (our G_1) and such that $\mathrm{LexCat}(Th, \mathrm{Glob}(G_1[C]))$ has an initial object (our G_2).

Because G_1 is initial, we automatically get a unique homomorphism $F_1 : G_1 \to \Gamma[G_2]$. And because G_2 is an initial G-internal gadget, we automatically get a unique G_1-internal homomorphism $F_2 : G_2 \to \Gamma[G_3]$ where $G_3 = F_1[G_2]$.

This setup is thus a geminal gadget internal to V (with the equations E_1 and E_2 of Geminal category, compact presentation (Definition 5.15) automatically satisfied by the uniqueness observations in the previous paragraph).

Indeed, this is the unique way to equip $\langle G_1, G_2\rangle$ as a geminal gadget.

In practice, when an initial gadget has an initial internal gadget like above, this is usually not just some accident (caused by a paucity of globally defined structures, say), but rather, is due to the theory of gadgets itself encoding the construction of an internal initial gadget:

DEFINITION 6.4 Suppose, as above, given some lexcategory Th (the theory of "gadgets"), along with a lexcategory C internal to Th (the underlying lexcategory of a gadget).

If every gadget has an initial internal gadget, and every gadget homomorphism preserves these initial internal gadgets, then we say the theory of gadgets is **self-initializing**.

In other words, Th is self-initializing if $\mathrm{LexCat}(Th, \mathrm{Glob}(C))$ has an initial object, and this initiality is preserved by $\mathrm{Induced}(f, C)$ for every lexfunctor f out of Th.

Thus, Construction 6.3 shows us how to equip the initial model of any self-initializing theory as a geminal category.

The above all admits a generalization worth noting:

Construction 6.5 Suppose given some lexfunctor $i : Th \to Th'$, along with a lexcategory C internal to Th. Call Th the theory of "gadgets", and Th' the theory of "supergadgets". Via i, every supergadget has an underlying gadget, and via C, every gadget has an underlying lexcategory.

Furthermore, suppose given an initial gadget G_1 with an initial internal supergadget G_2. That is, suppose given some lexcategory V such that $\mathrm{LexCat}(Th, V)$ has an initial object (our G_1) and such that $\mathrm{LexCat}(Th', \mathrm{Glob}(G_1[C]))$ has an initial object (our G_2).

Because G_1 is initial, we automatically get a unique gadget homomorphism $F_1 : G_1 \to \Gamma[G_2]$. And because G_2 is an initial G_1-internal supergadget, we automatically get a unique G_1-internal supergadget homomorphism $F_2 : G_2 \to \Gamma[G_3]$ where $G_3 = F_1[G_2]$.

This setup is thus a geminal gadget internal to V (with the equations E_1 and E_2 of Geminal category, compact presentation (Definition 5.15) automatically satisfied by the uniqueness observations in the previous paragraph).

Indeed, this is the unique way to equip $\langle G_1, G_2 \rangle$ as a geminal gadget $\langle G_1, G_2; F_1, F_2 \rangle$ such that F_2 comes from a supergadget homomorphism.

And again, in practice, when an initial gadget has an initial internal supergadget like above, this is usually not just some accident caused by a paucity of globally defined structures, but rather, is due to the theory of gadgets itself encoding the construction of an internal initial supergadget:

Definition 6.6 Suppose, as above, given some lexfunctor $i : Th \to Th'$, along with a lexcategory C internal to T. We call T the theory of "gadgets", and Th' the theory of "supergadgets". Via i, every supergadget has an underlying gadget, and via C, every gadget has an underlying lexcategory.

If every gadget has an initial internal supergadget, and every gadget homomorphism preserves these initial internal supergadgets, then we say the theory of gadgets (or more precisely, the extension of the theory of gadgets by the theory of supergadgets) is **super-initializing**.

In other words, this situation is super-initializing if $\mathrm{LexCat}(Th', \mathrm{Glob}(C))$ has an initial object, and this initiality is preserved by $\mathrm{Induced}(f, C)$ for every lexfunctor f out of Th.

Note in this case that Th' will itself be self-initializing, as every supergadget is a fortiori a gadget (thus having an initial internal supergadget), and every supergadget homomorphism is a fortiori a gadget homomorphism (thus preserving initial internal supergadgets).

The self-initializing situation is of course the special case of the super-initializing situation where $Th' = Th$ and i is the identity.

There are a number of self- and super-initializing theories in the wild, which thus immediately give us examples of geminal categories in the wild.

For example: It is straightforward to show that every NNO-topos has internal initial models of every finitely axiomatizable lex theory, preserved by every NNO-topos homomorphism. This was the observation of Theorem 1.62

It is a little more difficult, but also possible to show that more generally, every arithmetic universe has internal initial models of every finitely axiomatizable lex theory, preserved by every arithmetic functor. This was the observation of Theorem 1.63.

Thus, any finitely axiomatizable extension of the theory of arithmetic universes is self-initializing. More generally, given any Th extending the theory of arithmetic universes, and any finitely axiomatizable Th' extending Th, the extension of Th to Th' is super-initializing.

This immediately gives us many examples of geminal categories using the above construction. For example, as one random example among myriad, we can obtain a geminal category $\langle G_1, G_2 \rangle$ where G_1 is the initial cartesian closed arithmetic universe and G_2 is its internal initial NNO-topos satisfying the internal axiom of choice.

We have discussed all this just in the context of geminal categories, but this extends to give analogous constructions of introspective theories as well. We discuss these next.

6.3.2 The theory of initial models as an introspective theory

Throughout the following, we say an initial object in a category of the form $\mathrm{LexCat}(T, S)$ is an initial model of T internal to S. Given lexcategories S and S' both containing initial internal models of T, we say a lexfunctor from S to S' preserves initial models of T if its composition with the initial model of T in S is the initial model of T in S'.

By the 2-category $\mathrm{InitialModels}(T)$, we mean LexCat with its objects restricted to just those lexcategories with initial internal models of T, and its 1-cells restricted to just those lexfunctors which preserve initial models of T. (The 2-cells remain unchanged.)

THEOREM 6.7 $\mathrm{InitialModels}(Th)$ has an initial object, whenever Th is a set-sized lexcategory.

Proof. This is in exactly the same way that we have familiar constructions such as of the initial NNO-topos, the initial arithmetic universe, the initial lexcategory with countable products, etc.

