

OBVIOUS NATURAL MORPHISMS OF SHEAVES ARE UNIQUE

RYAN COHEN REICH

ABSTRACT. We prove that a large class of natural transformations (consisting roughly of those constructed via composition from the “functorial” or “base change” transformations) between two functors of the form $\cdots f^*g_* \cdots$ actually has only one element, and thus that any diagram of such maps necessarily commutes. We identify the precise axioms defining what we call a “geofibered category” that ensure that such a coherence theorem exists. Our results apply to all the usual sheaf-theoretic contexts of algebraic geometry. The analogous result that would include any other of the six functors remains unknown.

Commutative diagrams express one of the most typical subtle beauties of mathematics, namely that a single object (in this case an arrow in some category) can be realized by several independent constructions. The more interesting the constructions, the more insight is gained by carefully verifying commutativity, and it is tempting to take the inverse claim to mean that comparable arrows, constructed mundanely, should be expected to be equal and are therefore barely worth proving to be so. As half the purpose of a human-produced publication is to provide insight, there is some validity in thinking that any opportunity to reduce both length and tedium is worth taking, but this is at odds with the other half, which is to supply rigor. The goal of this paper is, therefore, to serve both ends by proving that a large class of diagrams commonly encountered in algebraic geometry, obtained from the fibered category nature of sheaves, are both interesting and necessarily commutative.

The goal of proving “all diagrams commute” began with the first “coherence theorem” of this kind, proved by Mac Lane [ML63]; further research, apparently, has not yet developed this particular application, as we are not aware of any coherence theorems that apply to fibered categories. Indeed, our main result [Theorem 2.4](#) is itself not entirely free of conditions, and our best unconditional result such as in [Theorem 2.3](#) is somewhat restricted in scope. A similar statement to the latter was obtained by Jacob Lurie [Lur05] concerning *arbitrary* tensor isomorphisms of functors on coherent sheaves over a stack; ours applies to less general morphisms but more general situations. The problem of proving a general coherence theorem for pullbacks and pushforwards in the context of *monoidal* functors was mentioned in Fausk–Hu–May [FHM03] and said to be both unsolved and desirable; we do not claim to have solved it, though certainly, we have done something. Those authors do not consider multiple maps f in combination, however, as we do.

This research was partially supported by NSF RTG grant DMS 0943832.

Received by the editors 2013-10-20 and, in revised form, 2014-04-05.

Transmitted by Ross Street. Published on 2014-04-15.

2010 Mathematics Subject Classification: Primary 14A15; Secondary 18D30, 18A25.

Key words and phrases: commutative diagrams, coherence theorem, string diagrams, pullback, pushforward.

© Ryan Cohen Reich, 2014. Permission to copy for private use granted.

The goal of rigorously verifying the diagrams of algebraic geometry has been pursued by several people, notably recently Brian Conrad in [Con00], who proved the compatibility of the trace map of Grothendieck duality with base change. Our theorems do not seem to apply to this problem, most significantly because almost every construction in that book intimately refers to details of abelian and derived categories and to methods of homological algebra, none of which are addressed here. Such a quality is possessed by many constructions of algebraic geometry and place many of its interesting compatibilities potentially out of reach of the results stated here.

Maps constructed in such a way, as concerns Conrad constantly in that book, can certainly differ by a natural sign, if not worse, and so the only hope for automatically proving commutativity of that kind of diagram is to separate the cohomological part from the categorical part and apply our results only to the latter. This is to say nothing of the explicit avoidance of “good” hypotheses (for example, flatness, such as we will consider) on the schemes and morphisms considered there, all of which place this particular compatibility on the other side of the line separating “interesting” from “mundane”.

In Section 1, we give (many) definitions in preparation for stating the main theorems; comments on the hypotheses considered there are given in Section 7, as are the acknowledgements. These theorems are given, without proof, in Section 2, and their proofs are delayed until Section 6. The style of these first two sections of this paper is quite formal; for a more relaxed overview, see Section A. Section 3 quickly presents our use of “string diagrams”, a computational tool we will use to give structure to some of our more arbitrary calculations; further details and comparison with others’ use of this tool is given in Section B. Afterwards, Sections 4 and 5 contain the core arguments upon which the eventual high-level proofs rely.

1. Definitions and notation

In this section we define the natural transformations we consider and will eventually prove are unique. Since they apply to a variety of similar but slightly different common situations in algebraic geometry, we have abstracted the essential properties into a formal object of category theory; the particular applications are indicated in Section 2.

ABSTRACT PUSH AND PULL FUNCTORS. Our results apply in a variety of similar situations with slightly different features and technical hypotheses, all unified by the formalism of pushforward and pullback. The following definition encapsulates the axiomatic properties we will require.

1.1. DEFINITION. *We define a geofibered category to be a functor $F: \mathbf{Sh} \rightarrow \mathbf{Sp}$ between two categories called, respectively, “shapes” and “spaces”, such that:*

- *The category \mathbf{Sp} has all finite fibered products.*
- *For every morphism $f: X \rightarrow Y$ of \mathbf{Sp} , there are functors*

$$f^*: \mathbf{Sh}_Y \longleftarrow \mathbf{Sh}_X : f_*$$

between the fibers over Y and X , which are an adjoint pair (f^*, f_*) .

- The assignment $X \mapsto \mathbf{Sh}_X$ and $f \mapsto f^*$ constitutes a pseudofunctor $\mathbf{Sp} \rightarrow \mathbf{Cat}$; i.e. a cleaved fibered category, as described below.
- Dually, the assignment $X \mapsto \mathbf{Sh}_X$ and $f \mapsto f_*$ constitutes a pseudofunctor $\mathbf{Sp} \rightarrow \mathbf{Cat}^{\text{op}}$.
- The above pseudofunctor data obey the compatibilities described in the remainder of this subsection.

The terminology was chosen in part because “bifibered category” has a different meaning, and also to indicate that this concept models the fibered category of sheaves in various geometries. For a description of pseudofunctors in the context of fibered categories, see Vistoli’s exposition [Vis05, §3.1.2].

The requirement that \mathbf{Sp} has fibered products is only strictly speaking necessary for some of the following definitions, but since that of the “roof” is among them, given the central nature of this concept in our main theorems, a geofibered category would be quite useless otherwise.

The rest of this subsection is devoted to introducing notation and terminology and, along side it, describing the compatibilities of the data of a geofibered category.

1.2. DEFINITION. *As per Definition 1.1, in any geofibered category, for any morphism $f: X \rightarrow Y$ of spaces we have adjoint functors (f^*, f_*) of shapes which we call basic standard geometric functors (the terminology intentionally references “geometric morphisms” of topoi)*

We define a standard geometric functor (SGF) to be any composition of basic standard geometric functors. Formally, an SGF is equivalent to a diagram of spaces and morphisms forming a directed graph that is topologically linear, together with an ordering of its vertices from one end of the segment to the other (i.e. we “remember” the terms of the composition as well as the resulting functor). For an SGF F from shapes on X to shapes on Y , we write $X = \mathbf{S}(F)$ and $Y = \mathbf{T}(F)$.

The pseudofunctor data consists of a number of natural transformations, which are the basis for the main objects of our study.

1.3. DEFINITION. *Between pairs of SGFs there are canonical natural transformations of the following three types that we call the basic standard geometric natural transformations:*

$$\left. \begin{array}{l} \mathbf{unit}(f): \mathrm{id} \rightarrow f_* f^* \\ \mathbf{counit}(f): f^* f_* \rightarrow \mathrm{id} \end{array} \right\} \text{Adjunctions} \quad (1.1a)$$

$$\left. \begin{array}{l} \mathbf{comp}^*(f, g): f^* g^* \xrightarrow{\sim} (gf)^* \\ \mathbf{comp}_*(f, g): g_* f_* \xrightarrow{\sim} (gf)_* \end{array} \right\} \text{Compositions} \quad (1.1b)$$

$$\left. \begin{array}{l} \mathbf{triv}_*(f): \mathrm{id}_* \xrightarrow{\sim} \mathrm{id} \\ \mathbf{triv}^*(f): \mathrm{id}^* \xrightarrow{\sim} \mathrm{id} \end{array} \right\} \text{Trivializations.} \quad (1.1c)$$

For the latter two types, we will use $^{-1}$ to denote their inverses (which will not arise as often in our arguments).

The origins, nature, and compatibilities of these transformations are described in the following subpoints.

ADJUNCTIONS. The maps $\mathbf{unit}(f)$ and $\mathbf{counit}(f)$ are equivalent to the (f^*, f_*) adjunction in the usual way and satisfy the familiar compatibility required of an adjunction, which we state at risk of being pedantic so as to provide a complete reference for the data of a geofibered category.

$$\left(f^* \xrightarrow{f^* \mathbf{unit}(f)} f^* f_* f^* \xrightarrow{\mathbf{counit}(f) f^*} f^* \right) = \mathrm{id} \quad (1.2a)$$

$$\left(f_* \xrightarrow{\mathbf{unit}(f) f_*} f_* f^* f_* \xrightarrow{f_* \mathbf{counit}(f)} f_* \right) = \mathrm{id} \quad (1.2b)$$

In the context of *natural transformations of functors* rather than morphisms of individual objects, we have a “reverse natural adjunction” also defined by the unit and counit; we use it on occasion to aid definitions. The construction and proof of bijectivity are left as an exercise for readers who desire it.

$$\mathrm{RNA}_f: \mathrm{Hom}(F f_*, G) \longleftrightarrow \mathrm{Hom}(F, G f^*). \quad (1.3)$$

COMPOSITIONS. The data of a pseudofunctor implies the existence of isomorphisms $\mathbf{comp}_*(f, g): (fg)_* \cong f_* g_*$ for every composable pair of morphisms f, g of spaces. Likewise, we have isomorphisms $\mathbf{comp}^*(f, g): (fg)^* \cong g^* f^*$. The compatibilities required of these data by a pseudofunctor express their associativity in triple compositions:

$$\left((fgh)_* \cong (fg)_* h_* \cong f_* g_* h_* \right) = \left((fgh)_* \cong f_*(gh)_* \cong f_* g_* h_* \right) \quad (1.4)$$

and similarly for pullbacks. It must be noted that this data for pushforwards or pullbacks alone determines such data for the other; for example, given pseudofunctoriality of f_* , for the adjoint f^* , we define $\mathbf{comp}^*(f, g)$ via category-theoretic formalism: $(fg)_*$ has the left adjoint $(fg)^*$ and the composition $f_* g_*$ has as left adjoint the composition $g^* f^*$; since we have $(fg)_* \cong f_* g_*$ we also get $(fg)^* \cong g^* f^*$ by uniqueness of left adjoints. In more explicit terms, this means that for any shapes \mathcal{F} and \mathcal{G} , we have an isomorphism

$$\begin{aligned} \mathrm{Hom}((fg)^* \mathcal{F}, \mathcal{G}) &\cong \mathrm{Hom}(\mathcal{F}, (fg)_* \mathcal{G}) \cong \mathrm{Hom}(\mathcal{F}, f_* g_* \mathcal{G}) \cong \mathrm{Hom}(f^* \mathcal{F}, g_* \mathcal{G}) \\ &\cong \mathrm{Hom}(g^* f^* \mathcal{F}, \mathcal{G}). \end{aligned} \quad (1.5)$$

We require as a compatibility relation that this is $\text{Hom}(\text{comp}^*(f, g)_{\mathcal{F}, \mathcal{G}})$.

TRIVIALIZATIONS. The trivialization isomorphism $\text{triv}_*: \text{id}_* \cong \text{id}$ is another part of the pseudofunctor data (and likewise for pullbacks), as well as its compatibility with composition:

$$\left(\text{comp}_*(\text{id}, f): \text{id}_* f_* \cong (\text{id}f)_* = f_*\right) = \left(\text{triv}_* f_*: \text{id}_* f_* \cong \text{id}f_* = f_*\right) \quad (1.6)$$

(and, again, the same for pullbacks). Given this data just for pullbacks or pushforwards, for example the latter, we could define an isomorphism $\text{id}^* \rightarrow \text{id}$ and its inverse by adjunction, similar to (1.5):

$$\text{Hom}(\text{id}^* \mathcal{F}, \mathcal{G}) \cong \text{Hom}(\mathcal{F}, \text{id}_* \mathcal{G}) \cong \text{Hom}(\mathcal{F}, \mathcal{G}). \quad (1.7)$$

We require that this isomorphism be equal to $\text{Hom}(\text{triv}_{\mathcal{F}}^*, \mathcal{G})$.

CLASSES OF TRANSFORMATIONS. Typically we do not consider transformations in the pure form above, but rather in horizontal composition with some identity maps.

BASIC CLASSES. The following are the classes of natural transformations consisting of just one of the basic ones in horizontal composition.

1.4. DEFINITION. *In this definition, F and G represent any SGFs; f and g represent any maps of spaces. We define:*

$$\text{unit} = \{F \text{ unit}(f)G\} \quad \text{counit} = \{F \text{ counit}(f)G\} \quad (1.8a)$$

$$\text{comp}_0 = \{F \text{ comp}_*(f, g)G\} \cup \{F \text{ comp}^*(f, g)G\} \quad \text{comp} = \text{comp}_0 \cup \text{comp}_0^{-1} \quad (1.8b)$$

$$\text{triv}_0 = \{F \text{ triv}_* G\} \cup \{F \text{ triv}^* G\} \quad \text{triv} = \text{triv}_0 \cup \text{triv}_0^{-1}. \quad (1.8c)$$

And now we may introduce the main objects of our study.

1.5. DEFINITION. *A standard geometric natural transformation (SGNT) is an element of the class (in which we use the notation $\langle S \rangle$ to denote the class generated by S via composition)*

$$\text{SGNT} = \langle \text{unit} \cup \text{counit} \cup \text{comp} \cup \text{triv} \rangle. \quad (1.9)$$

For any SGNT $\phi: F \rightarrow G$ between two SGFs, we write $F = \text{dom } \phi$ and $G = \text{cod } \phi$.

Another way of understanding this class is the following characterization: SGNT is the smallest category of natural transformations of SGFs containing all identity maps, **comp** maps, **triv** maps, and their inverses, and that is closed under horizontal and vertical composition and adjunction of $*$ and $*$.

The following type of SGNT is of fundamental importance throughout the paper; it arises seemingly of its own accord in a variety of situations and is also essential in simplifying SGFs to the point that our main theorem is provable.

1.6. DEFINITION. Consider a commutative diagram

$$\begin{array}{ccc}
 P & \xrightarrow{\tilde{f}} & Y \\
 \tilde{g} \downarrow & & \downarrow g \\
 X & \xrightarrow{f} & B
 \end{array} \tag{1.10}$$

We define the associated commutation morphism

$$\text{cd}(f, g; \tilde{f}, \tilde{g}): g^* f_* \rightarrow \tilde{f}_* \tilde{g}^* \tag{1.11}$$

to be the map corresponding by g -adjunction and \tilde{g} -reverse natural adjunction to the corner-swapping map

$$f_* \tilde{g}_* \xrightarrow{\text{comp}_*(\tilde{g}, f)} (f \tilde{g})_* = (g \tilde{f}_*) \xrightarrow{\text{comp}_*(\tilde{f}, g)^{-1}} g_* \tilde{f}_* \tag{1.12}$$

See (4.6) for a visualization of this definition.

The commutation morphisms are too general to be entirely useful, though it is helpful to retain the concept for manipulations when no further hypotheses are needed. Nonetheless, for our theorems such hypotheses are needed.

1.7. DEFINITION. When (1.10) is cartesian, we write $\text{bc}(f, g) = \text{cd}(f, g; \tilde{f}, \tilde{g})$ and use the aggregate notation, as before:

$$\text{bc} = \{F \text{bc}(f, g)G\}. \tag{1.13}$$

INVERTIBLE BASE CHANGES. In this subsection and the next, we introduce some further conditions on geofibered categories that support a larger class of natural transformations to which to apply our theorems. We include the proofs of a few of their simplest properties, some of which are necessary for the definitions to make sense and others are simply most appropriate when placed here.

We begin with an abstract structure on geofibered categories inspired by Definition 1.7.

1.8. DEFINITION. Let $\mathbf{Sh} \rightarrow \mathbf{Sp}$ be a geofibered category. We will call a class P of morphisms in \mathbf{Sp} push-geolocalizing if:

- It contains every isomorphism and is closed under composition.
- For each $f \in P$ and $g \in \mathbf{Sp}$, the base change $\text{bc}(f, g)$ is an isomorphism and $\tilde{f} \in P$.

Similarly, we define P to be pull-geolocalizing by swapping the roles of f and g . We say that a geofibered category is given a geolocalizing structure if it is equipped with a pair of push- and pull-geolocalizing classes in \mathbf{Sp} .