In more detail, the category of strict lexcategories with internal initial models of Th, and strict lexfunctors strictly preserving these internal initial models, is the category of models of an infinitary quasi-equational theory (whose infinitary operations have arity bounded by a set-sized cardinal dependent on the size of Th), and thus has an initial object. This initial strict structure furthermore is initial in the non-strict context, because all the relevant operations (finite limits, initial models of Th) are given by universal properties, so that any functor out of the initial strict structure preserving these in a non-strict sense is canonically isomorphic to a functor preserving these strictly on the nose. ∎

Construction 6.8 If Th is a self-initializing theory, then $\mathrm{InitialModels}(Th)$ is left comma-stable within LexCat, via Theorem 1.60. Furthermore, it has an initial object T via Theorem 6.7. This T by definition has an initial internal model of Th; that is, there is an initial $f \in \mathrm{LexCat}(Th, T)$. Furthermore, since Th is self-initializing, it contains an internal category C such that $\mathrm{Glob}(f[C])$ itself is an object of $\mathrm{InitialModels}(Th)$.

We can thus invoke Construction 6.1 to obtain a unique introspective theory $\langle T, f[C], \mathcal{S}, \mathcal{N}\rangle$ such that $\mathcal{S}$ is a map in $\mathrm{InitialModels}(Th)$.

There is an extension of the above construction to super-initializing theories. However, it is a bit trickier. The key issue is to construct, for a super-initializing lexfunctor $i : Th \to Th'$, a lexcategory which captures simultaneously the properties which are shared by initial models of Th and by initial models of Th'. We sketch out the construction as follows:

Construction 6.9 Let $i : Th \to Th'$ be a lexfunctor, such that models of Th are called gadgets, models of Th' are called supergadgets, and via i every supergadget is thought of as having an underlying gadget.

Given a lexcategory L, we will say that a "Th'-initial model of Th in L" is an internal gadget α in L, along with, for every internal supergadget β in L, a chosen gadget homomorphism from $h_\beta : \alpha \to \beta$, such that furthermore, these chosen homomorphisms are closed under postcomposition with supergadget homomorphisms (that is, for any supergadget homomorphism $f : \beta \to \beta'$ in L, we have that $f \circ h_\beta = h_{\beta'}$, as gadget homomorphisms). Note that this structure is NOT given by a universal property! There may be multiple non-equivalent ways to choose such structure within L. (In particular, both an initial internal gadget and an initial internal supergadget could be taken to be α.)

We may define a 2-category $Special$ whose objects are lexcategories along with a choice of Th'-initial models of Th in these lexcategories. The 1-cells of $Special$ will be lexfunctors which preserve these Th'-initial models of Th, in the sense of taking designated gadgets and homomorphisms to designated gadgets and homomorphisms. 2-cells between these are just ordinary natural transformations. There is an obvious forgetful 2-functor from this $Special$ to LexCat.

This $Special$ has an initial object A. This A contains a designated internal gadget G with an underlying lexcategory C. Because of the super-initializing property of gadgets and supergadgets, within $\mathrm{Glob}(C)$, there is an internal initial supergadget. Thus $\mathrm{Glob}(C)$ can uniquely be equipped as an object of $Special$ such that the designated gadget in $\mathrm{Glob}(C)$ is its internal initial supergadget. And thus there is a unique lexfunctor $\mathcal{S} : A \to \mathrm{Glob}(C)$ which takes G to the initial supergadget G' in $\mathrm{Glob}(C)$.

Finally, we shall show that there is a unique natural transformation $\mathcal{N} : \mathrm{id}_A \to \mathrm{Hom}_C(1, \mathcal{S}(-))$ such that action of this $\mathcal{N}$ restricted to G is a designated gadget homomorphism. The argument for this is completely analogous to Construction 6.1, with appropriate modification for the fact that Th'-initial models of Th are not given by a universal property and thus $Special$ is not merely a subcategory of LexCat.

Specifically, we will consider the comma category (id_A/f_Y) where $f_Y = \mathrm{Hom}_A(1, \mathcal{S}(-)) : A \to A$. Note that this f_Y, when applied to the designated gadget in A, yields the underlying gadget of a supergadget.

By a modification of The Comma-Kan Lemma (Theorem 1.47), we find a unique way to equip (id_A/f_Y) as having a Th'-initial model of Th which is preserved by both projections, such that the induced homomorphism on G is a designated homomorphism.

Then we may apply a modification of Theorem 1.56 to conclude that the identity on A has a unique natural transformation into f_Y whose induced action on G is a designated homomorphism.

6.3.3 A self-initializing theory with uncountable and uncomputable flavor

Note that all our arguments concerning self-initializing theories immediately adapt just as well when all instances of "lexcategory" are replaced by "lexcategory furthermore having countable products" and all instances of "lexfunctor" are replaced by "lexfunctor furthermore preserving countable products". We thus obtain the concept of self-initializing countably lex theories. The analogue of Construction 6.3 then shows us how to equip the initial model of any self-initializing countably lex theory as a geminal category, while the analogue of Construction 6.8 furthermore gives us a corresponding introspective theory.

In particular, the combination of Theorem 1.65 and Observation 1.67 tells us that the theory of toposes with countable products is a self-initializing countably lex theory. Thus:

Construction 6.10 The initial topos with countable products is naturally equipped as a geminal category. Indeed, there is a uniquely determined geminal category $\langle G_1, G_2; F_1, F_2 \rangle$ in which G_1 is the initial topos with countable products, G_2 is the initial topos with countable products internal to G_1, F_1 preserves the structure of a topos with countable products, and F_2 is in the internal logic of G_1 a map preserving the structure of a topos with countable products.

Note that this structure contains true arithmetic (in the sense of all true statements in the language of first-order arithmetic; note that determining membership in this set of statements is highly uncomputable), and has uncountable hom-sets (thus, too large to admit Gödel coding by mere natural numbers), but like any geminal category is still subject to Löb's theorem and Gödel's incompleteness theorems. There is no conflict between this structure being subject to Gödel's second incompleteness theorem and the fact that this structure contains all of true arithmetic, as this structure's consistency sentence isn't expressible in first-order arithmetic; that is, this structure is not definable in first-order arithmetic. Thus, this example demonstrates quite vividly that these Gödel-Löb phenomena are not constrained to having anything to do with computability or even countability.

6.4 The initial arithmetic universe

CONSTRUCTION 6.11 Let IAU be the initial arithmetic universe, and let C be any arithmetic universe internal to IAU. Then by the combination of Construction 6.1 and Corollary 1.55, we obtain an introspective theory $\langle \mathrm{IAU}, C, \mathcal{S}, \mathcal{N} \rangle$ in which $\mathcal{S} : \mathrm{IAU} \to \mathrm{Glob}(C)$ is the unique such arithmetic functor, and the natural transformation $\mathcal{N} : \mathrm{id}_T \to \mathrm{Hom}_C(1, \mathcal{S}(-))$ is uniquely determined.