We chose the word “localizing” in deference to the convention that to invert morphisms of a category is to localize it; the full term “geolocalizing” is in part to complement

“geofibered”, and in part to indicate that it is not the geolocalizing morphisms themselves that are inverted.

We note that, by definition, every geofibered category has a “trivial” geolocalizing structure consisting of just the isomorphisms of \mathbf{Sp} .

This concept matched by that of a “good” SGF or SGNT, based on the following concept. Its uniqueness claim will be proven later, to keep this section efficient.

1.9. DEFINITION. *We say that an SGF F is alternating if it is of the form $(f_*)g^*h_* \cdots$ (f optional), where none of f, g, h, \dots is the identity map. The alternating reduction of F is the unique (by Proposition 4.11) alternating SGF F' admitting an SGNT in $\langle \mathbf{comp}_0 \cup \mathbf{triv}_0 \rangle$ (thus, an isomorphism in SGNT) $F \rightarrow F'$.*

1.10. DEFINITION. *Let $\mathbf{Sh} \rightarrow \mathbf{Sp}$ be a geolocalizing geofibered category, and let F be an alternating SGF; we say that it is good if, either: every basic SGF in F of the form f^* is pull-geolocalizing; or, every basic SGF in F of the form f_* is push-geolocalizing. For any SGF, we say that it is good if it has a good alternating reduction. We say that an SGNT $\phi: F \rightarrow G$ is good if both F and G are good.*

1.11. LEMMA. *If F is good, then for any SGNT $\phi: F \rightarrow G$ with $\phi \in \mathbf{comp} \cup \mathbf{triv}$, we have G good as well. If F is alternating, the same is true for $\phi \in \mathbf{bc}$.*

PROOF. Let $\alpha: F \rightarrow F'$ and $\beta: G \rightarrow G'$ be the alternating reductions of F and G . If $\phi \in \mathbf{comp}_0 \cup \mathbf{triv}_0$, then $\beta\phi$ is also an alternating reduction of F , and thus by Definition 1.9 we have $F' = G'$; since F (and thus F') is good by hypothesis, G' (and thus G) is good. Likewise, if $\phi \in \mathbf{comp}_0^{-1} \cup \mathbf{triv}_0^{-1}$, then $\alpha\phi^{-1}$ is an alternating reduction of G and so we have $G' = F'$ and thus, again, that G is good.

Finally, for $\phi \in \mathbf{bc}$, it is clear first of all that for a map $\mathbf{bc}(f, g): g^*f_* \rightarrow \tilde{f}_*\tilde{g}^*$ itself, the right-hand side is good if the left is, because the hypotheses imposed by goodness on the maps of spaces are stable under base change. For a more general element $A\mathbf{bc}(f, g)B$ of \mathbf{bc} , if $F = Ag^*f_*B$ is good and alternating then $G = A\tilde{f}_*\tilde{g}^*B$ satisfies the conditions given in Definition 1.10 (but is not alternating); since the (push-, pull-) geolocalizing morphisms are closed under composition and contain the identity, these conditions are preserved by passage through any element of $\langle \mathbf{comp}_0 \cup \mathbf{triv}_0 \rangle$, so the same conditions are satisfied by the alternating reduction G' ; i.e. G' , hence G , is good. ■

This definition is complemented and completed by the following concept, upon which all significant results of this paper are ultimately based.

1.12. DEFINITION. *Let $\mathbf{Sh} \rightarrow \mathbf{Sp}$ be a geofibered category and let F be any SGF, viewed as a diagram of morphisms in \mathbf{Sp} . We define the roof of F , denoted $\mathbf{roof}(F)$, by constructing the final object in the category of spaces with maps to this diagram, which exists and is unique by the universal property of the fibered product.*

Then the roof is the space thus defined, also denoted $\mathbf{roof}(F)$, together with its “projection” morphisms

$$a_F: \mathbf{roof}(F) \rightarrow \mathbf{T}(F) \qquad b_F: \mathbf{roof}(F) \rightarrow \mathbf{S}(F). \qquad (1.14)$$

We also define $\text{roof}(F)$ as an SGF to be the functor $a_{F*}b_F^*$ (resp. a_{F*} if $b_F = \text{id}$, resp. b_F^* if $a_F = \text{id}$, resp. id), and in [Proposition 1.13](#) we will construct a canonical SGNT $F \rightarrow \text{roof}(F)$ that we will also call $\text{roof}(F)$. We will make an effort to eliminate ambiguity.

Just as for goodness, the concept of a roof comes associated with a natural morphism. We state this proposition here and prove it as [Propositions 4.17](#) and [4.18](#).

1.13. PROPOSITION. *For any SGF F , its roof is the unique SGF of the form a_*b^* admitting a map $\text{roof}(F): F \rightarrow \text{roof}(F)$ in SGNT_0^+ . This map exists and is uniquely determined by the properties that it factors through the alternating reduction of F and, if F is alternating, through an element of bc .*

1.14. COROLLARY. *Let $\mathbf{Sh} \rightarrow \mathbf{Sp}$ be a geolocalizing geofibered category. If F is any good SGF, then $\text{roof}(F)$ is also good, and $\text{roof}(F): F \rightarrow \text{roof}(F)$ is a natural isomorphism.*

PROOF. In the construction of [Proposition 1.13](#), the roof morphism is constructed in such a way that its factors in bc are applied only to an alternating source, with the other factors in comp_0 or triv_0 by [Definition 1.9](#). Therefore [Lemma 1.11](#) applies and each composand of $\text{roof}(F)$ is good. Then each base-change factor is, by definition of good, an isomorphism, while all the other factors are automatically so. ■

This motivates notation for classes of SGNTs including the inverses of invertible base changes.

1.15. DEFINITION. *We use the following class notation for the “balanced” elements of SGNT in which the units and counits only occur in pairs within a base change morphism, along with a “forward” variant:*

$$\text{SGNT}_0^+ = \langle \text{bc} \cup \text{comp}_0 \cup \text{triv}_0 \rangle \quad \text{SGNT}_0 = \langle \text{SGNT}_0^+ \cup (\text{SGNT}_0^+)^{-1} \rangle. \quad (1.15)$$

Here we use the inverse notation to refer to the class of inverses of only the actually invertible (as abstract natural transformations) elements of SGNT_0^+ .

INVERTIBLE UNIT MORPHISMS. We will also be allowing the inverses of certain units and counits, whose definition is more technical. The goodness hypothesis in this next definition is, strictly speaking, unnecessary for its formulation, but as it is required in our only nontrivial example of this concept, [Lemma 2.7](#), it seems likely that without it the definition would be invalid.

1.16. DEFINITION. *Let $\mathbf{Sh} \rightarrow \mathbf{Sp}$ be a geolocalizing geofibered category. We define an acyclicity structure on it to be a class C of pairs (a, b) of morphisms of \mathbf{Sp} having the properties:*

- *Every pair (a, i) or (i, b) , with a being push-geolocalizing and b being pull-geolocalizing, and where i is any invertible morphism in \mathbf{Sp} (and having the appropriate sources and targets, as below), is in C .*

- For any $(a, b) \in C$, we have $S(a_*) = T(b^*)$, and for every $f \in \mathbf{Sp}$ such that $a_* \mathbf{unit}(f)b^*$ is good and a natural isomorphism, either (left invertibility) $a_* \mathbf{unit}(f)$, or (right invertibility) $\mathbf{unit}(f)b^*$ is a natural isomorphism.
- C is closed under base change in the following sense: for any pair of maps $X \rightarrow T(a_*)$ and $Y \rightarrow S(b^*)$, the base change

$$(\tilde{a}, \tilde{b}) = X \times_{T(a_*)} (a, b) \times_{S(b^*)} Y \quad (1.16)$$

of the pair map (a, b) into $T(a_*) \times S(b^*)$ is in C .

This entails the derived concept of admissibility: an SGF F is admissible if $\mathbf{roof}(F) = (a_F, b_F)$ is in C . We will say, correspondingly, that any $(a, b) \in C$ is itself admissible.

We chose “acyclicity structure” in reference to a morphism $f: X \rightarrow Y$ being acyclic in geometry or topology when $\mathbf{unit}(f)$ is an isomorphism on certain sheaves (e.g. possibly only those of the form $b^*\mathcal{F}$), potentially after taking cohomology (i.e. applying derived a_*).

We note that, by definition, every geolocalizing geofibered category has a “trivial” acyclicity structure consisting of just the pairs (a, i) and (i, b) of the first point.

1.17. DEFINITION. Let $\mathbf{Sh} \rightarrow \mathbf{Sp}$ be a geolocalizing geofibered category with acyclicity structure. We define the class \mathbf{Unit} to be the class of all good SGNTs of the form $F\phi G$, where $\phi = A \mathbf{unit}(f)B$ is a natural isomorphism and $\mathbf{dom}(\phi)$ (which is AB) is admissible.

Until now we have ignored the counits, but for the most part, this is justifiable. We give the proof of the following lemma in [Section 6](#).

1.18. LEMMA. We have $\mathbf{counit} \subset \langle \mathbf{SGNT}_0^+ \cup \mathbf{unit} \rangle$ and, with \mathbf{Counit} as in [Definition 1.17](#) but with good $\mathbf{counit}(f)$ replacing arbitrary $\mathbf{unit}(f)$, we have $\mathbf{Counit} \subset \langle \mathbf{SGNT}_0^+ \cup \mathbf{Unit} \rangle$.

2. Main theorems

In this section we suppose the existence of an ambient geolocalizing geofibered category with an acyclicity structure, $\mathbf{Sh} \rightarrow \mathbf{Sp}$; as we have noted, any geofibered category can play this role with trivial structures; among our results is a description of some nontrivial ones. Our main results are of two types: the first contains a “quantitative” and comparatively technical result on SGNTs; the second contains a “qualitative” corollary. Proofs, if not indicated otherwise, are given in [Section 6](#).

The quantitative result is a classification of SGNTs:

2.1. THEOREM. Let $\phi: F \rightarrow G$ be in $\langle \mathbf{SGNT} \cup \mathbf{bc}^{-1} \cup \mathbf{Unit}^{-1} \rangle$, and denote $\mathbf{roof}(G) = a_{G*}b_G^*$. Then there exist maps of spaces f and g , such that $a_{G*} \mathbf{unit}(g)b_G^*$ is a natural

isomorphism, forming a commutative diagram:

$$\begin{array}{ccc}
 F & \xrightarrow{\text{roof}(F)} & \text{roof}(F) = a_{F*}b_F^* \\
 \downarrow \phi & & \downarrow (\text{comp}_*(f, a_F) \text{comp}^*(f, b_F)) \circ a_{F*} \text{unit}(f)b_F^* \\
 & & (a_F f)_*(b_F f)^* = (a_G g)_*(b_G g)^* \\
 & & \uparrow (\text{comp}_*(g, a_G) \text{comp}^*(g, b_G)) \circ a_{G*} \text{unit}(g)b_G^* \\
 G & \xrightarrow{\text{roof}(G)} & \text{roof}(G) = a_{G*}b_G^*
 \end{array} \tag{2.1}$$

The upward arrow can be omitted for $\phi \in \langle \text{SGNT} \cup \text{bc}^{-1} \rangle$.

The last sentence is [Proposition 5.2](#), and the rest is [Proposition 5.5](#). The use of $\langle \text{SGNT} \cup \text{bc}^{-1} \rangle$ rather than $\langle \text{SGNT}_0 \cup \text{unit} \rangle$ is justified by [Lemma 1.18](#).

Now we give criteria under which “all diagrams commute”; i.e. there exists only one SGNT between two given SGFs. First, we have the basic Corollaries [4.12](#) and [4.19](#):

2.2. THEOREM.

1. For any SGFs F and G , there exists at most one $\phi: F \rightarrow G$ in $\langle \text{comp} \cup \text{triv} \rangle$.
2. Let $G = a_*b^*$; then for any SGF F , there exists at most one $\phi: F \rightarrow G$ in SGNT_0 .

We have also found a curious conclusion that is not quite a corollary of the latter nor of the main theorem:

2.3. THEOREM. *Let $\phi: F \rightarrow G$ be in SGNT , where both F and G have at most one basic SGF. Then either both are trivial and $\phi \in \text{triv}_0^{-1} \text{triv}_0$, or neither is, so $F = G$ and $\phi = \text{id}$.*

More importantly, we have the following general theorem:

2.4. THEOREM. *Let $\phi: F \rightarrow G$ be a natural transformation of SGFs.*

Write $S(F) = S(G) = X$ and $T(F) = T(G) = Y$; denote $Z = X \times Y$ and let $b: \text{roof}(F) \times_Z \text{roof}(G) \rightarrow \text{roof}(G)$ be the projection map, where $\text{roof}(F) \rightarrow Z$ and $\text{roof}(G) \rightarrow Z$ are the pair maps (b_F, a_F) and (b_G, a_G) ; suppose that $a_{G} \text{unit}(b)b_G^*$ is an isomorphism.*

Suppose as well that the map $\text{roof}(G): G \rightarrow \text{roof}(G)$ is an isomorphism. If $\phi \in \langle \text{SGNT} \cup \text{bc}^{-1} \cup \text{Unit}^{-1} \rangle$, then it is the unique map $\psi: F \rightarrow G$ in that class.

We also have an auxiliary lemma giving sufficient conditions for the hypotheses of the theorem to hold. To state it, we use the term *weakly admissible* of an SGF F to mean that the pair morphism (a_F, b_F) of its roof is a universal monomorphism.

2.5. LEMMA. *We have $\text{roof}(G)$ is an isomorphism if G is good. The other condition of Theorem 2.4 holds if either:*

- *F is weakly admissible, and either there exists some $\phi \in \langle \text{SGNT}_0 \cup \text{unit} \rangle$, or its consequence: (b_G, a_G) factors through (b_F, a_F) ;*
- *G is weakly admissible and there exists some $\phi \in \langle \text{SGNT}_0 \cup \text{Unit}^{-1} \rangle$.*

Finally, we address the question of exhibiting geolocalizing and acyclicity structures on the geofibered categories that occur in practice. The intended application of this concept is to “sheaves” on “schemes” in various more or less geometric contexts:

Presheaves We may define $\mathbf{Sp} = \mathbf{Cat}^{\text{op}}$, the opposite category of all small categories, and for any category X , set \mathbf{Sh}_X to be the category of presheaves on X (with values in any fixed complete category; if it is abelian, then so is \mathbf{Sh}_X for any X). For any morphism $f: X \rightarrow Y$ of spaces (i.e. a functor $F: Y \rightarrow X$) and for any shapes (presheaves) $\mathcal{F} \in \mathbf{Sh}_X$ and $\mathcal{G} \in \mathbf{Sh}_Y$, we take the usual pushforward and pullbacks:

$$f_*(\mathcal{F})(y) = \mathcal{F}(F(y)) \qquad f^*(\mathcal{G})(x) = \lim_{x \rightarrow F(y)} \mathcal{G}(y) \qquad (2.2)$$

This is far more general than actual sheaves on schemes, but because of the flexibility in the concept of geolocalizing and acyclicity structures, our results apply uniformly to it (if with potentially less power should these structures be too trivial).

Quasicoherent sheaves Each \mathbf{Sh}_X is the category of quasicoherent sheaves in the Zariski topology on $X \in \mathbf{Sp}$ being a scheme of finite type over a fixed locally noetherian base scheme; pushforwards and pullbacks are those of quasicoherent sheaves.

Constructible étale sheaves Each \mathbf{Sh}_X is the category of ℓ -torsion or ℓ -adic “sheaves” in the étale topology on X being a scheme of finite type over a fixed locally noetherian base scheme; pushforwards and pullbacks are those of such “sheaves” (ultimately, inherited from actual sheaf operations) We will not recall the definition here, nor the definitions of any of the functors on it, but work purely with the associated formalism.

Constructible complex sheaves Each \mathbf{Sh}_X is the category of constructible sheaves of complex vector spaces in the classical topology on X being a complex-analytic variety of finite type over a fixed base variety. Pushforwards and pullbacks are those of sheaves of complex vector spaces.

Derived categories With \mathbf{Sp} being any of the above categories of spaces, we may take \mathbf{Sh}_X to be the derived category of the corresponding abelian category of shapes on a space X . Pushforwards and pullbacks are, respectively, the right-derived pushforward and left-derived pullback.

The noetherian hypotheses were suggested by Brian Conrad to ensure the good behavior of the sheaf theory, as we have attempted to encapsulate in [Definition 1.1](#) and subsequent definitions. Presumably this list, as varied as it is, is incomplete; for instance, most likely sheaves on the crystalline site, D-modules, and other such categories belong on it as well.

We regret that we have been unsuccessful in locating references that explicitly prove, in all of these contexts, that the functors described (which *are* defined very carefully) actually have the properties that we have called a geofibered category. In all cases, the pseudofunctor structure of pushforwards is either totally obvious (as for presheaves and sheaves, given [\(2.2\)](#)) or formal (as for ℓ -adic and derived sheaves), and in lieu of existing literature on the category-theoretic minutiae, we feel free to simply declare that the structure for pullbacks should be *determined* by adjunction and the required compatibilities; see the discussion following [Definition 1.1](#).

In order to exhibit geolocalizing structures on these categories, we simply recall the base change theorems of algebraic geometry, together with standard properties of the types of morphism.

2.6. LEMMA. (*Proper, smooth, and flat base change.*) *In the étale or complex (possibly derived) contexts, the class of proper morphisms of schemes is push-geolocalizing and the class of smooth morphisms is pull-geolocalizing. In the quasicohherent (possibly derived) context, the class of flat morphisms is pull-geolocalizing. ■*

A stunningly general flat base change theorem for derived quasicohherent sheaves on any algebraic spaces over any scheme can be found at [\[Stacks, Tag 08IR\]](#). The proper and smooth base change theorems in étale cohomology were proven for torsion sheaves by Artin [\[SGA IV₃, Exp. xii, xiii, xvi\]](#); the ℓ -adic and derived versions follow formally. The recent preprint [\[LZ12\]](#) of Liu and Zhang presents an extension of these theorems to the derived categories on Artin stacks in the lisse-étale topology, as well.

As for acyclicity structures, in general we can only offer the trivial one, but in the étale context or its derived analogue (and presumably by the same token, the complex one) we can do better using a theorem from SGA4.

2.7. LEMMA. *In the étale or derived étale contexts, the class*

$$C = \{(a, b): X \rightarrow Y \times Z \mid (a, b) \text{ is an immersion}\} \quad (2.3)$$

is an acyclicity structure for the geolocalizing structure defined in [Lemma 2.6](#).

3. String diagrams

In the course of executing the general strategy of [Section 5](#) we will need to do a few specific computations with SGNTs. As these have little intrinsic meaning, the work would be unintelligible using traditional notation, so we have chosen to express it visually using “string diagrams”.

For the convenience of readers familiar with such depictions of categorical algebra, in the present section we will give only the essential definitions and results that will be cited

in our later proofs. A more conversational introduction to this topic of string diagrams, together with the proofs of the mostly routine facts shown here, are left to the appendix.

In summary, in our diagrams, vertical edges represent basic geometric functors, and are marked with upward or downward arrows to distinguish, respectively, f_* from f^* . The shapes shown in [Figure 1](#) generate all our string diagrams by horizontal (natural transformation) and vertical (functor) composition, which correspond to horizontal (left-to-right) and vertical (bottom-to-top) concatenation of diagrams. We use a doubled-line convention for our edges, which has no mathematical meaning but does improve aesthetics and (with some imagination) topologically justifies most of our string diagram identities as being mere topological deformations in the plane.

An example of the correspondence between string diagrams and SGNTs is given in [Figure 2](#), but we will never be so thorough in labeling them in actual use.

We emphasize that it is important for the correspondence between string diagrams and natural transformations that the string diagram be *labeled*; i.e. for the edges and components to have the meanings assigned to them by [Figure 1](#). For if not, then the operation of reflecting the string diagram vertically produces another planar graph that is valid combinatorially, but does not necessarily correspond to any SGNT (the corresponding operation on functors $f^*g_*h^*\cdots$ is to swap upper $*$ and lower $*$, but the resulting basic SGFs are no longer composable). We thank Mitya Boyarchenko for this last observation, however much it forced the restructuring of this paper.

Although the diagram acquires its unique identity as an SGNT only after labeling all the edges, we will almost always omit these labels. We will never assert the identity of an SGNT corresponding to an unlabeled diagram without indicating how we would label it, which can (in our applications) always be deduced from the ends by propagating through the various transformations.

STRING DIAGRAM IDENTITIES. In this subsection we record all the identities satisfied by string diagrams with two shapes. These can basically be considered the relations in the category whose objects are SGFs and whose morphisms are the string diagrams between two given SGFs. The proofs, which are either direct translation of the corresponding symbolic equations or simple manipulation, are left to the appendix for the convenience of readers who would prefer to get to the point.

3.1. LEMMA. (*Adjunction identities.*)



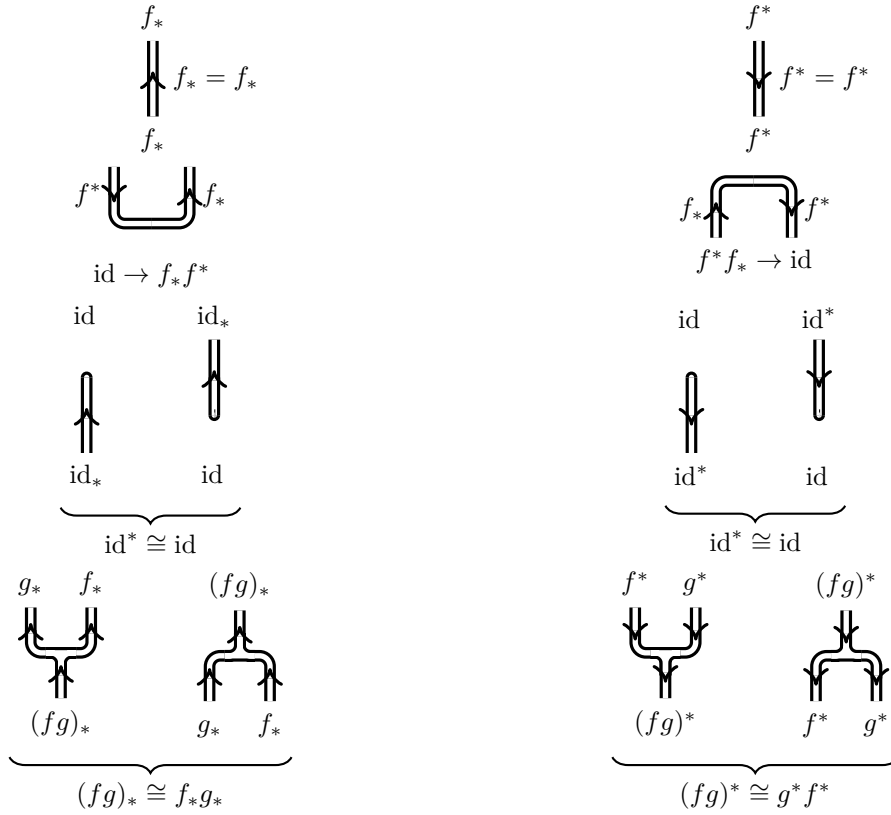
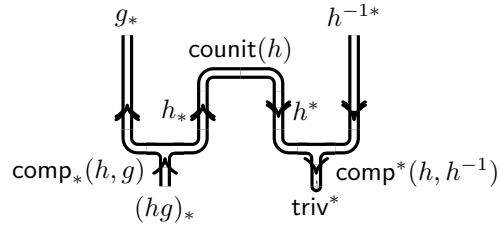


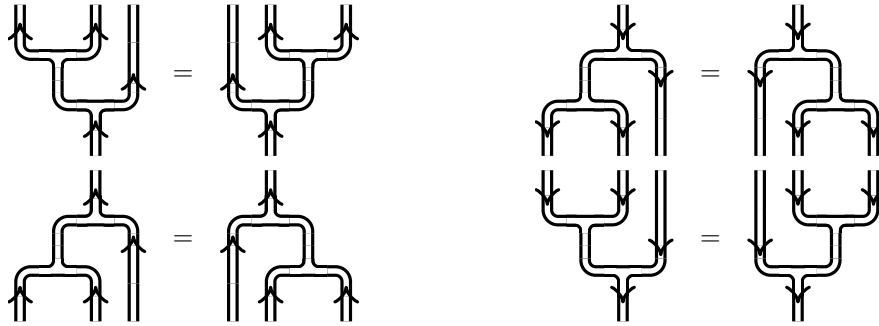
Figure 1: The basic SGNTs as string diagrams



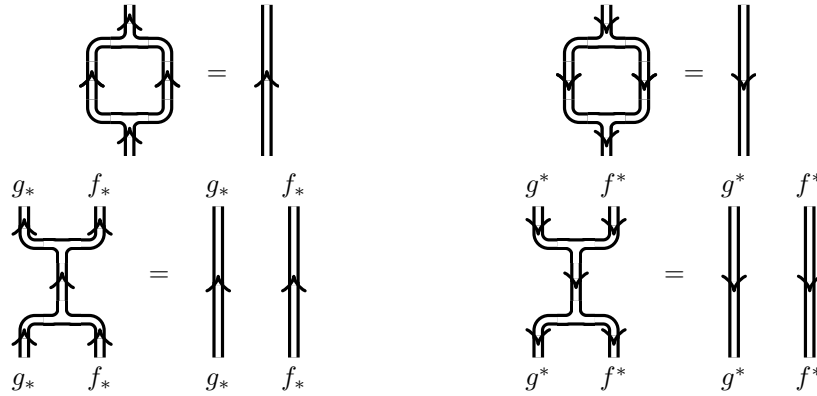
$$\begin{aligned}
 & h^{-1*} \text{counit}(h) g_* \circ h^{-1*} h^* \text{comp}_*(h, g) \circ \text{comp}^*(h^{-1}, h)(hg)_* \circ \text{triv}^*(hg)_* \\
 (hg)_* & \xrightarrow{\text{triv}^*} \text{id}^*(hg)_* = (hh^{-1})^*(hg)_* \xrightarrow{\text{comp}^*(h, h^{-1})} h^{-1*} h^*(hg)_* \\
 & \xrightarrow{\text{comp}_*(h, g)} h^{-1*} h^* h_* g_* \xrightarrow{\text{counit}(h)} h^{-1*} g_*
 \end{aligned}$$

Figure 2: Example of a string diagram with corresponding SGNT expressed several ways

3.2. LEMMA. (*Composition identities (associativity).*)



3.3. LEMMA. (*Composition identities (inverses).*)



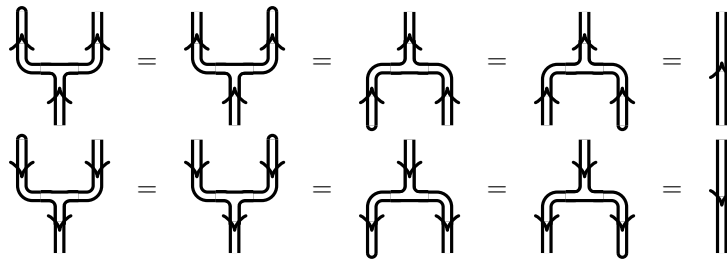
3.4. LEMMA. (*Trivialization identities (trivializations).*)



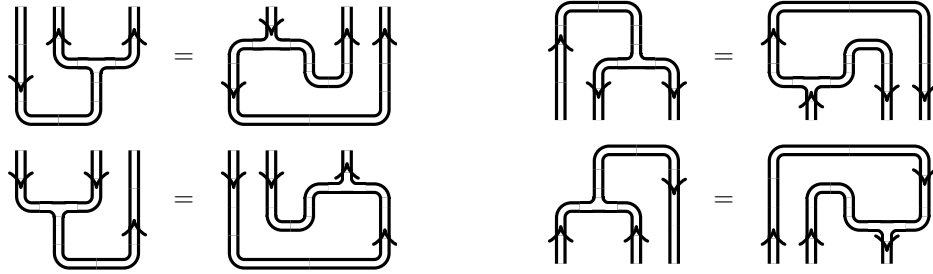
3.5. LEMMA. (*Trivialization identities (adjunctions).*)



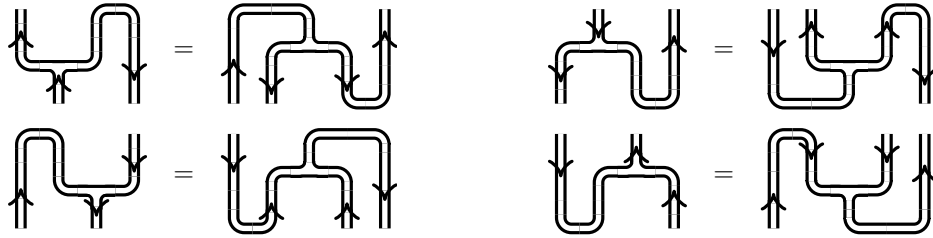
3.6. LEMMA. (*Trivialization identities (compositions).*)



3.7. LEMMA. (*Adjunction-composition identities (part 1).*)



3.8. LEMMA. (*Adjunction-composition identities (part 2).*)



3.9. LEMMA. *Suppose we have maps of spaces $f, g, h,$ and k and another one l such that $g = lk$ and $h = fl$ (resp. $f = hl$ and $k = lg$). Then we have the first equality (resp. the second equality) below, and similarly for the $*$ version:*

(3.1)

We note that [McC12, Definition 4] establishes a *Frobenius algebra* as an object in a monoidal category satisfying precisely the above diagrammatic constraints, except that here, the ends of that diagram are not all the same object. This lemma therefore shows that the basic SGFs of a geofibered category form a generalization of a Frobenius algebra, presumably a “Frobenius algebroid” in the same sense as a groupoid, though we have not been able to find this term in use.

3.10. LEMMA. *We have the equivalence (for either direction of edges and all possible assignments of maps of spaces compatible with the depicted compositions):*

(3.2)

4. Uniqueness of reduced forms

We have defined (Definitions 1.9 and 1.12) two reduced forms for an SGF and claimed that they are unique. In this section we prove these claims; the first one is relatively straightforward and does not require string diagrams, but the manipulations of the second one are simplified by their use, so we have placed both of them at this point. Only one general relation needs to be noted here, expressing the “commutation” of unrelated natural transformations.

4.1. LEMMA. *Let $F = AF_1BF_2C$ be a composition of functors; let $\phi: F_1 \rightarrow G_1$ and $\psi: F_2 \rightarrow H_2$ be natural transformations. Then the following diagram commutes:*

$$\begin{array}{ccc}
 F = AF_1BF_2C & \xrightarrow{AF_1B\psi C} & H = AF_1BH_2C \\
 A\phi BF_2C \downarrow & & \downarrow A\phi BH_2C \\
 G = AG_1BF_2C & \xrightarrow{AG_1B\psi C} & K = AG_1BH_2C
 \end{array} \quad \blacksquare \quad (4.1)$$

UNIQUENESS OF THE ALTERNATING REDUCTION. Showing that the alternating reduction is unique is a matter of enforcing an inductive structure on its construction, as encapsulated by the following lengthy definition.

4.2. DEFINITION. *Let $\phi: F \rightarrow G$ be an SGNT; we define a staging structure on it to be the following data:*

- A sequence of SGNTs

$$F = F_0 \xrightarrow{\gamma_1} F_1 \xrightarrow{\alpha_2} F_2 \cdots \xrightarrow{\alpha_{2n}} F_{2n} = G \quad (4.2)$$

together with representations of $\gamma_i \in \langle \mathbf{comp}_0 \rangle$ and $\alpha_i \in \langle \mathbf{triv}_0 \rangle$ as products of factors that are basic SGNTs.

- *This data must satisfy some conditions. We make reference to the terms of an SGF, its basic SGF composands; a term is trivial if it is of the form id_* or id^* ; a composable pair is a sequence of two consecutive terms of the form f_*g_* or f^*g^* . We define the stage of any trivial term or composable pair in any of the intermediate SGFs of the staging structure, together with conditions:*

- *The stage of any composable pair of F_0 is 0, and of any trivial term is 1.*
- *Let α be any factor of α_i , acting locally as $t_1tt_2 \rightarrow t_1t_2$, where t is trivial; we require that t have stage $i - 1$ and that both pairs t_1t and tt_2 (when composable) have stage i . If t_1t_2 is a composable pair, we define its stage to be i . We say that t , t_1t_2 , and t_1t and tt_2 are affected by α .*
- *Let γ be any factor of γ_i , acting locally as $t_0t_1t_2t_3 \rightarrow t_0tt_3$, where t_1t_2 is composable. We require that t_1t_2 have stage $i - 1$, and that if either of these terms is trivial, then that term has stage i . We define the stages of t_0t or tt_1 (when composable), to*

be those of t_0t_1 and t_1t_2 each; the stage of t (if it is trivial) is i , unless t_1 and t_2 are both trivial, in which case the stage of t is the larger of their stages. We say that t and t_1t_2 are affected by γ .

We say that the staging structure is complete at stage s when s is odd if F_s has no composable pairs, and when s is even if F_s has no trivial terms.

The following fact is trivial:

4.3. LEMMA. *In a staged SGNT $\phi: F \rightarrow G$, any trivial term or composable pair in F_s has stage at most s , if $s > 0$. A term of stage s in G is not affected by any factor of any α_i or γ_i with $i > s$. ■*

In the next few lemmas we establish the strong properties of a complete staging.