OBSERVATION 6.12 Note that the above construction can be applied using ANY arithmetic universe internal to IAU. One natural choice is where C is taken to be the initial arithmetic universe IAU' internal to IAU (which exists thanks to Theorem 1.63 and Observation 1.66). In this case, the natural transformation $\mathcal{N}$ we obtain is the same as the one constructed in Lemma 5.15 of [DO20].

Note in this case also that $\mathcal{S} : \mathrm{IAU} \to \mathrm{Glob}(\mathrm{IAU}')$ is an equivalence. This is because the global sections functor $\mathrm{Hom}_{\mathrm{IAU}}(1, -) : \mathrm{IAU} \to \mathrm{Set}$ is an arithmetic functor, by Theorem 1.80. Thus, as arithmetic functors preserve the initial internal models obtained by Theorem 1.63, we have that $\mathrm{Glob}(\mathrm{IAU}') = \mathrm{IAU}$. Thus, the unique arithmetic functor from IAU to $\mathrm{Glob}(\mathrm{IAU}')$ [which we have taken as $\mathcal{S}$] is an equivalence.

Note that little was uniquely special about the initial arithmetic universe IAU for being able to be equipped as an introspective theory in this manner. It was just an invocation of our general construction Construction 6.1. We could similarly construct introspective theories using initial objects of any kind of structure left comma-stable over LexCat, given any structure of the same kind internal to the initial one. What's noteworthy about IAU is just that it happens to actually contain interesting internal structures (such as internal categories corresponding to Peano Arithmetic, to ZFC, to the initial arithmetic universe, etc), whereas the initial lexcategory, or initial regular category, or initial lexcategory with finite pullback-stable colimits, or such things, all have a paucity of interesting internal structures.

We now use Construction 6.11 to give a fuller account of our original guiding example of an introspective theory based on traditional logical theories, from Construction 2.10.

First, we must observe a lemmatic construction, on localizing introspective theories:

CONSTRUCTION 6.13 If $\langle T, C, \mathcal{S}, \mathcal{N} \rangle$ is an introspective theory, and $f : T \to T[M^{-1}]$ is a lex localization in the sense of Definition 1.73, and every morphism in M is sent to an isomorphism by $\mathcal{S}$, then f acts as an introspective theory homomorphism (in the sense of the non-strict analogue of Definition 5.10) from T to a uniquely determined introspective theory $\langle T[M^{-1}], f(C) \rangle$.

Furthermore, given any introspective theory homomorphism $h : \langle T, C \rangle \to \langle T_2, C_2 \rangle$ such that h sends every morphism in M to an isomorphism, this h factors uniquely through f by an introspective theory homomorphism from $\langle T[M^{-1}], f(C) \rangle$ to $\langle T_2, C_2 \rangle$. In this sense, $\langle T[M^{-1}], f(C) \rangle$ is the localization qua introspective theory of $\langle T, C \rangle$ at M.

In particular, for any introspective theory, we can apply the above taking M to be the set of all morphisms sent to isomorphisms by $\mathcal{S}$. We may call the result the **maximal localization** of our original introspective theory.

Proof. If $f : T \to T[M^{-1}]$ is to act as an introspective theory homomorphism, it must be to some introspective theory $\langle T[M^{-1}], C', \mathcal{S}', \mathcal{N}' \rangle$. We will show that each of these components are uniquely determined by the requirements of Definition 5.10.

The requirement on C' in Definition 5.10 directly determines it as $f(C)$.

The requirement on $\mathcal{S}'$ is that $\mathcal{S}' \circ f = \mathrm{Induced}(f, C) \circ \mathcal{S}$. Note that the right hand side of this equation sends all morphisms in M to isomorphisms (since $\mathcal{S}$ already does so). Thus, by the defining property of the localization $f : T \to T[M^{-1}]$, this uniquely determines $\mathcal{S}'$ as a functor, and indeed this $\mathcal{S}'$ will be a lexfunctor by Lemma 1.75.

Finally, the requirement on $\mathcal{N}'$ is that the whiskering $\mathcal{N}' f$ is equal to the whiskering $f\mathcal{N}$. By Lemma 1.72, this uniquely determines $\mathcal{N}'$.

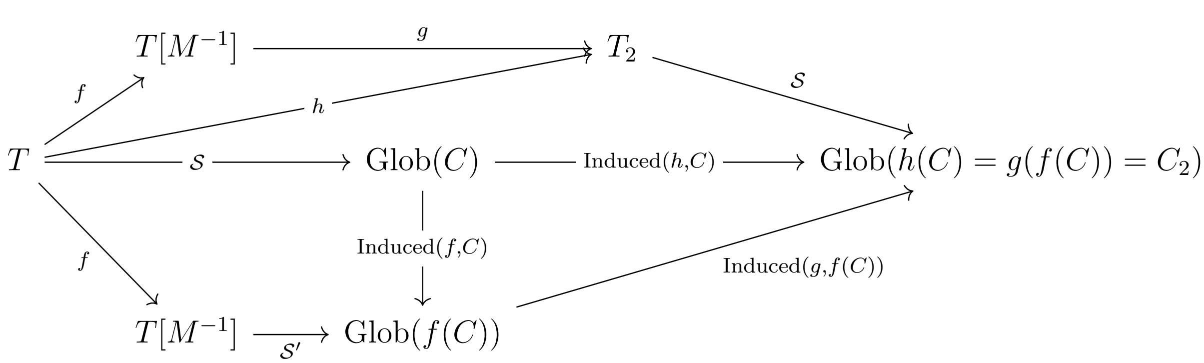
Next, we show the unique factorization property. Let h be a map as described. As $f : T \to T[M^{-1}]$ is a lex localization, we have that there is a unique lexfunctor $g : T[M^{-1}] \to T_2$ such that $h = g \circ f$. All that remains is to show that this g is an introspective theory homomorphism from $\langle T[M^{-1}, f(C), \mathcal{S}', \mathcal{N}' \rangle$ to $\langle T_2, C_2, \mathcal{S}_2, \mathcal{N}_2 \rangle$, in that it satisfies the conditions of Definition 5.10.

The introspective theory homomorphism condition $h(C) = C_2$ gives us the corresponding introspective theory homomorphism condition $g(f(C)) = C_2$.