4.4. LEMMA. *If an SGNT with staging $\phi: F \rightarrow G$ is complete at any stage $s \geq 3$, then it is also complete at stage $s - 2$.*

PROOF. Suppose s is odd, and consider a composable pair t_1t_2 in F_{s-2} ; it has stage $\leq s - 2$ by Lemma 4.3. It is thus not possible for either term to be affected by a factor of α_{s-1} , so this pair persists into F_{s-1} . There, since t_1t_2 still has stage $\leq s - 2$, it cannot be acted on by a factor of γ_s so this space in F_{s-1} will remain in a composable pair in F_s .

Suppose s is even, and consider a trivial term t in F_{s-2} . By Lemma 4.3 it has stage $\leq s - 2$, so it is not affected by any factor of γ_{s-1} , so it persists into F_{s-1} . There, it has the same stage and so cannot be affected by any factor of α_s , so it persists into F_s , forming a trivial term there. ■

4.5. COROLLARY. *If $\phi: F \rightarrow G$ has a staging such that G has neither composable pairs nor trivial terms, then it is complete at every stage $s \geq 1$.*

PROOF. It is vacuously compatible with the definition of a staging to define $F_{2n+2} = F_{2n+1} = F_{2n} = G$ with $\alpha_{2n+2} = \gamma_{2n+1} = \text{id}$, with which convention ϕ is complete at stages $2n + 1$ and $2n$, so by induction on Lemma 4.4, at every stage $s \geq 1$. ■

4.6. LEMMA. *If $\phi: F \rightarrow G$ has a staging that is complete at every stage $s \geq 1$, then all the intermediate SGFs F_i and SGNTs α_i and β_i are uniquely determined by F .*

PROOF. It is clear that if ϕ is complete at F_1 , then F_1 must be obtained by composing all composable pairs of F_0 ; this is well-defined regardless of order by (1.4) and gives γ_1 uniquely. Then it is also clear that if ϕ is complete at F_2 , it must be obtained by deleting all trivial terms of F_1 , which is well-defined regardless of order by Lemma 4.1 and gives α_2 uniquely. By induction, the lemma follows. ■

Now we turn to the construction of a staging on a given SGNT.

4.7. LEMMA. *Let $\phi: F \rightarrow G$ have a staging; then any reordering of the factors of any γ_i by associativity, as in (1.4), is a valid staging.*

PROOF. Given two overlapping compositions:

$$t_0((t_1t_2)t_3)t_4 \qquad t_0(t_1(t_2t_3))t_4 \qquad (4.3)$$

suppose that the first grouping satisfies the conditions of a staging. Thus, t_1t_2 has stage $\geq i - 1$ and each trivial t_j has stage i . Let t be their composition, so in particular, tt_3 has the same stage as t_2t_3 and, as it is acted on by another factor of γ_i , thus has stage $\geq i - 1$. It follows that t_2t_3 has stage $\geq i - 1$, and likewise that if t_3 is trivial, then it has stage i . Therefore the second grouping also satisfies the conditions of a staging. Furthermore, the product of the first group has stage i unless all three of t_1 , t_2 , and t_3 are trivial, which is the same exception as for the second group, in which case its stage is the maximum of their stages. So the results of each grouping are identical. ■

4.8. LEMMA. *Let $\phi: F \rightarrow G$ have a staging and let $t_0t_1t_2t_3t_4t_5$ be a sequence of composable terms in G . Suppose that the staging can be extended by composing t_2t_3 or by composing both t_1t_2 and t_3t_4 . Then it can be extended by the sequence of compositions*

$$t_0(t_1((t_2t_3)t_4))t_5. \qquad (4.4)$$

PROOF. We suppose that we are constructing stage $2n + 1$ of ϕ . By the first assumption, the stage of t_2t_3 is $2n$. By construction, $t = (t_2t_3)$, if trivial, has stage $2n + 1$ and both pairs t_1t and tt_4 have the same stages as, respectively, t_1t_2 and t_3t_4 . By the second assumption, the latter stages are $2n$ and each $t_{1,4}$ (when trivial) has stage $2n + 1$. Therefore, the pair tt_4 may be composed in a staging, with value u (if trivial) of stage $2n + 1$ and the pair t_1u having stage that of t_1t , which is the same as that of t_1t_2 , which is $2n$. So the pair t_1u may be composed as well. ■

4.9. COROLLARY. *Let $\phi: F \rightarrow G$ have a staging ending at stage $2n + 1$, and let $\psi \in \mathbf{comp}_0$ have domain G affecting a pair t_1t_2 of stage $\leq 2n - 2$. Define a factorization of γ_{2n+1} , using Lemma 4.3 and associativity (1.4), as $\gamma^1\gamma^0$, where γ^0 does not affect t_1t_2 and γ^1 is a pair of two \mathbf{comp}_0 factors having values t_1 and t_2 respectively. Let ϕ' be the portion of ϕ up to stage $2n$ and let ψ' be ψ applied to the codomain of γ^0 . Then, if $\psi'\gamma^0\phi'$ has a staging, so does $\psi\phi$.*

PROOF. By Lemma 4.7, the SGNT $\gamma^1\gamma^0\phi'$ has a staging, so by definition, so does $\gamma^0\phi'$, and then Lemma 4.8 applies. ■

4.10. LEMMA. *Let $\phi: F \rightarrow G$ have a staging, and let ψ be a basic SGNT either in \mathbf{comp}_0 or \mathbf{triv}_0 composable with ϕ . Then $\psi\phi$ has a staging.*

PROOF. The proof is by descending induction, with the two terminal cases:

1. $\psi \in \text{triv}_0$ acts as $t_1 t_2 \rightarrow t_1 t_2$, where the triple satisfies the conditions of a staging at stage $2n$;
2. $\psi \in \text{comp}_0$ acts as $t_1 t_2 \rightarrow t$, where the composable pair satisfies the conditions of a staging at stage $2n + 1$;

In the first, we can append ψ to α_{2n} , satisfying the definition of a staging. In the second, we can begin γ_{2n+1} with ψ , satisfying the definition of a staging. Note that these each apply, respectively, at stages 2 and 1, so the induction is well-founded.

Suppose that $\psi \in \text{triv}_0$ acts on a trivial term t ; if (1) does not apply, then either t has stage $< 2n - 1$ or one of the pairs including t is composable; if the conditions for a composition in a staging do not apply to it, then either the pair has stage $< 2n$ or the other term is trivial of stage $< 2n - 1$.

Suppose that $\psi \in \text{comp}_0$ acts on a composable pair $t_1 t_2$; if (2) does not apply, then either its stage is $< 2n$, or one of the terms t_i is trivial of stage $< s$.

In either case, we have identified a pair $t_1 t_2$ such that either at least one term is trivial and has stage $< 2n - 1$, or $t_1 t_2$ itself has stage $< 2n$; when $\psi \in \text{triv}_0$ it acts on one of the t_i , and when $\psi \in \text{comp}_0$ it acts on $t_1 t_2$. Then by (1.6), we can replace ψ with either the composition of $t_1 t_2$ or a trivialization of either trivial t_i . In the latter case, by Lemma 4.1, we may “commute” ψ with all β_i and γ_i down to F_s . In the former case, we do this with $\psi \in \text{comp}_0$ using Corollary 4.9 and Lemma 4.1. Either way, the proof follows by induction. ■

4.11. PROPOSITION. *For any SGF F , there exists a unique map $F \rightarrow G$ in $\langle \text{comp}_0 \cup \text{triv}_0 \rangle$ with G alternating.*

PROOF. Any such SGNT has a staging by induction on Lemma 4.10, and G has neither trivial terms nor composable pairs by definition of an alternating SGF. Therefore, by Corollary 4.5, it is complete at every stage, so by Lemma 4.6 it is uniquely determined by F . Given F alone, such a map exists by the construction in the proof of that lemma. ■

4.12. COROLLARY. *If $\phi: F \rightarrow G$ is in $\langle \text{comp} \cup \text{triv} \rangle$, then it is unique there, and F and G have the same alternating reduction.*

PROOF. Write ϕ as an alternating composition of $\langle \text{comp}_0 \cup \text{triv}_0 \rangle$ and its inverses. By Proposition 4.11, either one preserves both the alternating reduction and the map to it, so this is true of ϕ as a whole by induction. If there is another such map ϕ' , then $\phi^{-1}\phi'$ is a self-map of F preserving the map to its alternating reduction. Since that map is an isomorphism, we have $\phi^{-1}\phi' = \text{id}$. ■

RELATIONS. To deal with the roof and its additional complications arising from the base change morphisms, we prove a number of “commutation relations” among \mathbf{comp}_0 , \mathbf{triv}_0 , and \mathbf{bc} , beginning with rewriting some especially trivial transformations in terms of simpler ones. In this subsection, we use \mathbf{cd} instead of \mathbf{bc} to emphasize the fundamentally diagrammatic nature of the arguments; following that, we will be forced for practical reasons to specialize.

4.13. LEMMA. *We have identities*

$$\mathbf{cd}(f, \text{id}; f, \text{id}) = \mathbf{triv}_*^{-1} f^* \circ f^* \mathbf{triv}_* : f^* \text{id}_* \rightarrow \text{id}_* f^* \quad (4.5a)$$

$$\mathbf{cd}(\text{id}, f; \text{id}, f) = f_* \mathbf{triv}^{*-1} \circ \mathbf{triv}^* f_* : \text{id}^* f_* \rightarrow f_* \text{id}^*. \quad (4.5b)$$

PROOF. First, we note that according to the definition (1.12), we have as a string diagram

$$\mathbf{cd}(f, g; \tilde{f}, \tilde{g}) = \begin{array}{c} \tilde{g}^* \quad \tilde{f}_* \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ f_* \quad g^* \end{array} \quad (4.6)$$

Now, for the claimed identities, moving the right-hand side terms to the left, they are equivalent to the two string diagram identities

$$\begin{array}{c} \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \end{array} \quad \begin{array}{c} \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \end{array} \quad (4.7)$$

the first of which follows from Lemma 3.5 and then Lemma 3.6 (twice each), and the second of which follows from Lemma 3.6 (twice) and then Lemma 3.1. ■

For those nontrivial SGNTs that do “interact”, we have the following relations:

4.14. LEMMA. *The following “commutation relations” hold in \mathbf{SGNT}_0 :*

a. *We have, for any composable maps f and g of spaces:*

$$\begin{aligned} (f_* \text{id}_* g_* \xrightarrow{\mathbf{triv}_*} f_* g_* \xrightarrow{\mathbf{comp}_*(g,f)} (fg)_*) \\ = (f_* \text{id}_* g_* \xrightarrow{\mathbf{comp}_*(g,\text{id})} f_* g_* \xrightarrow{\mathbf{comp}_*(g,f)} (fg)_*) \end{aligned} \quad (4.8a)$$

$$\begin{aligned} (f_* \text{id}^* g_* \xrightarrow{\mathbf{triv}^*} f_* g_* \xrightarrow{\mathbf{comp}_*(g,f)} (fg)_*) \\ = (f_* \text{id}^* g_* \xrightarrow{\mathbf{cd}(g,\text{id};g,\text{id})} f_* g_* \text{id}^* \xrightarrow{\mathbf{comp}_*(g,f)} (fg)_* \text{id}^* \xrightarrow{\mathbf{triv}^*} (fg)_*) \end{aligned} \quad (4.8b)$$

$$\begin{aligned} (f^* \text{id}^* g^* \xrightarrow{\mathbf{triv}^*} f^* g^* \xrightarrow{\mathbf{comp}^*(f,g)} (gf)^*) \\ = (f^* \text{id}^* g^* \xrightarrow{\mathbf{comp}^*(\text{id},g)} f^* g^* \xrightarrow{\mathbf{comp}^*(f,g)} (gf)^*) \end{aligned} \quad (4.8c)$$

recall our convention on omitting labels from diagrams; there should be no ambiguity provided that one recalls the labels of the ends.

We begin by applying [Lemma 3.6](#) to the one triv, simplifying it to the first diagram below, which then transforms using [Lemma 3.8](#) on the blue subdiagram:

$$(4.13)$$

Finally, we break the identity (middle lower) string according to [Lemma 3.4](#) and remove the associated unit-triv combination using [Lemma 3.5](#) and then [Lemma 3.6](#) again, leaving the first figure of [\(4.12\)](#), as desired. The same computation applies to [\(4.8d\)](#) (or, to avoid repeating the same work: take the above computation, reverse all the arrows, and reflect it horizontally).

For [\(4.11a\)](#), we again render the two transformations as diagrams, which are somewhat more complex:

$$(4.14)$$

To parse the first diagram, the blue sub-diagram is $\text{comp}_*(h, f)$ and the red one is $\text{cd}(fh, g; \tilde{f}\tilde{h}, k)$. To parse the second diagram, the blue sub-diagram is $\text{cd}(f, g; \tilde{f}, \tilde{g})$, the red one is $\text{cd}(h, \tilde{g}; \tilde{h}, k)$, and the black one is $\text{comp}_*(\tilde{h}, \tilde{f})$. Simplifying this requires a number of steps but as a first major goal we eliminate the loop. We use matching colors to indicate changes in the diagrams, where violet denotes a subdiagram that is both blue and red (i.e. is changed both from and to the adjacent diagrams).

$$(4.15)$$

In the first equality we use [Lemma 3.7](#), and in the second, we use [Lemma 3.10](#). Now we

paste this into the rest of (4.14):

$$\text{Diagram 1} = \text{Diagram 2} = \text{Diagram 3} \quad (4.16)$$

where the first equality is Lemma 3.2 again and the second is Lemma 3.7. The last diagram is the left diagram of (4.14), as desired. The proof of (4.11b) is the same (that is, with arrows reversed and the diagrams reflected horizontally). ■

UNIQUENESS OF THE ROOF. Now we pursue a “normal form” for the roof similar to the “staging” defined for the alternating reduction. It is much less complicated, however.

4.15. LEMMA. *We have $\langle \text{comp}_0 \cup \text{triv}_0 \cup \text{bc} \rangle = \langle \text{triv}_0 \rangle \langle \text{comp}_0 \rangle \langle \text{bc} \rangle$.*

PROOF. Before proceeding, note that using Lemma 4.14(b) from “left to right” requires making a choice of the individual morphisms \tilde{f} , \tilde{g} , and \tilde{h} given only the composition $\tilde{h}\tilde{f}$; this may not, in general, be possible, but is in fact canonical if we assume the outer rectangle is cartesian (then we may take \tilde{g} to be the base change of g). This is why we specialize to bc in this corollary, aside from the applications.

Now, it follows from Lemma 4.1 and the cases (a) and (b) of Lemma 4.14 that $\langle \text{comp}_0 \cup \text{triv}_0 \cup \text{cd} \rangle = \langle \text{triv}_0 \rangle \langle \text{comp}_0 \cup \text{cd} \rangle$, and it follows from case (c) that we have $\langle \text{comp}_0 \cup \text{cd} \rangle = \langle \text{comp}_0 \rangle \langle \text{cd} \rangle$. ■

Although this proves that any $\phi \in \langle \text{comp}_0 \cup \text{triv}_0 \cup \text{bc} \rangle$ can be written as $\alpha\beta\gamma$ with $\alpha \in \langle \text{triv}_0 \rangle$, $\beta \in \langle \text{comp}_0 \rangle$, and $\gamma \in \langle \text{bc} \rangle$, the intermediate functors $\text{dom } \alpha = \text{cod } \beta$ and $\text{dom } \beta = \text{cod } \gamma$ are not canonical. In the general case this is unfixable, though the next lemma remains valid. In the case of the roof, this is fortunately all that is required.

4.16. LEMMA. *Let F be any SGF and let $\phi: G \rightarrow F$ be in $\langle \text{bc} \cup \text{comp}_0 \cup \text{triv}_0 \rangle$. Then there exists an SGF $F_{\text{triv}} = F_l F F_r$ and an SGNT $\phi_{\text{triv}} = \phi_l F \phi_r: F_{\text{triv}} \rightarrow F$, having the properties that:*

- $F_l = \text{id}$ if and only if $\phi_l = \text{id}$; otherwise, $F_l = \text{id}_*$ and $\phi_l = \text{triv}_*$, and unless $F = f^* F_1$ starts with a $*$, we have $F_l = \text{id}$,
- $F_r = \text{id}$ if and only if $\phi_r = \text{id}$; otherwise, $F_r = \text{id}^*$ and $\phi_r = \text{triv}^*$, and unless $F = F_1 f_*$ ends with a $*$, we have $F_r = \text{id}$,

and a $\psi: G \rightarrow F_{\text{triv}}$ in $\langle \text{bc} \cup \text{comp}_0 \rangle$ such that $\phi = \phi_{\text{triv}} \psi$.