The next condition we must establish is that this diagram commutes:

$$\begin{CD} T[M^{-1}] @>{g}>> T_2 \\ @V{\mathcal{S}'}VV @VV{\mathcal{S}}V \\ \mathrm{Glob}(f(C)) @>>{\mathrm{Induced}(g, f(C))}> \mathrm{Glob}(g(f(C))) = \mathrm{Glob}(C_2) \end{CD}$$

Because of the uniqueness of factorizations through localizations, it suffices to establish that both paths here become the same when preceded by the localization $f : T \to T[M^{-1}]$. And that can be seen via the following commuting diagram:



Finally, the last condition we must establish is that the whiskerings $\mathcal{N}g$ and $g\mathcal{N}'$ are equal. By Lemma 1.72, it suffices to establish that the whiskerings $\mathcal{N}gf$ and $g\mathcal{N}'f$ are equal. This can be seen via the chain of equations $\mathcal{N}gf = \mathcal{N}h = h\mathcal{N} = gf\mathcal{N} = g\mathcal{N}'f$. ■

THEOREM 6.14 The introspective theory described in Construction 2.10 is the maximal localization, in the sense of Construction 6.13, of an introspective theory produced by Construction 6.11.

Proof. Recall the categories Z and Z_{Σ_1} from Construction 2.10. Here, Z is an exact category whose objects and morphisms correspond to definable classes and graphs of functions between these in the theory ZF-Finite, with morphisms taken modulo provable equality in ZF-Finite. While Z_{Σ_1} is the subcategory of Z where the definability conditions are further restricted to Σ_1-definability.

It is readily verified that Z is an arithmetic universe. Thus, there is a unique arithmetic functor $!_Z : \mathrm{IAU} \to Z$. Let M be the set of morphisms in IAU which are taken to isomorphisms by this $!_Z$. By Lemma 1.78, this $!_Z$ factors uniquely through the arithmetic localization $\mathrm{IAU}[M^{-1}]$. Using Lemma 1.77, it is straightforwardly, if tediously, verified that this $\mathrm{IAU}[M^{-1}]$ is in fact Z_{Σ_1}, with $!_Z$ thus being the unique arithmetic functor from IAU to Z_{Σ_1} followed by the inclusion from Z_{Σ_1} to Z. That is to say, the role played by the Σ_1 constraints in defining Z_{Σ_1} is precisely to make Z_{Σ_1} an arithmetic localization of IAU.

Note also that, as Z and Z_{Σ_1} are both computably enumerable arithmetic universes internal to Set, we find, in keeping with Observation 1.45, that these are the images in Set of arithmetic universes internal to the initial arithmetic universe IAU. That is, letting $\mathrm{Glob}_{\mathrm{IAU}}$ be the unique arithmetic functor from IAU to Set (which is the same as the global sections functor $\mathrm{Hom}_{\mathrm{IAU}}(1, -)$, thanks to Theorem 1.80), we have arithmetic universes $\mathrm{Glob}_{\mathrm{IAU}}^{-1}[Z]$ and $\mathrm{Glob}_{\mathrm{IAU}}^{-1}[Z_{\Sigma_1}]$ such that the images of these under $\mathrm{Glob}_{\mathrm{IAU}}$ are Z and Z_{Σ_1}, respectively.

Via Construction 6.1, we thus obtain an introspective theory $\langle \mathrm{IAU}, \mathrm{Glob}_{\mathrm{IAU}}^{-1}[Z]\rangle$, whose $\mathcal{S} : \mathrm{IAU} \to \mathrm{Glob}(\mathrm{Glob}_{\mathrm{IAU}}^{-1}[Z]) = Z$ is the unique arithmetic functor from IAU to Z. Thus the set of morphisms in IAU sent to isomorphisms by this $\mathcal{S}$ is the same as the M defined above.

Now let $\langle Z_{\Sigma_1}, Z'\rangle$ be the introspective theory described in Construction 2.10.

It is readily verified that Z′ and $\mathrm{Glob}_{\mathrm{IAU}}^{-1}[Z]$ can be chosen so that the former is the image of the latter under the unique arithmetic functor $!_{Z_{\Sigma_1}} : \mathrm{IAU} \to Z_{\Sigma_1}$. Furthermore, it is readily verified that $\mathcal{S} : Z_{\Sigma_1} \to \mathrm{Glob}(Z')$ is an arithmetic functor. Thus by Construction 6.1, the unique arithmetic functor $!_{Z_{\Sigma_1}} : \mathrm{IAU} \to Z_{\Sigma_1}$ is in fact an introspective theory homomorphism from $\langle \mathrm{IAU}, \mathrm{Glob}_{\mathrm{IAU}}^{-1}[Z]\rangle$ to $\langle Z_{\Sigma_1}, Z'\rangle$.

Since $!_{Z_{\Sigma_1}} : \mathrm{IAU} \to Z_{\Sigma_1}$ was, as noted above, the same as the arithmetic localization $\mathrm{IAU} \to \mathrm{IAU}[M^{-1}]$, we may invoke Construction 6.13 to conclude that the introspective theory homomorphism from $\langle \mathrm{IAU}, \mathrm{Glob}_{\mathrm{IAU}}^{-1}[Z]\rangle$ to $\langle Z_{\Sigma_1}, Z'\rangle$ is the same as the localization of the introspective theory $\langle \mathrm{IAU}, \mathrm{Glob}_{\mathrm{IAU}}^{-1}[Z]\rangle$ at M, which by the observation three paragraphs ago is the maximal localization of this introspective theory.

This concludes the proof. ■

OBSERVATION 6.15 Clearly, there is nothing special about ZF-Finite in the above. From any traditional computably enumerable logical theory extending, say, Peano Arithmetic (though even this is much stronger than necessary), we get a computably enumerable arithmetic universe in the style of Z', which (a la Observation 1.45) is coded by some arithmetic universe C internal to IAU. For example, we can do this with ZFC, or vNBG, or ZFC + "ZFC is not consistent", or any such thing. For each of these, we get correspondingly an introspective theory $\langle \mathrm{IAU}, C\rangle$ via Construction 6.11, whose maximal localization (in the sense of Construction 6.13) is perfectly analogous to the introspective theory $\langle \mathrm{Z}_{\Sigma_1}, \mathrm{Z}'\rangle$ from Construction 2.10.

6.5 Models based on presheaf categories

In this section, we will develop an introspective theory construction which unifies and vastly generalizes Construction 2.13 and Construction 2.16.