PROOF. By Lemma 4.15, it suffices to assume $\phi \in \langle \text{triv}_0 \rangle$. By (4.5a), any triv_* in the configuration $f^* \text{triv}_* : f^* \text{id}_* \rightarrow f^*$ can be replaced by $\text{bc}(f, \text{id})$ followed by $\text{triv}_* f^* : \text{id}_* f^* \rightarrow f^*$. Inductively, then, any configuration $f_1^* \cdots f_n^* \text{triv}_*$ is equal to $\text{triv}_* f_1^* \cdots f_n^*$ following a sequence of base change morphisms. Similarly, by (4.5b) we may convert $\text{triv}^* f_{1*} \cdots f_{n*}$ to $f_{1*} \cdots f_{n*} \text{triv}^*$ following a sequence of base changes.

Analogously, by (1.6), any triv_* in the configuration $\text{triv}_* f_*$ or $f_* \text{triv}_*$ can be replaced wholesale with simply $\text{comp}_*(f, \text{id})$ or $\text{comp}_*(\text{id}, f)$ respectively. Likewise for triv^* .

Now, by Lemma 4.1, all triv_* and triv^* morphisms “commute”, so we may assume that those covered in the first paragraph occur first on the composition of ϕ , followed by those covered in the second paragraph. It follows that the only trivializations that cannot be eliminated by this combination are those appearing in the configuration $\text{triv}_* f^*$ at the left end, or $f_* \text{triv}^*$ at the right end of the composition, giving F_{triv} and ϕ_{triv} , and the above construction furnishes ψ . \blacksquare

Now we can prove Proposition 1.13.

4.17. PROPOSITION. *For any SGF F , its roof $\text{roof}(F)$ is the unique SGF of the form $a_* b^*$ admitting a map $\text{roof}(F) : F \rightarrow a_* b^*$ in SGNT_0^+ , and this map is unique.*

PROOF. Let $\phi : F \rightarrow \text{roof}(F)$ be in SGNT_0^+ ; that is, in $\langle \text{bc} \cup \text{comp}_0 \cup \text{triv}_0 \rangle$. Lemma 4.15 then places it in the form $\alpha \gamma \beta$ with $\alpha \in \langle \text{triv}_0 \rangle$, $\gamma \in \langle \text{comp}_0 \rangle$, and $\beta \in \langle \text{bc} \rangle$, and furthermore by Lemma 4.16, $\alpha = \text{id}$, since $\text{roof}(F) = a_{F_*} b_F^*$.

Write $\beta : F \rightarrow G$; since $\gamma : G \rightarrow a_{G_*} b_G^*$ is in comp_0 , we must have G of the form $(a_{1*} \cdots a_{i*})(b_j^* \cdots b_1^*)$ and $\beta \in \langle \text{bc} \rangle$; thus, the maps a_k and b_k furnish the projections of the final object mapping to the diagram of F described in Definition 1.12, and in particular, G is unique. We will write $G = F_{\text{bc}}$.

The proof then reduces to two claims: that the map $\gamma : F_{\text{bc}} \rightarrow \text{roof}(F)$ is unique in $\langle \text{comp}_0 \rangle$, and that the map $\beta : F \rightarrow F_{\text{bc}}$ is unique in $\langle \text{bc} \rangle$. The former is simple: since $\text{roof}(F) = a_{F_*} b_F^*$, the map γ must be of the form $(\text{comp}_*(?, ?) \cdots)(\text{comp}^*(?, ?) \cdots)$; i.e. the $*$ and $*$ compositions do not interact. Then by (1.4), both factors are fully associative, so γ is unique.

For the latter, the proof follows directly from Lemma 4.1 but only with the right words. By definition, $\beta \in \langle \text{bc} \rangle$ is a composition of base change morphisms, which we may view as rewriting the string of basic SGFs F : each $\text{bc}(f, g)$ replaces $g^* f_*$ with $\tilde{f}_* \tilde{g}^*$; we will call the two-letter space it affects its *support*. We will say that given any particular representation of β as a composition, the basic SGFs of F itself have *level 1* and that each replacement increases the level of each basic SGF by 1. We will say that the *level* of a specific $\text{bc}(f, g)$ factor is n if that is the larger of the levels of f and g .

By definition of level, if a factor of level 1 follows any other factor, then their supports must be disjoint, and therefore, they are subject to Lemma 4.1. Therefore, β may be written with all the factors of level 1 coming first; i.e. $\beta = \beta \beta_0$, where β_0 is a composition of level-1 factors and β , a composition of higher-level factors.

The set of possible level-1 factors is the set of possible base change morphisms out of F , each of which correspond to a configuration $g^* f_*$ in F . Since in $\text{roof}(F)$, no such

configurations remain, each one must be the support of some factor of β , and therefore necessarily some level-1 factor. Furthermore, since their supports are disjoint their order is irrelevant by [Lemma 4.1](#). Therefore, β_0 is uniquely determined by F .

Now, letting $F' = \text{cod } \beta_0$, this functor is uniquely determined by F and we may apply induction to $\beta: F' \rightarrow \text{roof}(F') = \text{roof}(F)$ to conclude that β is equal to a uniquely defined ordered product of its factors of each level, and is therefore unique, as claimed. ■

The following proposition gives a less laborious construction of the roof morphism.

4.18. PROPOSITION. *The roof morphism of F factors through the alternating reduction of F and admits a factorization $(\phi_n \beta_n) \cdots (\phi_1 \beta_1)$, where each $\beta_i \in \mathbf{bc}$ and each ϕ_i is an alternating reduction.*

PROOF. To define the roof morphism of F , it suffices to define it for the alternating reduction F' , since then by [Proposition 4.17](#), they will have the same roof. We thus define $\beta_1 = \text{id}$. Let β_2 be any element of \mathbf{bc} defined on F' , and let ϕ_2 be the alternating reduction of its codomain. We claim that we may complete the proof by induction applied to the codomain of ϕ_2 . Indeed, we have decreased the number of terms of F' if that number was at least three, and if it has exactly two terms, then the construction is completed by a single additional base change. ■

We finish with a few other results extending the uniqueness of the roof to a larger class of SGNTs.

4.19. COROLLARY. *If $\phi: F \rightarrow G$ is in SGNT_0 , then $\text{roof}(F) = \text{roof}(G)$ as SGFs and $\text{roof}(G)\phi = \text{roof}(F)$ as SGNTs.*

PROOF. For the former, write ϕ as an alternating composition of SGNT_0^+ and its inverses. By [Proposition 4.17](#) both preserve the roof (as in the proof of [Lemma 1.11](#)).

For the latter, again write $\phi = \phi_0 \alpha$ with $\alpha \in \text{SGNT}_0^+ \cup (\text{SGNT}_0^+)^{-1}$ and $H = \text{dom}(\phi_0) = \text{cod}(\alpha)$, we have $\text{roof}(G)\phi_0 = \text{roof}(H)$ by induction. If $\alpha \in \text{SGNT}_0^+$ then we get $\text{roof}(G)\phi = \text{roof}(H)\alpha = \text{roof}(F)$ by [Proposition 4.17](#). If $\alpha \in (\text{SGNT}_0^+)^{-1}$, then we get $\text{roof}(F)\alpha^{-1} = \text{roof}(H) = \text{roof}(G)\phi_0$. ■

5. Simplification via the roof

Finally we can employ the device of the roof to simplify an arbitrary SGNT that may contain unit morphisms.

5.1. PROPOSITION. *Let $\phi: FG \rightarrow Ff_*f^*G$ be the SGNT $F \text{unit}(f)G$. Then there exists a map of spaces \tilde{f} and a commutative diagram:*

$$\begin{array}{ccc}
 FG & \xrightarrow{\text{roof}(FG)} & \text{roof}(FG) = a_*b^* \\
 \phi \downarrow & & \downarrow a_* \text{unit}(\tilde{f})b^* \\
 Ff_*f^*G & \longrightarrow & a_*\tilde{f}_*\tilde{f}^*b^*
 \end{array} \tag{5.1}$$

where the lower edge is in SGNT_0 and is independent of ϕ .

PROOF. For notation, write $\text{roof}(F) = a_{F*}b_F^*$, $\text{roof}(G) = a_{G*}b_G^*$, and $\text{roof}(FG) = a_*b^*$ as in the statement (we used $a_{FG*}b_{FG}^*$ in [Definition 1.12](#)). Consider the following diagram:

$$\begin{array}{ccccc}
 & & \text{roof}(FG) & & \\
 & & \swarrow p_G & & \searrow p_F \\
 & \text{roof}(G) & & A & \text{roof}(F) \\
 & \swarrow b_G & & \downarrow f & \searrow a_F \\
 Z & & & Y & X \\
 & \searrow a_G & & \swarrow b_F & \\
 & & & &
 \end{array} \tag{5.2}$$

where $X = T(F)$, $Y = S(F) = T(G)$, and $Z = S(G)$, the middle square is cartesian, and $a = a_F p_F$, $b = b_F p_G$. We observe the following formal identity:

$$\text{roof}(FG) \times_Y A \cong (\text{roof}(G) \times_Y A) \times_A (A \times_Y \text{roof}(F)) \tag{5.3}$$

since $\text{roof}(FG) \cong \text{roof}(G) \times_Y \text{roof}(F)$ by [\(5.2\)](#). This is represented by the following diagram (rotated from the above for compactness):

$$\begin{array}{ccccc}
 & & A \times_Y \text{roof}(F) & \xrightarrow{\tilde{f}_F} & \text{roof}(F) \\
 & \nearrow \pi_F & & \searrow \tilde{b}_F & \downarrow b_F \\
 \text{roof}(FG) \times_Y A & & & & A \xrightarrow{f} Y \\
 & \searrow \pi_G & & \nearrow \tilde{a}_G & \downarrow a_G \\
 & & \text{roof}(G) \times_Y A & \xrightarrow{\tilde{f}_G} & \text{roof}(G)
 \end{array} \tag{5.4}$$

We define \tilde{f} to be the projection $\text{roof}(FG) \times_Y A \rightarrow \text{roof}(FG)$; then, since $\text{roof}(FG) \cong \text{roof}(F) \times_Y \text{roof}(G)$, we have the following compositions, using the projections from diagram [\(5.2\)](#):

$$\tilde{f}_F \pi_F = p_F \tilde{f} \qquad \tilde{f}_G \pi_G = p_G \tilde{f}. \tag{5.5}$$

Using all this notation, we can define the multipart composition for the bottom arrow of [\(5.1\)](#):

$$F f_* f^* G \xrightarrow{\text{roof}(F), \text{roof}(G)} a_{F*} b_F^* f_* f^* a_{G*} b_G^* \tag{5.6a}$$

$$\left. \begin{array}{l}
 \xrightarrow{\text{bc}(f, b_F), \text{bc}(a_G, f)} a_{F*} \tilde{f}_{F*} \tilde{b}_F^* \tilde{a}_{G*} \tilde{f}_G^* b_G^* \\
 \xrightarrow{\text{bc}(\tilde{a}_G, \tilde{b}_F)} a_{F*} \tilde{f}_{F*} \pi_{F*} \pi_G^* \tilde{f}_G^* b_G^* \\
 \xrightarrow{\text{comp}_*(\pi_F, \tilde{f}_F), \text{comp}^*(\pi_G, \tilde{f}_G)} a_{F*} (\tilde{f}_F \pi_F)_* (\tilde{f}_G \pi_G)^* b_G^* \\
 = a_{F*} (p_F \tilde{f})_* (p_G \tilde{f})^* b_G^* \\
 \xrightarrow{\text{comp}_*(\tilde{f}, p_F)^{-1}, \text{comp}^*(\tilde{f}, p_G)^{-1}} a_{F*} p_{F*} \tilde{f}_* \tilde{f}^* p_G^* b_G^* \\
 \xrightarrow{\text{comp}_*(p_F, a_F), \text{comp}^*(p_G, b_G)} a_* \tilde{f}_* \tilde{f}^* b^*.
 \end{array} \right\} \tag{5.6b}$$

$$\tag{5.6c}$$

We have braced the middle lines for comparison with $\text{roof}(FG)$, which by [Corollary 4.19](#) may be written as:

$$FG \xrightarrow{\text{roof}(F), \text{roof}(G)} a_{F*} b_F^* a_{G*} b_G^* \quad (5.7a)$$

$$\xrightarrow{\text{bc}(a_G, b_F)} a_{F*} p_{F*} p_G^* b_G^* \quad (5.7b)$$

$$\xrightarrow{\text{comp}_*(p_F, a_F), \text{comp}^*(p_G, b_G)} a_* b^*. \quad (5.7c)$$

To show that (5.1) commutes with (5.6) as the lower edge, we have to show that (using the numbers as names) $(5.6) \circ F \text{unit}(f)G = a_* \text{unit}(\tilde{f})b^* \circ (5.7)$. According to [Lemma 4.1](#), we have both of:

$$(5.6a) \circ F \text{unit}(f)G = a_{F*} b_F^* \text{unit}(f) a_{G*} b_G^* \circ (5.7a) \quad (5.8)$$

$$(5.6c) \circ a_{F*} p_{F*} \text{unit}(\tilde{f}) p_G^* b_G^* = a_* \text{unit}(\tilde{f}) b^* \circ (5.7c) \quad (5.9)$$

so since $(5.6) = (5.6c)(5.6b)(5.6a)$ and $(5.7) = (5.7c)(5.7b)(5.7a)$ it suffices to show that:

$$(5.6b) \circ a_{F*} b_F^* \text{unit}(f) a_{G*} b_G^* = a_{F*} p_{F*} \text{unit}(\tilde{f}) p_G^* b_G^* \circ (5.7b) \quad (5.10)$$

We can omit the a_{F*} and b_G^* on the ends and move the comp_* and comp^* inverses in (5.6b) to the other side, rendering both sides as maps of SGFs

$$b_F^* a_{G*} \rightarrow (p_F \tilde{f})_* (p_G \tilde{f})^*. \quad (5.11)$$

We show these are equal using string diagrams. First, the two sides of (5.10) are:

$(\tilde{f}_G \pi_G)^*$
 $(\tilde{f}_F \pi_F)_*$

$(p_G \tilde{f}_G)^*$
 $(p_F \tilde{f}_F)_*$

a_{G*}

b_F^*

$a_{G*} b_F^*$

(5.12)

Note that $\tilde{f}_F \pi_F = p_F \tilde{f}$ and the same for G , by (5.5). In the left diagram, the blue portion is $\text{unit}(f)$; the red portion is $\text{bc}(f, b_F) \text{bc}(a_G, f)$; the brown portion is $\text{bc}(\tilde{a}_G, \tilde{b}_F)$; and the yellow portion is $\text{comp}_*(\pi_F, \tilde{f}_F) \text{comp}^*(\pi_G, \tilde{f}_G)$. In the right diagram, the blue is $\text{unit}(\tilde{f})$; the red is $\text{bc}(a_G, b_F)$; and the brown is $\text{comp}_*(\tilde{f}, p_F) \text{comp}^*(\tilde{f}, p_G)$.

Despite the complexity of these diagrams we claim that both are equivalent to that of $\text{cd}(a_G, b_F; p_G \tilde{f}, p_F \tilde{f})$. First, the second one, where we match blue and red in consecutive

pictures to track regions that are altered; violet means a shape that is both blue and red.

(5.13)

by [Lemma 3.10](#), and this is exactly the desired cd diagram. For the larger diagram we have to do only scarcely more:

(5.14)

We have used [Lemma 3.1](#) on the blue diagram (with the cyan diagram unchanged for comparison), and [Lemma 3.7](#) on the red diagram. This rather extended result is now amenable to [Lemma 3.10](#) applied twice:

(5.15)

where, finally, we have used [Lemma 3.3](#) on the second diagram. The third is once again a cd diagram, necessarily $\text{cd}(a_G, b_F; p_G \tilde{f}, p_F \tilde{f})$ because the ends are correct. This completes the proof. ■

5.2. PROPOSITION. *Let $\phi: F \rightarrow G$ be in $\langle \text{SGNT}_0 \cup \text{unit} \rangle$; then there exists a map of spaces g making the following diagram commute:*

$$\begin{array}{ccc}
 F & \xrightarrow{\text{roof}(F)} & \text{roof}(F) = a_* b^* \\
 \phi \downarrow & & \downarrow (\text{comp}_*(g,a) \text{comp}^*(g,b)) \circ a_* \text{unit}(g) b^* \\
 G & \xrightarrow{\text{roof}(G)} & \text{roof}(G) = (ag)_* (bg)^*
 \end{array}
 \tag{5.16}$$

PROOF. We apply induction on ϕ : thus, suppose that $\phi = \alpha\phi_0$, where diagram (5.16) exists for $\phi_0: F \rightarrow F'$ and $\alpha \in \text{SGNT}_0 \cup \text{unit}$. If $\alpha \in \text{SGNT}_0$, then we can augment the ϕ_0 diagram simply:

$$\begin{array}{ccc}
 F & \xrightarrow{\text{roof}(F)} & \text{roof}(F) = a_*b^* \\
 \downarrow \phi_0 & & \downarrow \text{comp}_*(g,a) \text{comp}^*(b,g) \circ a_* \text{unit}(g)b^* \\
 F' & \xrightarrow{\text{roof}(F')} & \text{roof}(F') = (ag)_*(bg)^* = \text{roof}(G) \\
 \downarrow \alpha & \nearrow \text{roof}(G) & \\
 G & &
 \end{array} \tag{5.17}$$

where $\text{roof}(F') = \text{roof}(G)$ by [Corollary 4.19](#); the triangle commutes by [Corollary 4.19](#). If, alternatively, $\alpha \in \text{unit}$, then we write $F' = AB$ and augment the ϕ_0 diagram with (5.1):

$$\begin{array}{ccc}
 F & \xrightarrow{\text{roof}(F)} & \text{roof}(F) = a_*b^* \\
 \downarrow \phi_0 & & \downarrow (\text{comp}_*(g,a) \text{comp}^*(g,b)) \circ a_* \text{unit}(g)b^* \\
 F' = AB & \xrightarrow{\text{roof}(G)} & \text{roof}(AB) = (ag)_*(bg)^* \\
 \downarrow \alpha & & \downarrow (ag)_* \text{unit}(f)(bg)^* \\
 G = Af_*f^*B & \longrightarrow & (ag)_*f_*f^*(bg)^* \xrightarrow{\text{comp}_*(f,ag) \text{comp}^*(f,bg)} (agf)_*(bgf)^* = \text{roof}(G)
 \end{array} \tag{5.18}$$

The lower edge is, by [Corollary 4.19](#), equal to $\text{roof}(G)$; since both squares commute and the triangle commutes by construction, the large diagram commutes. We claim that the right edge is equal to $(\text{comp}_*(gf, a) \text{comp}^*(gf, b)) \circ a_* \text{unit}(gf)b^*$. We prove this using string diagrams:

$$\begin{array}{ccc}
 \begin{array}{c} (bgf)^* \quad (agf)^* \\ \downarrow \downarrow \\ (bg)^* \quad (ag)^* \\ \downarrow \downarrow \\ b^* \quad a_* \end{array} & = & \begin{array}{c} \downarrow \downarrow \\ \text{fg} \\ \downarrow \downarrow \end{array} \\
 & & = \begin{array}{c} \downarrow \downarrow \\ \text{fg} \\ \downarrow \downarrow \end{array}
 \end{array} \tag{5.19}$$

where we have used first [Lemma 3.2](#) and then (B.2) from the proof of [Lemma 3.7](#). ■

This is the ultimate theorem for unit morphisms alone; now we extend it to include inverse units.

5.3. LEMMA. *In [Proposition 5.1](#), if ϕ is in Unit, then so is the unit in the right edge.*

PROOF. For concurrency of notation, replace F and G in (5.1) with FA and BG respectively, where we assume as in Definitions 1.16 and 1.17 that $A \mathbf{unit}(f)B$ is a natural isomorphism with AB admissible and ϕ itself good. Let $\psi = a_* \mathbf{unit}(\tilde{f})b^*$ for brevity.

As usual, we write $\mathbf{roof}(F) = a_{F*}b_F^*$ and $\mathbf{roof}(G) = a_{G*}b_G^*$, and similarly $\mathbf{roof}(AB) = a_{AB*}b_{AB}^*$; by hypothesis, we have (a_{AB}, b_{AB}) admissible. Finally, we write $\mathbf{roof}(b_F^*a_{AB*}b_{AB}^*a_{G*}) = a_{0*}b_0^*$, so by Corollary 4.19 (together with Proposition 1.13) we have the alternating reduction $\mathbf{roof}(a_{F*}a_{0*}b_0^*b_G^*) = \mathbf{roof}(FABG) = a_*b^*$. We claim that $a_* \mathbf{unit}(\tilde{f})b^*$ is in **Unit**.

First, we verify that $a_* \mathbf{unit}(\tilde{f})b^*$ is good. Indeed, we have already shown that $a_*b^* = \mathbf{roof}(FABG)$, which is good by hypothesis on ϕ and Corollary 1.14; likewise, the alternating reduction of $a_*\tilde{f}_*\tilde{f}^*b^*$ is equal to $\mathbf{roof}(FAf_*f^*BG)$ by the same token and the lower edge of Proposition 5.1, so is also good.

Next, we verify that $a_{0*} \mathbf{unit}(\tilde{f})b_0^*$ is an isomorphism. Indeed, if we write the diagram corresponding to Proposition 5.1 with $F \leftrightarrow b_F^*A$ and $G \leftrightarrow Ba_{G*}$:

$$\begin{array}{ccc} (b_F^*A)(Ba_{G*}) & \longrightarrow & a_{0*}b_0^* \\ \downarrow & & \downarrow \\ (b_F^*A)f_*f^*(Ba_{G*}) & \longrightarrow & a_{0*}\tilde{f}_*\tilde{f}^*b_0^* \end{array} \quad (5.20)$$

then it appears as the right edge. Both horizontal edges are good, as the alternating reduction of $b_F^*ABa_{G*}$ is that of $FABG$ with a_{F*} and b_G^* removed, so by Corollary 1.14 they are isomorphisms. Since the left edge contains $A \mathbf{unit}(f)B$, which is an isomorphism by hypothesis, the right edge is an isomorphism, as claimed.

Finally, we verify that the pair map (a_0, b_0) is admissible. Indeed, if we write the diagram

$$\begin{array}{ccccc} X & & Y & & Z \\ & \searrow^{b_{AB}} & & \searrow^{a_{AB}} & \\ a_G \swarrow & & U & & V & \swarrow_{b_F} \\ & & & & \end{array} \quad (5.21)$$

then the map from its roof to the product of its two projections X and Z is

$$(b_0, a_0) = (X \times_U Y \times_V Z \rightarrow X \times Z) = X \times_U (Y \rightarrow U \times V) \times_V Z \quad (5.22)$$

and is therefore the base change of an admissible map, so admissible. \blacksquare

5.4. LEMMA. *Let $F = a_*b^*$ and let $\phi = a_* \mathbf{unit}(f)b^*$ be a left or right isomorphism. Then for any $\psi = a_* \mathbf{unit}(g)b^*$, it is possible to write $\psi\phi^{-1} = \alpha^{-1}\beta$ for some α, β with α invertible.*

PROOF. The proof is drawn from the ‘‘calculus of fractions’’; $\psi\phi^{-1}$ is represented by the upper-left corner of the following two diagrams, and we take α and β to be the other two

and therefore we can apply diagram (5.16) to both halves, giving a larger diagram:

$$\begin{array}{ccccc}
 a_{F*}b_F^* & \xrightarrow{\text{comp}^2 \circ \text{unit}} & * & \xleftarrow{\text{comp}^2 \circ \text{unit}} & a_{G*}b_G^* \\
 \parallel & & \uparrow & & \parallel \\
 \text{roof}(F) & \xrightarrow{\psi} & \text{roof}(H) & \xrightarrow{\beta_0} & * & \xleftarrow{\alpha_0^{-1}} & \text{roof}(G) \\
 \text{roof}(F) \uparrow & & \uparrow \text{roof}(H) & & \uparrow & & \uparrow \text{roof}(G) \\
 F & \xrightarrow{\phi_1} & H & \xrightarrow{\phi_0} & G \\
 & & \searrow & \nearrow & \\
 & & & \phi &
 \end{array} \tag{5.26}$$

which is what we want. Now, suppose that $\phi_1 \in \text{Unit}^{-1}$; then we rewrite the top line of (5.25) as

$$\begin{aligned}
 a_{F*}b_F^* & \xleftarrow[\phi_1^{-1}]{\text{comp}_*(f, a_F) \text{comp}^*(f, b_F) \circ a_{F*} \text{unit}(f)b_F^*} (a_F f)_*(b_F f)^* = a_{H*}b_H^* = \\
 & = (a_K g)_*(b_K g)^* \xrightarrow[\beta_0]{\text{comp}_*(g, a_K) \text{comp}^*(g, b_K) \circ a_{K*} \text{unit}(g)b_K^*} a_{K*}b_K^* \xrightarrow{\alpha_0} \text{roof}(G); \tag{5.27}
 \end{aligned}$$

Leaving the **comps** on the outside, the two units form the combination considered in Lemma 5.4, where by Lemma 5.3, we have $a_{F*} \text{unit}(f)b_F^* \in \text{Unit}$ and so, by definition of acyclicity structure, is either a left- or right-isomorphism. Therefore we can replace them with two different elements of **unit** (with a common target different from H). Since the **comps** are invertible, that means that we can rewrite (5.25) as

$$\begin{array}{ccccc}
 & & * & & \\
 & \swarrow & & \searrow & \\
 \text{roof}(F) & \xleftarrow[\phi_1^{-1}]{} & \text{roof}(H) & \xrightarrow{\beta_0} & \text{roof}(K) & \xleftarrow[\alpha_0^{-1}]{} & \text{roof}(G) \\
 \text{roof}(F) \uparrow & & \uparrow & & \uparrow & & \uparrow \text{roof}(G) \\
 F & \xrightarrow{\phi} & & & & & G
 \end{array} \tag{5.28}$$

The second diagonal arrow is, as indicated, invertible by Lemma 5.4. Then, as before, we may apply Proposition 5.2 to the left and to the composition of the two right arrows on the top of this diagram to complete the proof. ■

6. Proofs of the main theorems

Here are the proofs of the remaining main results and supporting lemmas.

PROOF OF LEMMA 1.18. Let $\text{counit}(f): f^* f_* \rightarrow \text{id}$ be a counit morphism, and consider the following diagram:

$$\begin{array}{ccccc}
 X & \xrightarrow{\Delta} & X \times_Y X & \xrightarrow{\tilde{f}_1} & X \\
 & \searrow & \downarrow \tilde{f}_2 & & \downarrow f \\
 & & X & \xrightarrow{f} & Y
 \end{array} \tag{6.1}$$

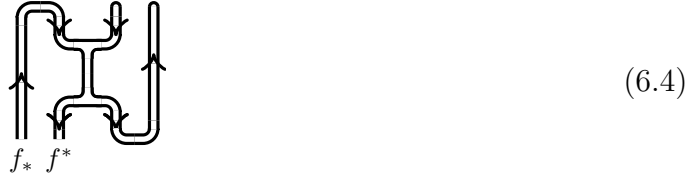
Then, in short, we have the following sequence of maps whose composition is an SGNT $f^* f_* \rightarrow \text{id}$ in $\langle \text{SGNT}_0^+ \cup \text{unit} \rangle$.

$$\begin{aligned}
 f^* f_* &\xrightarrow{\text{bc}(f, f)} \tilde{f}_{1*} \tilde{f}_2^* \xrightarrow{(\tilde{f}_1)_* \text{unit}(\Delta) \tilde{f}_2^*} \tilde{f}_{1*} \Delta_* \Delta^* \tilde{f}_2^* \\
 &\xrightarrow{\text{comp}_*(\Delta, \tilde{f}_1) \text{comp}^*(\Delta, \tilde{f}_2)} (\tilde{f}_1 \Delta)_* (\tilde{f}_2 \Delta)^* \\
 &= \text{id}_* \text{id}^* \xrightarrow{\text{triv}_* \text{triv}^*} \text{id} \text{id} = \text{id}. \quad (6.2)
 \end{aligned}$$

To see that this coincides with $\text{counit}(f)$, we do a string diagram computation. Below is the diagram of the map constructed in (6.2):



where the red portion is $\text{bc}(f, f)$, the blue portion is $\text{unit}(\Delta)$, the brown portion is $\text{comp}_*(\Delta, \tilde{f}_1) \text{comp}^*(\Delta, \tilde{f}_2)$, and the black portion is $\text{triv}_* \text{triv}^*$. This is precisely the second diagram considered in (5.12), with a_G and b_F replaced by f and the upper ends replaced by id^* and id_* and two triv s applied. Accounting for the change in notation, diagram (6.3) is equivalent to $\text{triv}_* \text{triv}^* \circ \text{cd}(f, f; \text{id}, \text{id})$:



By Lemma 3.5 and Lemma 3.6, this becomes merely $\text{counit}(f)$, as desired.

Suppose now that $\psi = F\phi G$ is good, where $\phi = A \text{counit}(f) B$ an isomorphism, $\text{counit}(f)$ is good, and $\text{dom}(\phi) = Af^* f_* B$ admissible as in Definition 1.17. Since $f^* f_*$ is good, the factor $\text{bc}(f, f)$ is an isomorphism by Definition 1.8. Thus, the composition (6.2) contains only one potentially non-isomorphism, namely the term $A\tilde{f}_{1*} \text{unit}(\Delta) \tilde{f}_2^* B$, which it follows is an isomorphism as well. It is good, even after composing with FA and BG : for its domain $FA\tilde{f}_{1*} \tilde{f}_2^* BG$, this follows from Corollary 1.14 and Proposition 4.17, since $\text{bc}(f, f) \in \text{SGNT}_0^+$; for its codomain, the entire trailing part of the diagram is its partial alternating reduction to $FABG$, which is assumed to be good. Finally, by uniqueness of the roof from Proposition 1.13:

$$\text{roof}(A\tilde{f}_{1*} \tilde{f}_2^* B) = \text{roof}(Af^* f_* B) = \text{roof}(\text{dom } \phi), \quad (6.5)$$

so $A\tilde{f}_{1*} \tilde{f}_2^* B$ is admissible. Thus, $A\tilde{f}_{1*} \text{unit}(\Delta) \tilde{f}_2^* B \in \text{Unit}$, so $\psi \in \langle \text{SGNT}_0^+ \cup \text{Unit} \rangle$, as claimed. ■

PROOF OF [THEOREM 2.3](#). This follows from examining [Proposition 5.2](#). Clearly both the upper and lower edges are either the identity or a single trivialization each, while the right edge must be the identity since any $\text{unit}(f)$ would incur both a $*$ and a $*$ in G , not both of which are present.

PROOF OF [THEOREM 2.4](#). By [Lemma 1.18](#) we may use $\langle \text{SGNT}_0 \cup \text{unit} \rangle$ in place of $\langle \text{SGNT} \cup \text{bc}^{-1} \rangle$. We show uniqueness by applying [Proposition 5.5](#); if the bottom edge is a natural isomorphism then it suffices to show that the right edge is independent of ϕ . We assume that both arrows in this edge occur; the case in which only one does is treated in [Lemma 2.5](#). We denote the right edge by $\alpha^{-1}\beta$.

Write $X = \text{S}(F) = \text{S}(G)$ and $Y = \text{T}(F) = \text{T}(G)$, and let $A = \text{roof}(F)$ and $B = \text{roof}(G)$, with projections $a_F: A \rightarrow X$, $b_F: A \rightarrow Y$, and similarly for G . Both maps f and g necessarily have the same source C ; we have $f: C \rightarrow A$ and $g: C \rightarrow B$. In order for $a_F f = a_G g$ and $b_F f = b_G g$, it is equivalent that the composites $(b_F, a_F)f = (b_G, a_G)g$ into $X \times Y$ be equal. Such a pair of maps $C \rightarrow A, B$ is equivalent once again to a single map $h: C \rightarrow A \times_{X \times Y} B$. Let a and b be the two projections of this fibered product.

We have $f = ah$ and $g = bh$, so by diagram [\(B.2\)](#) in the proof of [Lemma 3.7](#), we have

$$\text{unit}(f) = \left(\text{id} \xrightarrow{\text{unit}(a)} a_* a^* \xrightarrow{a_* \text{unit}(h) a^*} a_* h_* h^* a^* \xrightarrow{\text{comp}_*(h,a) \text{comp}^*(h,a)} f_* f^* \right) \quad (6.6)$$

and similarly for g . After composing with a_{F*} and b_F^* (resp. a_{G*} and b_G^*), applying $\text{comp}_*(f, a_F) \text{comp}^*(f, b_F)$ to the end is the same as the following, by [Lemma 4.1](#):

$$a_{F*} b_F^* \rightarrow a_{F*} a_* a^* b_F^* \rightarrow (a_F a)_* (b_F a)^* \rightarrow (a_F a)_* h_* h^* (b_F a)^* \rightarrow (a_F a h)_* (b_F a h)^*, \quad (6.7)$$

and similarly for g , with G replacing F and b replacing a . In the latter situation, assuming that $a_{G*} \text{unit}(b) b_G^*$ is an isomorphism, so is $(a_G b)_* \text{unit}(h) (b_G b)^*$ as the only potentially non-isomorphism in [\(6.7\)](#), and since $a_F a = a_G b$ and $b_F a = b_G b$, the last two steps of both are identical and so cancel out in $\alpha^{-1}\beta$. Thus, we may assume $f = a$ and $g = b$, which are uniquely determined by F and G , making ϕ canonical. ■

PROOF OF [LEMMA 2.5](#). It follows from [Corollary 1.14](#) that $\text{roof}(G)$ is an isomorphism if G is good.