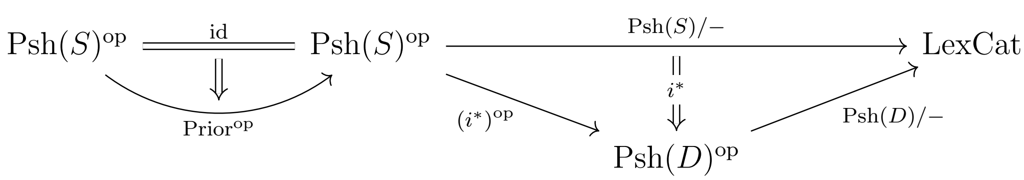
6.5.1 The general construction yielding locally introspective theories

CONSTRUCTION 6.16 Let $i : D \to S$ be an arbitrary functor between set-sized categories D and S.

Furthermore, suppose given some subset of the morphisms of S which is closed under composition on either side with arbitrary morphisms. That is, suppose given some bifunctor $\mathrm{SpecialHom}_S : S^{op} \times S \to \mathrm{Set}$ along with an inclusion map from $\mathrm{SpecialHom}_S$ to $\mathrm{Hom}_S : S^{op} \times S \to \mathrm{Set}$.[1]

By currying, we may read this $\mathrm{SpecialHom}_S$ as a functor from S to $\mathrm{Psh}(S)$ which is a subfunctor of the Yoneda embedding. As the Yoneda embedding exhibits $\mathrm{Psh}(S)$ as the free cocompletion of S under set-sized colimits, we can uniquely extend this subfunctor of the Yoneda embedding to an endofunctor Prior on $\mathrm{Psh}(S)$ which is a subfunctor of the identity and which preserves set-sized colimits. By the adjoint functor theorem, this Prior is a left adjoint.

We may now define a locally introspective theory $\langle \mathrm{Psh}(S), C\rangle$ like so:



[1] This construction would work just as well for any bifunctor $\mathrm{SpecialHom}_S$ with a map to Hom_S. It is not actually necessary that this map be monic. We use the language of "subfunctor" and "inclusion" here just for linguistic convenience, and because our archetypal examples happen to be of this form.

The unlabelled 2-cell on the left of the above diagram is the one given by the inclusion of Prior into the identity (which is turned around when considered as endodunctors of $\mathrm{Psh}(S)^{\mathrm{op}}$ instead of $\mathrm{Psh}(S)$).

Here, we take the $\mathrm{Psh}(S)$-indexed lexcategory C to be the bottom composite path, and of course our $\mathcal{F}$ is the composite 2-cell from top to bottom. The local representableness of this C follows via Theorem 1.20 from the observations that Prior and i^* both have right adjoints and $\mathrm{Psh}(D)/-$ is locally representable (i.e., $\mathrm{Psh}(D)$ is locally cartesian closed).

(Pedantically, we note that in the above diagram, LexCat must be understood as containing not just set-sized lexcategories but also large lexcategories, so that this LexCat may serve as the target of the self-indexings of the large categories $\mathrm{Psh}(S)$ and $\mathrm{Psh}(D)$.)

This construction can be seen as Example 2.3 applied to $\mathrm{Psh}(D)$, followed by Construction 2.18 using the pullback-preserving functor $i^* : \mathrm{Psh}(S) \to \mathrm{Psh}(D)$, followed by Construction 2.17 using a map derived from the inclusion of Prior into the identity on $\mathrm{Psh}(S)$.

Observation 6.17 Our archetypal examples of locally introspective theories Construction 2.12 and Construction 2.14 were each instances of Construction 6.16.

Specifically, Construction 2.12 was the instance where i is the identity functor on the poset ω of natural numbers, with $\mathrm{SpecialHom}(a, b)$ being uniquely inhabited when $a < b$ and otherwise empty.

And Construction 2.14 was the instance where i is the inclusion of $|P|$ into Q (where P is an arbitrary poset, $|P|$ is its underlying discrete set, and Q is P augmented with a new maximum element), and again $\mathrm{SpecialHom}(a, b)$ was taken to be uniquely inhabited when $a < b$ and otherwise empty.

Unfortunately, this Construction 6.16 does not in general yield a fully introspective theory. We cannot expect this C to be representable.

But by passing to suitable full sublexcategories of $\mathrm{Psh}(S)$ and $\mathrm{Psh}(D)/-$ a la Construction 2.20 and Construction 2.21, we may hope to obtain an introspective theory, and indeed we shall always be able to do so in a convenient way whenever $\mathrm{SpecialHom}$ satisfies a certain well-foundedness condition. The details of this process are described in the next sections.

6.5.2 Presheaves with varying cardinality constraints (aka, ramps)

The details in this section may seem like a lot. Bear with me! We are simply abstracting the same kind of cardinality constraints used in Construction 2.13 and Construction 2.16. Note that this amounts to a generalization of the construction from [HS99], to allow for constraints which vary over the objects of the indexing category rather than remaining constant.

Let X be a category internal to Set, and let K be a function from $\mathrm{Ob}(X)$ to sets of sets[2]. We call any choice of such K a **ramp** on X.

By a K-presheaf on X, we mean a presheaf E on X such that $E(x) \in K(x)$ for each $x \in X$. These comprise a full subcategory of $\mathrm{Psh}(X)$, which we may call $\mathrm{Psh}_K(X)$.

More generally, given an arbitrary presheaf P on X, we define the notion of a K-**decoration** on P. This is a function f which assigns to each $x \in X$ and $p \in P(x)$ an element $f(x,p) \in K(x)$, along with a presheaf E on X and a map $\pi : E \to P$ in $\mathrm{Psh}(X)$ satisfying the condition that $E(x) = \coprod_{p \in P(x)} f(x,p)$ with $\pi_x : E(x) \to P(x)$ being the corresponding projection, for each $x \in X$. (Thus, all that is left to specify in E is its reindexing maps, in a compatible fashion with the morphism structure of X and with the projection map to P.)

That is, a K-decoration of P is a presheaf over P whose fibers at each aspect $P(x)$ of P are each given by elements of the corresponding $K(x)$. Note that the K-decorations of P comprise a set (they can be compared for equality, and it is readily observed that in size they comprise a set rather than a proper class).

Note that K-decorations of the terminal presheaf amount to the same thing as objects of $\mathrm{Psh}_K(X)$.

Given a map $m : P_1 \to P_2$ in $\mathrm{Psh}(X)$, we straightforwardly can pull a K-decoration of P_2 back to a K-decoration of P_1. This is strictly functorial, and thus we get a contravariant map from $\mathrm{Psh}(X)$ to Set which assigns to any $P \in \mathrm{Psh}(X)$ the set of K-decorations of P, with reindexings as just described.

It is also straightforward to observe that this map takes set-sized colimits in $\mathrm{Psh}(X)$ to set-sized limits in Set. Thus, by Theorem 1.84, this is in fact $\mathrm{Psh}(X)$-representable. That is to say, we have a particular object $K{-}\mathrm{Dec}(X)$ in $\mathrm{Psh}(X)$ and a K-decoration of $K{-}\mathrm{Dec}(X)$, such that any K-decoration of any object in $\mathrm{Psh}(X)$ is the reindexing of this one along a unique morphism. In particular, this gives us a map into $K{-}\mathrm{Dec}(X)$ such that every other map with K-sized fibers, so to speak, is a pullback of this one.

Construction 6.18 We may now consider the full subcategory of the self-indexing $\mathrm{Psh}(X)/-$ restricted to pullbacks of this generic map with K-sized fibers. Call this $\mathrm{RPsh}_K(X)$. This $\mathrm{Psh}(X)$-indexed category is locally representable (as it is a full subcategory of $\mathrm{Psh}(X)/-$, which is locally representable by the local cartesian closure of $\mathrm{Psh}(X)$), and it can also be taken as having a representable set of objects (given by $K{-}\mathrm{Dec}(X)$). Thus, it can be taken to be a $\mathrm{Psh}(X)$-internal category, which is to say, an X-indexed set-sized strict category.

Note that the global aspect of $\mathrm{RPsh}_K(X)$ is $\mathrm{Psh}_K(X)$, by our previous observation about K-decorations of the terminal presheaf.

Observe that if each $K(x)$ for $x \in X$ is closed under finite products and subsets, then each $K(x)$ comprises the objects of a full sublexcategory of Set, and $\mathrm{RPsh}_K(X)$ is in fact a full sublexcategory of the self-indexing $\mathrm{Psh}(X)/-$, with $\mathrm{Psh}_K(X)$ accordingly being a

[2]It might be better to say that the outputs of K are set-indexed sets. At any rate, we shall think of each $K(x)$ as a genuine set, whose elements both can be compared for equality and have associated sets.

full sublexcategory of $\mathrm{Psh}(X)$. From now on, we will always make this assumption on our ramps K. [It would actually suffice for our purposes to make the slightly weaker presumption that each $K(x)$ comprises the objects of a full sublexcategory of Set, but for convenience, we go ahead and presume here closure under arbitrary subsets.]

When ramps are closed under finite limits in this way, observe that any map in $\mathrm{Psh}_K(X)$ is in fact a K-decoration (or rather, isomorphic to one as an object of the corresponding slice category). That is, given $f : Q \to P$ in $\mathrm{Psh}_K(X)$, we have for each $x \in X$ and each $p \in P(x)$ that the fiber $f^{-1}(p)$ is in $K(x)$ (as it can be defined as a pullback $1 \times_{P(x)} Q(x)$). Thus, the self-indexing $\mathrm{Psh}_K(X)/-$ is an indexed full sublexcategory of $\mathrm{RPsh}_K(X)$.[3]

Observation 6.19 Observe that the conditions defining a K-decoration of $P \in \mathrm{Psh}(X)$ only depend on the values of $K(x)$ at $x \in X$ for which $P(x)$ is inhabited. The values of K at x for which $P(x)$ is empty play no role. In other words, the aspect of the $\mathrm{Psh}(X)$-internal category $\mathrm{RPsh}_K(X)$ at presheaf P only depends on the value of K at $x \in X$ for which $P(x)$ is inhabited.

When we consider $\mathrm{RPsh}_K(X)$ instead as an X-indexed structure, its aspect at $x \in X$ is the same as its aspect as a $\mathrm{Psh}(X)$-internal structure at $\mathrm{yoneda}(x)$. By the above, this only depends on the values of K at y for which $\mathrm{yoneda}(x)(y) = \mathrm{Hom}_X(y, x)$ is inhabited.

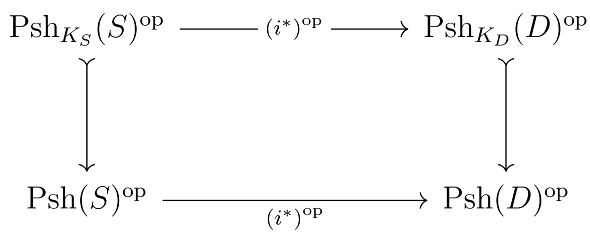
6.5.3 Having ramps on two categories

Observe that if we are given an arbitrary functor $i : D \to S$, along with ramps K_S on S and K_D on D satisfying the compatibility condition that $K_S(i(d)) \subseteq K_D(d)$ for each $d \in D$, then applying $i^* : \mathrm{Psh}(S) \to \mathrm{Psh}(D)$ to a presheaf in $\mathrm{Psh}_{K_S}(S)$ yields a presheaf in $\mathrm{Psh}_{K_D}(D)$. We may refer to this restricted action by the same name $i^* : \mathrm{Psh}_{K_S}(S) \to \mathrm{Psh}_{K_D}(D)$. Given our presumption of closure under finite limits on the ramps, this is a lexfunctor between lexcategories.

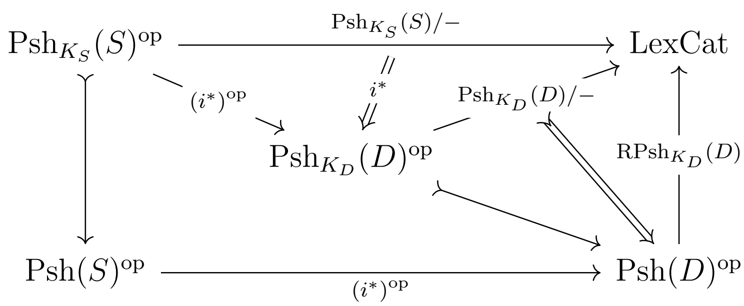
Definition 6.20 For convenience, we may go ahead and even define K_D as $K_S \circ i$ to ensure the compatibility condition, though this is stronger than needed for our purposes. To simplify our exposition, we will from now on presume K_D is defined from K_S in this way.

We thus have the following commutative square of lexfunctors, where the unlabelled arrows are the inclusion lexfunctors:

[3] This inclusion needn't be an equivalence. There may be further K-decorations of $P \in \mathrm{Psh}_K(X)$ whose domain (qua morphism into P) is not in $\mathrm{Psh}_K(X)$, as we've made no presumption that each $K(x)$ be closed under sums indexed by any set in $K(x)$. But we have this one direction of inclusion.