For the second condition, first note that if $\phi \in \langle \text{SGNT}_0 \cup \text{unit} \rangle$, then by [\(5.16\)](#), (b_G, a_G) factors through (b_F, a_F) . Assuming that factorization, we have a graph morphism $\text{roof}(G) \rightarrow \text{roof}(F) \times_Z \text{roof}(G)$ forming a section of b . Since (b_F, a_F) is a universal monomorphism, so b is a monomorphism, and therefore that section is an isomorphism; thus, $a_{G*} \text{unit}(b) b_G^*$ (in fact, $\text{unit}(b)$ itself) is an isomorphism.

Observe that we can, in this case, fill in an α^{-1} to go with $\beta = \phi$, in the notation of the proof of [Theorem 2.4](#). Namely, we take $\alpha = \text{id} = \text{comp}_*(\text{id}, a_F) \text{comp}^*(\text{id}, b_F) \circ \text{unit}(\text{id})$, so the formal setup of the previous proof applies.

If we have a $\phi \in \langle \text{SGNT}_0 \cup \text{Unit}^{-1} \rangle$, then by diagram [\(5.16\)](#) taken upside down, we find that (b_F, a_F) factors through (b_G, a_G) by some map g ; when G is weakly admissible, the same argument applies and shows that $\text{roof}(F) \times_Z \text{roof}(G) \cong \text{roof}(F)$, with the projection

onto $\text{roof}(G)$ being g . Thus the right edge of the diagram is $a_{G*} \text{unit}(b)b_G^*$, and is also invertible. As before, we can assume that $\phi = \alpha^{-1}$ is complemented by a trivial β for notational purposes. ■

PROOF OF LEMMA 2.7. The SGA4 result that we require is the following criterion for an invertible unit morphism; we assume the same hypotheses on schemes as in the description of the étale context.

6.1. LEMMA. ([SGA IV₃, Exp. xv, Th. 1.15]) *Let $f: X \rightarrow Y$ be separated and of finite type, as well as locally acyclic (for example, smooth). Let \mathcal{F} be an ℓ -torsion (or, therefore, ℓ -adic) sheaf on Y . Then the unit morphism of sheaves*

$$\text{unit}(f)_{\mathcal{F}}: \mathcal{F} \rightarrow f_* f^* \mathcal{F} \quad (6.8)$$

is an isomorphism if and only if, for every algebraic geometric point $g: y \rightarrow Y$ with fiber $f_y: X_y \rightarrow y$, the unit morphism

$$\text{unit}(f_y)_{g^* \mathcal{F}}: g^* \mathcal{F} \rightarrow f_{y*} f_y^* g^* \mathcal{F} \quad (6.9)$$

is an isomorphism. ■

Now we proceed to the proof. There are three statements to verify, of which two are trivial:

- If i is an isomorphism, then for any morphisms a or b , the pair map (a, i) or (i, b) is isomorphic to the graph of a or b , which is an immersion.
- The base change of any immersion is again an immersion.

For the third statement, we must verify that if $\phi = a_* \text{unit}(f)b^*$ is good and a natural isomorphism with (a, b) an immersion, then ϕ is a left or right isomorphism. The goodness hypothesis entails that either both af and a are proper, or bf and b are smooth. Let us write $a: Y \rightarrow A$ and $b: Y \rightarrow B$, so $f: X \rightarrow Y$ as in the statement of Lemma 6.1.

Consider the proper case, and let p be any (geometric) point of A . We will show that the stalk of $a_* \text{unit}(f)$ at p is an isomorphism, and therefore that $a_* \text{unit}(f)$ is itself an isomorphism since p is arbitrary. Let \tilde{p} denote the fiber of a over p and let \tilde{p}_f denote that of af over p . Using Proposition 5.1 on $p^* a_*$ and $p^*(af)_*$, the stalk of $a_* \text{unit}(f)b^*$ at p is the unit map from $p^* a_* b^* \xrightarrow{\sim} a|_{p^*} (b\tilde{p})^*$ to

$$\begin{aligned} p^* a_* f_* f^* b^* &\cong p^*(af)_* f^* b^* \xrightarrow{\sim} (a|_p f|_{\tilde{p}})_* \tilde{p}_f^* f^* b^* \\ &\cong a|_{p^*} f|_{\tilde{p}^*} (f\tilde{p}_f)^* b^* \cong a|_{p^*} f|_{\tilde{p}^*} (\tilde{p}_f|_{\tilde{p}})^* b^* \\ &\cong a|_{p^*} f|_{\tilde{p}^*} f|_{\tilde{p}}^* \tilde{p}^* b^* \cong a|_{p^*} f|_{\tilde{p}^*} f|_{\tilde{p}}^* (b\tilde{p})^*, \end{aligned} \quad (6.10)$$

where by Lemma 2.6 the arrows are isomorphisms since a and af are proper. Since (a, b) is an immersion, $b\tilde{p}$ is also an immersion (the base change of (a, b) along p) and therefore $(b\tilde{p})^*(b\tilde{p})_* \xrightarrow{\sim} \text{id}$. Applying $(b\tilde{p})_*$ to the right above, we find that $a|_{p^*} \text{unit}(f|_{\tilde{p}})$ is an

isomorphism. The same computation shows that this is the stalk of $a_* \mathbf{unit}(f)$ at p , as desired.

Consider the smooth case. Then for any point q of B we again have isomorphisms

$$a_* b^* q_* \xrightarrow{\sim} (a\tilde{q})_* b|_q^* \qquad a_* f_* f^* b^* q_* \xrightarrow{\sim} (a\tilde{q})_* f|_{\tilde{q}^* f^*} b|_q^* \qquad (6.11)$$

in which $a\tilde{q}$ is an immersion and thus $(a\tilde{q})^*(a\tilde{q})_* \xrightarrow{\sim} \text{id}$. Applying that pullback, we find that $\mathbf{unit}(f|_{\tilde{q}})b|_q^*$ is an isomorphism. Since f is smooth, it is locally acyclic, so by [Lemma 6.1](#) all its fibers $\mathbf{unit}(f|_p)p^*b|_q^* = \mathbf{unit}(f|_p)b|_p^*$ are isomorphisms, over all points p of X . Applying it again, this means that $\mathbf{unit}(f)b^*$ is an isomorphism.

7. Comments and acknowledgements

Owing to the high level of abstraction in our presentation and the precise formulation of our definitions and theorems, some analysis of the limitations of this line of investigation is in order.

COMMENTS AND COUNTEREXAMPLES. The conditions of [Theorem 2.4](#) may require some explanation. Invertibility of the roof morphism is of course technically necessary in the proof, and the “good” property of [Lemma 2.5](#) gives convenient access to it, but some such condition is actually necessary, as the following example due to Paul Balmer shows:

7.1. EXAMPLE. *Let X be the scheme $(\mathbb{A}^1 \setminus \{0\}) \sqcup \{0\}$; that is, the affine line with the origin detached, and let $f: X \rightarrow \mathbb{A}^1$ be the natural map that is the identity on each connected component of X . There are two SGNTs from $f^* f_* f^*$ to itself: the identity map, and the composition $\phi = f^* \mathbf{unit}(f) \circ \mathbf{counit}(f) f^*$. They are not equal, as can be seen by computing them on the constant sheaf \mathcal{C} of rank 1 on \mathbb{A}^1 (this works for any kind of sheaf):*

- *$f^* \mathcal{C}$ is again the constant sheaf; $f_* f^* \mathcal{C}$ has rank 2 on every neighborhood of $\{0\}$; therefore $f^* f_* f^* \mathcal{C}$ has rank 2 on $\{0\}$.*
- *The map $\mathbf{counit}(f) f^*$ already has to map something of rank 2 to something of rank 1, so is not injective; therefore ϕ cannot be an isomorphism, much less the identity.*

Since f is neither proper nor even flat, of course [Lemma 2.5](#) does not apply; this example illustrates the necessity of gaining control of the pathologies of the maps along which the functors are taken. In fact, the map $\text{roof}(F)$ is not an isomorphism either: we have $\text{roof}(F) = \tilde{f}_{2}(\tilde{f}_1^*)^*$, where \tilde{f}_i are the projections of $X \times_{\mathbb{A}^1} X$ onto X , and one can see that, applied to \mathcal{C} on \mathbb{A}^1 , it yields a sheaf on X with rank 4 at $\{0\}$ and rank 2 elsewhere, which is nowhere isomorphic to FC as computed above.*

Our second comment concerns the specific and careful definition of the class \mathbf{Unit} . The ultimate goal was to be able to prove [Lemma 5.4](#), which requires only the property of being a “left or right isomorphism” (see [Definition 1.16](#)) but whose partner [Lemma 5.3](#) was easily proven only for SGNTs of the simple form allowed by the “trivial” acyclicity

structure (this is actually a simplified version of the very involved history of this research). We felt that more general invertible “units” were likely to occur in reality, and eventually arrived at the statement of [Lemma 2.7](#), which is the key ingredient in the expanded class, by pondering the following example:

7.2. *EXAMPLE. Let $X = \mathbb{A}^1$ and let $f: \{0\} \rightarrow X$ and $g: \{1\} \rightarrow X$ be the closed immersions of two points, and consider (5.23). Denote by p the map from X to a point, and let $a = b = p$. Then although both squares commute and their common left vertical arrow is an isomorphism, their right vertical arrows are both zero.*

Of course, in this example, the map $(p, p): X \rightarrow \text{pt} \times \text{pt}$ is far from an immersion. But it illustrates how the failure of this condition can cause problems, morally speaking by subtracting information from the unit morphism embedded between a_* and b^* to the point that it becomes an isomorphism when it should not; note that in Example 7.2, the map $p^* \rightarrow f_* f^* p^*$ alone is very much not an isomorphism. This map corresponds to the actual immersion $(\text{id}, p): X \rightarrow X \times \text{pt}$, in which the key [Lemma 5.4](#) actually does hold.

As for the other condition of [Theorem 2.4](#), we have no particular insight into its general meaning, but we do note that even for maps $a_* b^* \rightarrow c_* d^*$, the theorem can fail if we have $(c, d) = (a, b)g$ for multiple maps g , giving not necessarily equal SGNTs $a_* \text{unit}(g)b^*$. This is forbidden by the weak admissibility hypothesis of [Lemma 2.5](#).

Finally, we comment on our choice of terminology for “standard geometric functors”. It is easy to imagine trying to prove theorems similar to the above involving not only f_* and f^* but also $f_!$ and $f^!$ (the “exceptional” pushforward and pullback), and indeed, this was the original intention of this paper. Unfortunately, we were unable to identify the correct context for such results; it seems likely that they will need to include, as well, the bifunctors Hom and \otimes , filling out the full complement of the six functors, in order to adequately express the relationship between $f^!$ and f^* . Furthermore, the techniques of this paper appear inadequate, as a functor such as $f_* g_!$ is not alternating but apparently has no alternating reduction (hence no roof), and is seemingly incomparable with $f_! g_*$.

ACKNOWLEDGEMENTS. This paper would probably not have been finished were it not for Mitya Boyarchenko’s encouragement and his astute, if disruptive, reading of the first draft and its several implicit errors. I am also grateful to Brian Conrad for commenting on that draft and for advocating the next section, whose title was another of Mitya’s suggestions. During the sophomoreic stages of this research, Paul Balmer was very generous in giving me much seminar time for it, as well as disproving the original theorem and, therefore, motivating my formulation of everything that is now in the paper. Finally, I want to thank the anonymous referee for requesting additional organizational clarity and precision of language that led to my formulation of [Definition 1.1](#).

A. User guide

This section is an informal description of the intuition and use of the main theorem [Theorem 2.4](#) aimed at readers hoping to find a connection with familiar appearances

of the so-called geometric functors. We begin with a non-rigorous reformulation of our definitions:

DEFINITION. (*Imprecise.*) Let F and G be two functors of the form $\cdots f_* g^* \cdots$ of sheaves; we consider natural transformations $\phi: F \rightarrow G$ contained in the smallest class closed under the inclusion of those of the following three types:

Functorial

- The identity transformations;
- Functoriality isomorphisms $f_* g_* \cong (fg)_*$ and $g^* f^* \cong (fg)^*$;
- Functoriality isomorphisms $\text{id}_* \cong \text{id}$ and $\text{id}^* \cong \text{id}$;
- Compositions of transformations;
- Applications of f_* or f^* to either side of a transformation;

Adjunction

- Transformations corresponding under adjunction of $*$ and $_*$, or equivalently, the units $\text{id} \rightarrow f_* f^*$ and counits $f^* f_* \rightarrow \text{id}$ of such adjunctions;

Inverse

- The inverses of all invertible base change transformations $g^* f_* \xrightarrow{\sim} \tilde{f}_* \tilde{g}^*$, as in diagram (1.10).
- The inverses of all invertible transformations $f_* \rightarrow f_* g_* g^*$ or $f^* \rightarrow g_* g^* f^*$ derived from adjunction units.

To illustrate the functors in question, one such is by definition equivalent to the data of a zigzag diagram of spaces

$$\begin{array}{ccc}
 & \cdots & \\
 & \swarrow f_n & \searrow g_m \\
 & \cdots & \\
 X & \swarrow f_1 & \searrow g_1 & Y
 \end{array} \tag{A.1}$$

corresponding to the functor $(g_{1*} \cdots g_{m*}) \cdots (f_n^* \cdots f_1^*)$, where the central ellipsis (“ \cdots ”) corresponds to further zigzags in the peak ellipsis of the picture. This notation is intended to encompass the many variants such as $f^* g_*$ or $f_n^* \cdots f_1^*$ by omitting some of the terms from either side of the composition (respectively, maps from the diagram).

Each of these functors has an associated “roof” (Definition 1.12), depicted diagrammatically as follows:

$$F = \begin{array}{ccccc}
 & X & & Y & \\
 & \swarrow h & \searrow f & \swarrow g & \searrow \\
 A & & B & & \\
 & & & & \cdots & \\
 & & & & Z & \\
 & & & & \swarrow & \searrow k \\
 & & & & & C
 \end{array} \tag{A.2a}$$

$$\text{roof}(F) = \begin{array}{c} X \times_B Y \times \cdots \times Z \\ \swarrow \text{pr}_X \quad \searrow \text{pr}_Z \\ X \qquad \qquad Z \\ \swarrow h \quad \searrow k \\ A \qquad \qquad C \end{array} \begin{array}{l} \xrightarrow{b_F} \\ \xrightarrow{a_F} \end{array} = a_{F*} b_F^* \quad (\text{A.2b})$$

Given the above context, our main theorems mostly claim the following:

THEOREM. *A transformation $\phi: F \rightarrow G$ as above is unique if either:*

1. *$F = G$ is of the form f_* or g^* and ϕ draws from the functorial and adjunction transformations and inverse base changes; or, if only G is of the form $f_* g^*$ and ϕ draws only from the functorial transformations, base changes, and inverse base changes.*
2. *ϕ is arbitrary, if G is isomorphic to its roof $a_{G*} b_G^*$, if the pair of roof maps (b_G, a_G) of (A.2b) factors through the pair (b_F, a_F) , and if the latter is an immersion into $A \times C$ (in the notation of the diagram).*

The potential isomorphism mentioned in (2) is canonically defined in Definition 1.12; it is effectively the sequence of base changes corresponding to (A.2b). We have strengthened the hypotheses unnecessarily to make it more straightforward. See the example of “cohomological pullback” for a discussion.

Functors and transformations of this nature are found throughout geometry. Part (1) is the more common application, and signifies that a diagram can be shown to commute by simple manipulation. Part (2) indicates at least a slightly domain-specific computation that (as its proof ultimately shows) is not *entirely* symbolic manipulation.

COHERENCE OF TENSOR PRODUCTS. Recall that the *tensor product* of sheaves \mathcal{F} and \mathcal{G} on a scheme X satisfies the relations

$$\mathcal{F} \otimes \mathcal{G} = \Delta^*(\mathcal{F} \boxtimes \mathcal{G}), \quad \mathcal{F} \boxtimes \mathcal{G} = \text{pr}_1^* \mathcal{F} \otimes \text{pr}_2^* \mathcal{G},$$

where $\text{pr}_{1,2}$ are the projections $X \times X \rightarrow X$ and $\Delta: X \rightarrow X \times X$ is the diagonal morphism. Taking the latter, “outer” tensor product as the fundamental object allows the convenient formulation of tensor product identities entirely in terms of the functors described by the theorem. We obtain the commutativity and associativity constraints,

$$\mathcal{F} \otimes \mathcal{G} \cong \mathcal{G} \otimes \mathcal{F} \qquad (\mathcal{F} \otimes \mathcal{G}) \otimes \mathcal{H} \cong \mathcal{F} \otimes (\mathcal{G} \otimes \mathcal{H}),$$

using the functoriality of $*$ on the analogous identities:

$$\Delta = \text{sw } \Delta \qquad (\Delta \times \text{id})\Delta = (\text{id} \times \Delta)\Delta,$$

where $\text{sw}: X \times X \rightarrow X \times X$ is the coordinate swap.