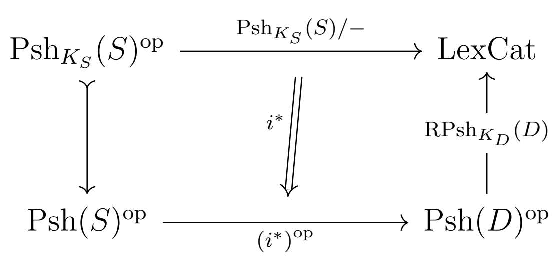
Indeed, this commutative diagram sits within the following diagram of lexfunctors and indexed lexfunctors:



In the above diagram, unlabelled arrows are canonical inclusions of full sublexcategories. In particular, the unlabelled right 2-cell is the way in which $\mathrm{Psh}_{K_D}(D)/-$ is an indexed full sublexcategory of $\mathrm{RPsh}_{K_D}(D)$.

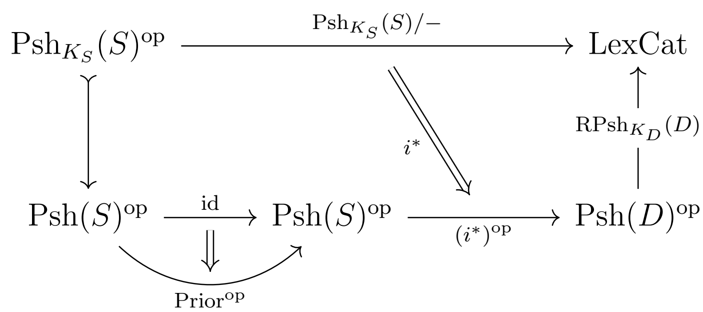
The top 2-cell i^* is the one obtained from $i^* : \mathrm{Psh}_{K_S}(S) \to \mathrm{Psh}_{K_D}(D)$ by Construction 1.85. The bottom left "triangle" (or square drawn as triangle) is our just previously mentioned commutative square.

Overloading names yet again, we may compress this last diagram into a composite 2-cell which we shall also name i^*, like so:



6.5.4 Cardinality-constraining the general construction to yield introspective theories

Construction 6.21 From the above, we have a pre-introspective theory $\langle \mathrm{Psh}_{K_S}(S), C, \mathcal{F} \rangle$ in which $\mathcal{F}$ is given by the composite 2-cell in the following diagram, and C is given by the codomain of this $\mathcal{F}$:



Observation 6.22 Our archetypal examples of introspective theories Construction 2.13 and Construction 2.16 were both instances of Construction 6.21, in the same manner as in Observation 6.17.

We now consider the question of when the pre-introspective theory described in Construction 6.21 is in fact an introspective theory. This happens precisely if the C we have defined is $\mathrm{Psh}_{K_S}(S)$-representable. As $\mathrm{RPsh}_{K_D}(D)$ is $\mathrm{Psh}(D)$-representable (by the comments at Construction 6.18), and i^* and Prior both have right adjoints, we automatically have (by Theorem 1.10) that our C corresponds to a lexcategory internal to $\mathrm{Psh}(S)$ (aka, an S-indexed lexcategory). However, the inclusion of $\mathrm{Psh}_{K_S}(S)$ into $\mathrm{Psh}(S)$ will not in general have a right adjoint, so we cannot conclude that C is $\mathrm{Psh}_{K_S}(S)$-representable in this same way. We may think of $\mathrm{Mor}(C)$ as an object within $\mathrm{Psh}(S)$, but do not know that this lives within its full subcategory $\mathrm{Psh}_{K_S}(S)$. This will happen just in case, for every $s \in S$, the s-aspect of $\mathrm{Mor}(C)$ is contained in $K_S(s)$.

Under suitable conditions, we can arrange for a ramp K_S such that this happens. For $t, s \in S$, let $t < s$ mean that $\mathrm{SpecialHom}_S(t, s)$ is inhabited. We have the following:

Observation 6.23 The s-aspect of C, where $s \in S$, depends only on the values of the ramp K_S at objects $< s$.

Proof. By definition, the s-aspect of C is the aspect of $\mathrm{RPsh}_{K_D}(D)$ at $i^*(\mathrm{Prior}(\mathrm{yoneda}(s)))$. By Observation 6.19, this depends only on the value of K_D at those objects in D at which the presheaf $i^*(\mathrm{Prior}(\mathrm{yoneda}(s)))$ is inhabited.

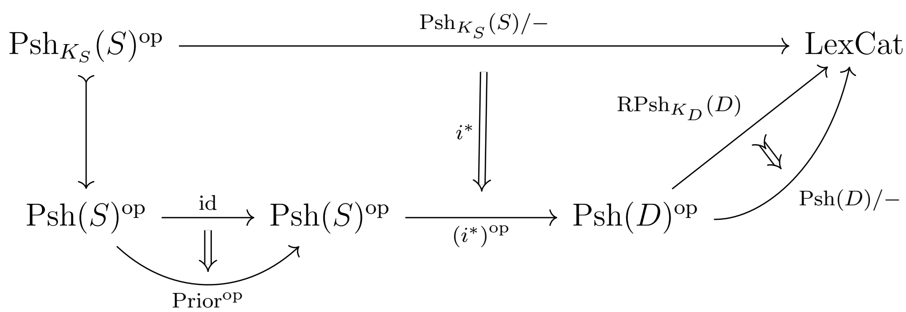
And by definition, $\mathrm{Prior}(\mathrm{yoneda}(s))$ is a presheaf on S which is inhabited only at $t < s$. Thus, $i^*(\mathrm{Prior}(\mathrm{yoneda}(s)))$ is a presheaf on D which is inhabited only at d for which $i(d) < s$.

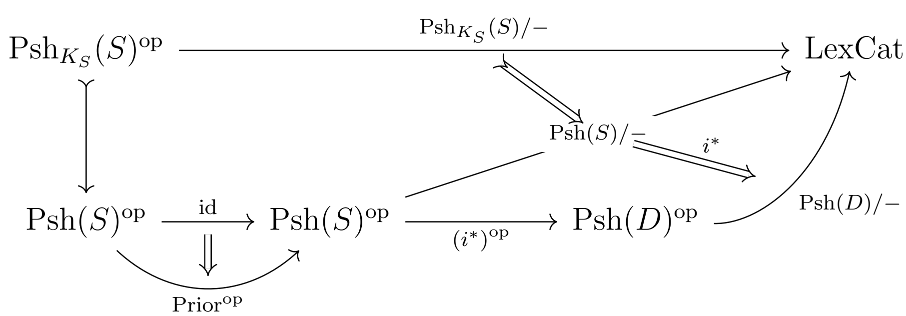
Thus, the s-aspect of C depends only the values of the ramp K_D at $d \in D$ for which $i(d) < s$. Since $K_D(d)$ was defined (at Definition 6.20) as $K_S(i(d))$, this is to say that the dependence is only on the values of the ramp K_S at objects $< s$. ■

Thus, we will have that Construction 6.21 gives an introspective theory just in case for each $s \in S$, we have that $K_S(s)$ contains a certain set (the s-aspect of $\mathrm{Mor}(C)$) determined by the values of $K_S(t)$ at $t < s$. If the $<$ relation is well-founded, we can easily recursively choose the values of K_S to always satisfy this condition, by simply choosing at each stage the minimal full sublexcategory of Set generated by the required set. (Indeed, we could just as well arrange in essentially the same way for each $K_S(s)$ to also contain any set-sized number of other sets of interest.)

Construction 6.24 Thus, when the $<$ relation corresponding to $\mathrm{SpecialHom}_S$ is well-founded, we can choose a ramp K_S such that Construction 6.21 yields an introspective theory (and we can furthermore do so in such a way that $\mathrm{Psh}_{K_S}(S)$ includes any other fixed set of desired objects of $\mathrm{Psh}(S)$).

Observation 6.25 The introspective theory of Construction 6.24 is straightforwardly a sub-introspection, in the sense of Construction 2.22, of the locally introspective theory given by Construction 6.16. This can be seen via the following two diagrams:





In the above two diagrams, unlabelled arrows are canonical inclusions. Note that the composite 2-cells from $\mathrm{Psh}_{K_S}(S)/-$ to $\mathrm{Psh}(D)/i^*(-)$ are the same in both diagrams.

Bibliography

[Bir+11] Lars Birkedal et al. "First Steps in Synthetic Guarded Domain Theory: Step-Indexing in the Topos of Trees". In: *2011 IEEE 26th Annual Symposium on Logic in Computer Science*. IEEE. 2011, pp. 55–64.

[BM13] Lars Birkedal and Rasmus Ejlers Møgelberg. "Intensional Type Theory with Guarded Recursive Types qua Fixed Points on Universes". In: *2013 28th Annual ACM/IEEE Symposium on Logic in Computer Science*. 2013, pp. 213–222. DOI: 10.1109/LICS.2013.27.

[Boo95] George Boolos. *The Logic of Provability*. Cambridge University Press, 1995.

[Bor94] Francis Borceux. *Handbook of Categorical Algebra 1 – Basic Category Theory*. Cambridge Univ. Press, 1994.

[DO20] Joost van Dijk and Alexander Gietelink Oldenziel. "Gödel Incompleteness through Arithmetic Universes after A. Joyal". In: *arXiv preprint arXiv:2004.10482* (2020).

[Göd31] Kurt Gödel. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". In: *Monatshefte für mathematik und physik* 38 (1931), pp. 173–198.

[Gra+21] Daniel Gratzer et al. "Multimodal Dependent Type Theory". In: *Logical Methods in Computer Science* Volume 17, Issue 3 (2021). ISSN: 1860-5974. DOI: 10.46298/lmcs-17(3:11)2021. URL: http://dx.doi.org/10.46298/lmcs-17(3:11)2021.

[GZ67] P. Gabriel and M. Zisman. *Calculus of fractions and homotopy theory*. Springer, 1967.

[HS99] Martin Hofmann and Thomas Streicher. "Lifting Grothendieck Universes". In: (1999). URL: https://www2.mathematik.tu-darmstadt.de/~streicher/NOTES/lift.pdf.

[Jac99] Bart Jacobs. *Categorical Logic and Type Theory*. Elsevier, 1999.

[KS06] Masaki Kashiwara and Pierre Schapira. *Categories and Sheaves*. Springer Berlin Heidelberg, 2006. DOI: 10.1007/3-540-27950-4. URL: https://link.springer.com/book/10.1007/3-540-27950-4.

[Law69] F William Lawvere. "Diagonal Arguments and Cartesian Closed Categories". In: *Category theory, Homology Theory and their Applications II*. Springer, 1969, pp. 134–145.

[Löb55] Martin Hugo Löb. "Solution of a problem of Leon Henkin". In: *The Journal of Symbolic Logic* 20.2 (1955), pp. 115–118.

[Mai03] Maria Emilia Maietti. "Joyal's arithmetic universes via type theory". In: *Electronic Notes in Theoretical Computer Science* 69 (2003). CTCS'02, Category Theory and Computer Science, pp. 272–286. ISSN: 1571-0661. DOI: https://doi.org/10.1016/S1571-0661(04)80569-3. URL: https://www.sciencedirect.com/science/article/pii/S1571066104805693.

[Mai05] Maria Emilia Maietti. "Reflection Into Models of Finite Decidable FP-sketches in an Arithmetic Universe". In: *Electronic Notes in Theoretical Computer Science* 122 (2005). Proceedings of the 10th Conference on Category Theory in Computer Science (CTCS 2004), pp. 105–126. ISSN: 1571-0661. DOI: https://doi.org/10.1016/j.entcs.2004.06.054. URL: https://www.sciencedirect.com/science/article/pii/S1571066105000356.

[Mai10] Maria Emilia Maietti. "Joyal's arithmetic universe as list-arithmetic pretopos". In: *Theory and Applications of Categories* 24 (2010), pp. 39–83.

[nLa23] nLab authors. *tiny object*. https://ncatlab.org/nlab/show/tiny+object#in_presheaf_categories. Revision 39. Jan. 2023.

[PT89] Andrew M. Pitts and Paul Taylor. "A Note on Russell's Paradox in Locally Cartesian Closed Categories". In: *Studia Logica: An International Journal for Symbolic Logic* 48.3 (1989), pp. 377–387. ISSN: 00393215, 15728730. URL: http://www.jstor.org/stable/20015448.

[PV07] E. Palmgren and S.J. Vickers. "Partial Horn logic and cartesian categories". In: *Annals of Pure and Applied Logic* 145.3 (2007), pp. 314–353. ISSN: 0168-0072. DOI: https://doi.org/10.1016/j.apal.2006.10.001. URL: https://www.sciencedirect.com/science/article/pii/S0168007206001229.

[Rob21] David Michael Roberts. *Substructural fixed-point theorems and the diagonal argument: theme and variations*. 2021. arXiv: 2110.00239 [math.CT].

[Vic20] Steve Vickers. *Arithmetic universes: Home of free algebras*. https://www.cs.bham.ac.uk/~sjv/PalmgrenMem.pdf. 2020.