As an easy consequence, we obtain the conclusion of Mac Lane’s coherence theorem for the tensor category of sheaves: any natural transformations of two parenthesized multiple

tensor products constructed only from commutativity and associativity constraints are equal. Indeed, all such parenthesized products are repeated pullbacks g^* for various maps g , so such a transformation is a map

$$g^* \cong g_1^* \cdots g_n^* \rightarrow h_1^* \cdots h_n^* \cong h^*,$$

where the outside isomorphisms are by functoriality of pullback. Therefore part (1) of the theorem applies.

PROJECTION FORMULA AND COMPATIBILITY DIAGRAMS. As a more interesting example, we consider the *projection formula* morphism for a map $f: X \rightarrow Y$ and sheaves \mathcal{F} and \mathcal{G} on X and Y respectively:

$$f_*\mathcal{F} \otimes \mathcal{G} \rightarrow f_*(\mathcal{F} \otimes f^*\mathcal{G}).$$

It can be expediently defined by first forming the cartesian diagram

$$\begin{array}{ccccc} & & \Gamma_f=(\text{id},f) & & \\ & & \curvearrowright & & \\ X & \xrightarrow{\Delta_X} & X \times X & \xrightarrow{(\text{id} \times f)} & X \times Y \\ \downarrow f & & & & \downarrow (f \times \text{id}) \\ Y & \xrightarrow{\Delta_Y} & Y \times Y & & \end{array}$$

and then rewriting the projection formula as

$$\Delta_Y^*(f \times \text{id})_*(\mathcal{F} \boxtimes \mathcal{G}) \rightarrow f_*\Gamma_f^*(\mathcal{F} \boxtimes \mathcal{G})$$

and realizing it as a base change morphism. Here we have used the fact that $f_*\mathcal{F} \boxtimes \mathcal{G} \cong (f \times \text{id})_*(\mathcal{F} \boxtimes \mathcal{G})$.

Examples of diagrams of the projection formula are diagrams (2.1.11) and (2.1.12) of [Con00], which are correctly said to be trivial but are nonetheless rather tedious, and in fact automatically commute. For verification we reproduce them here, using our notation. The first one is:

$$\begin{array}{ccc} f_*\mathcal{F} \otimes (\mathcal{G} \otimes \mathcal{H}) & \longrightarrow & (f_*\mathcal{F} \otimes \mathcal{G}) \otimes \mathcal{H} \\ \downarrow & & \downarrow \\ f_*(\mathcal{F} \otimes f^*(\mathcal{G} \otimes \mathcal{H})) & & f_*(\mathcal{F} \otimes f^*\mathcal{G}) \otimes \mathcal{H} \\ \downarrow & & \downarrow \\ f_*(\mathcal{F} \otimes (f^*\mathcal{G} \otimes f^*\mathcal{H})) & \longrightarrow & f_*((\mathcal{F} \otimes f^*\mathcal{G}) \otimes f^*\mathcal{H}) \end{array} \quad (2.1.11)$$

where the horizontal maps are associativity, the lower-left vertical map is the isomorphism $f^*(\mathcal{G} \otimes \mathcal{H}) \cong f^*\mathcal{G} \otimes f^*\mathcal{H}$ that is easily deduced from the outer product formula, and the others are projection formula maps. Here, we have $f: Y \rightarrow X$ for schemes X and Y , where \mathcal{F} is a sheaf on X and \mathcal{G} and \mathcal{H} are on Y . If we write Δ_X and Δ_Y

for the respective diagonal morphisms, then the two directions around the diagram are transformations

$$\Delta_X^*(f \times \text{id})_*(\text{id} \times \Delta_Y)^* \rightarrow f_*\Delta_Y^*(\text{id} \times f)^*(\Delta_Y \times \text{id})^*((\text{id} \times f) \times \text{id})^*,$$

and the latter is of the form f_*g^* after applying functoriality isomorphisms to the chain of pullbacks, so part (1) of the theorem applies.

The second diagram is:

$$\begin{array}{ccc} (gf)_*\mathcal{F} \otimes \mathcal{G} & \longrightarrow & (gf)_*(\mathcal{F} \otimes (gf)^*\mathcal{G}) \\ \downarrow & & \downarrow \\ g_*(f_*\mathcal{F}) \otimes \mathcal{G} & \longrightarrow & g_*f_*(\mathcal{F} \otimes (gf)^*\mathcal{G}) \\ \downarrow & & \downarrow \\ g_*(f_*\mathcal{F} \otimes g^*\mathcal{G}) & \longrightarrow & g_*f_*(\mathcal{F} \otimes f^*g^*\mathcal{G}) \end{array} \quad (2.1.12)$$

where the upper and lower horizontal and lower-left vertical maps are projection formulas, the other vertical maps are functoriality, and the middle map is defined by inverting either the upper-left or lower-right vertical maps and composing; to check the small squares it suffices to check the big one. Here, we have $f: Z \rightarrow Y$ and $g: Y \rightarrow X$, with sheaves \mathcal{F} on Z and \mathcal{G} on X , and the two ways around the diagram are transformations

$$\Delta_X^*(gf \times \text{id})_* \rightarrow g_*f_*\Delta_Z^*(\text{id} \times f)^*(\text{id} \times g)^*,$$

which again can be brought into the form treated by part (1) of the theorem by condensing the pullbacks and pushforwards.

COHOMOLOGICAL PULLBACK. An obvious way of involving units of adjunction is to introduce the *change of space* or *cohomological pullback* maps, defined as follows. Let $*$ denote the base (final) scheme and for any X with its canonical map $p: X \rightarrow *$, let $H^*(X, ?) = p_*$ denote global cohomology. Cohomological pullback is defined for any map $f: Y \rightarrow X$ as the natural transformation (written with reference to a sheaf \mathcal{F} on X)

$$H^*(X, \mathcal{F}) = p_*\mathcal{F} \rightarrow p_*f_*f^*\mathcal{F} \cong (pf)_*f^*\mathcal{F} = H^*(Y, f^*\mathcal{F})$$

since pf is the structure map for X . It follows easily from part (1) of the theorem that any composition of cohomological pullbacks from X to Z connected by some number of composable maps is independent of which intermediate maps are used; i.e. that pullback is functorial in the same sense as $*$ itself. Indeed, pullback is a transformation of the form $p_{X*} \rightarrow p_{Z*}f^*$, and by applying “reverse natural adjunction” (see (1.3) for this not-widely-mentioned operation) is equivalent to a transformation $p_{X*}f_* \rightarrow p_{Z*}$, which by a functoriality isomorphism is equivalent to a transformation $(p_X f)_* \rightarrow p_{Z*}$, which is unique.

More interestingly, let us say that f is a *cohomological isomorphism* if its associated pullback is an isomorphism. Then we can compose pullbacks and the *inverses* of pullbacks

that are isomorphisms to get maps between the global cohomology of spaces that are *not* connected by composable maps, but merely zigzags (similar to (A.2a)) where the zigs are invertible. Nonetheless, these are transformations to which part (2) of the theorem applies. For example, if we use the exact diagram shown there, we get transformations $p_{X*} \rightarrow p_{Z*}$.

The criterion given by the theorem as stated above for these transformations to be unique is somewhat restrictive: we would require that there be an actual map $f: Z \rightarrow X$ commuting with the structure maps p , in which case the zigzag pullback coincides with the actual pullback along f . A slightly less restrictive hypothesis is given in [Theorem 2.4](#): that the projection map $X \times Z \rightarrow Z$ be a cohomological isomorphism; then the universal representative of zigzag pullbacks from X to Z passes through $X \times Z$.

CUP PRODUCTS AND COHOMOLOGY ALGEBRAS. A final application of the theorem is again to tensor products: the *cup product* on cohomology, defined in the following two-step manner. For any sheaf \mathcal{F} on X , writing just $H^*(\mathcal{F})$ rather than $H^*(X, \mathcal{F})$, we have a map

$$H^*(\mathcal{F}) \otimes H^*(\mathcal{F}) \rightarrow H^*(\mathcal{F} \otimes \mathcal{F})$$

defined as the cohomological pullback along $\Delta: X \rightarrow X \times X$

$$(p \times p)_*(\mathcal{F} \boxtimes \mathcal{F}) \rightarrow (p \times p)_*\Delta_*\Delta^*(\mathcal{F} \boxtimes \mathcal{F}) \cong p_*(\mathcal{F} \otimes \mathcal{F})$$

We have used the fact that $(p \times p)_*(\mathcal{F} \boxtimes \mathcal{F}) \cong p_*\mathcal{F} \boxtimes p_*\mathcal{F} = p_*\mathcal{F} \otimes p_*\mathcal{F}$. Then, if \mathcal{F} is a ring sheaf with multiplication $m: \mathcal{F} \otimes \mathcal{F} \rightarrow \mathcal{F}$, we compose with the induced map $H^*(m): H^*(\mathcal{F} \otimes \mathcal{F}) \rightarrow H^*(\mathcal{F})$ to obtain the cup product

$$\smile: H^*(\mathcal{F}) \otimes H^*(\mathcal{F}) \rightarrow H^*(\mathcal{F}).$$

It is easy to show, from the functoriality of cohomological pullback, that whenever we have a map $f: X \rightarrow Y$ and thus we have a commutative diagram of spaces, we get one of natural transformations:

$$\begin{array}{ccc} X & \xrightarrow{\Delta_X} & X \times X \\ f \downarrow & & \downarrow f \times f \\ Y & \xrightarrow{\Delta_Y} & Y \times Y \end{array} \qquad \begin{array}{ccc} H^*(f^*\mathcal{F}) & \xleftarrow{\smile_X} & H^*(f^*\mathcal{F}) \otimes H^*(f^*\mathcal{F}) \\ \uparrow & & \uparrow \\ H^*(\mathcal{F}) & \xleftarrow{\smile_Y} & H^*(\mathcal{F}) \otimes H^*(\mathcal{F}) \end{array}$$

Indeed, we need only adjust the second diagram to bring the ring multiplication map of \mathcal{F} out. But this follows from naturality, i.e. the left square in the diagram below commutes:

$$\begin{array}{ccccc} & & \xleftarrow{\smile_X} & & \\ & & \xleftarrow{\quad} & & \\ H^*(f^*\mathcal{F}) & \xleftarrow{H^*(f^*m)} & H^*(f^*\mathcal{F} \otimes f^*\mathcal{F}) & \xleftarrow{\quad} & H^*(f^*\mathcal{F}) \otimes H^*(f^*\mathcal{F}) \\ \uparrow f^* & & \uparrow f^* & & \uparrow \\ H^*(\mathcal{F}) & \xleftarrow{H^*(m)} & H^*(\mathcal{F} \otimes \mathcal{F}) & \xleftarrow{\quad} & H^*(\mathcal{F}) \otimes H^*(\mathcal{F}) \\ & & \xleftarrow{\smile_Y} & & \end{array}$$

The rest of the proof is just that the two paths in the right half of the diagram are a pair of alternative cohomological pullbacks from $X \times X$ to Y , hence equal.

By a similar token, we can show that the cup product is an algebra structure on $H^*(\mathcal{F})$. We will gloss over the details that the ring multiplication map can be factored out of the diagrams expressing commutativity, unitarity, and associativity, after which they all express aspects of the functorial nature of cohomological pullback again.

We can go further in the case that $X = G$ is a group scheme, whose multiplication map induces a *coalgebra* structure on $H^*(G, \mathcal{F})$ for any sheaf \mathcal{F} , with comultiplication defined by cohomological pullback along $G \times G \rightarrow G$ and counit being pullback along $*$ $\rightarrow G$. Once again, functoriality of the pullback proves that this is indeed a coalgebra. Taken together with the cup product we easily combine the structures into a Hopf algebra, since (modulo writing the naturality diagrams that pull $H^*(m)$ out) all the compatibilities express equality of various iterated cohomological pullbacks.

This is standard algebraic topology material and nothing in this example has actually cited the main theorem directly; we include it only to note that it rests on a fact, functoriality of cohomological pullbacks, that is a corollary of our main theorem, as well as to indicate the relevance of this setup to well-known constructions.

B. Fundamentals of string diagrams

Here is an overview of the use of pictorial notation for category-theoretic computations, with an emphasis on applications to this paper, together with proofs of the diagram equivalences presented in [Section 3](#).

String diagrams are used to depict the algebra of category theory visually, and there appears to be a variety of styles in which they are drawn. These topological representations of natural transformations (introduced by Penrose [[Pen71](#), [PR87](#)] and whose first introduction into the pure mathematics literature seems to have been in [[JS91](#)]; an instructive introduction is given by the video series [[Cat](#)]) seem to have the same mysterious effectiveness in category theory that the Leibniz notation does in calculus, often indicating algebraic truths through visual intuition. It does not appear that the topological data is intrinsically important, though some experience with string diagrams suggests that removing loops is an essential first step in reducing them.

OVERVIEW OF STRING DIAGRAMS. String diagrams are a notational paradigm for representing natural transformations and, in particular, for making the following concepts from category theory more amenable to intuitive manipulation.

NOTATION. Functors and natural transformations can be combined in composition in several ways, reflecting the bicategorical structure of the category of small categories. Two functors may of course be composed, which we denote by pure juxtaposition without the symbol \circ . If $\phi: A \rightarrow B$ is a natural transformation of two functors A and B , and if F is a functor such that FA and FB are defined, then we write $F\phi$ to mean the associated natural transformation $FA \rightarrow FB$ such that $F\phi_x = F(\phi_x)$ for any object x . Likewise,

if AF and BF are defined, then we write ϕF to mean the functor $AF \rightarrow BF$ such that $\phi F_x = \phi_{F(x)}$.

We use the common term ‘‘horizontal composition’’ of natural transformations $\phi: A \rightarrow B$ and $\psi: C \rightarrow D$ to mean the following natural transformation $AC \rightarrow BD$, which we will write as simply $\phi\psi$ without ambiguity:

$$\begin{array}{c} A \\ \xrightarrow{\phi} B \\ C \xrightarrow{\psi} D = D \end{array} = \begin{array}{c} A \\ \xrightarrow{\phi} B \\ C \xrightarrow{\psi} D \end{array} = \begin{array}{c} A \\ \xrightarrow{\phi} B = B \\ C = C \xrightarrow{\psi} D \end{array} \quad (\text{B.1})$$

INTRODUCTION TO STRING DIAGRAMS. Every string diagram depicts a single natural transformation of functors, where all compositions (of both functors and transformations) are made explicit. Its main feature is a web of continuous paths recording the history of each functor as it is transformed; these transformations occur at the intersections. The connected components of the complement of this web represent the categories related by these functors, with the direction of composition being the same as in any diagram $X \rightarrow Y \rightarrow Z$. Thus, each functor is like a river flowing between two banks, and to apply the functor is to cross the river. The intersections of paths are natural transformations, with the direction of composition being upwards. In this paper, the allowable intersections are those given in [Figure 1](#) representing as marked the basic SGNTs of [Definition 1.3](#). We give a precise definition of the string diagrams we use:

B.1. DEFINITION. *A string diagram is a planar graph that is the union of shapes of the general form shown in [Figure 1](#), where the various segments may have any lengths. These shapes may only intersect at their ends, and conversely, any end of a shape in the diagram must either join another shape, or be continued upward or downward to an infinite vertical ray. The vertices of the graph are either the univalent caps of the triv^* or triv_* shapes, the trivalent intersections of the comp^* or comp_* shapes, or the right-angle turns of the various shapes; the edges are the connected line segments in the complement of the vertices. A string diagram may also have bi-infinite, directed vertical lines disjoint from the other portions.*

We write our diagrams with their edges doubled. This has no practical meaning, but aside from making the diagrams more aesthetic, it provides some topological intuition that will be explained later. There is a correspondence between SGNTs and string diagrams labeled as in [Figure 1](#).

B.2. DEFINITION. *Let D be a string diagram with its edges labeled; then any horizontal slice not containing a vertex of D determines an SGF by composing the labels left-to-right (rather than right-to-left, as is traditional in algebraic notation). In particular, its eventual slices in the two vertical directions determine two SGFs F and G . We describe how to obtain an SGNT $\phi_D: F \rightarrow G$; it satisfies the following properties:*

- *Each of the diagrams of [Figure 1](#) has the interpretation as an SGNT given there.*

- If D is split by any simple path extending to horizontal infinity in both directions and not containing any vertices or crossing any horizontal edge (the latter, a restriction born of our choice of visual style), then both the lower and upper parts D_1 and D_2 are string diagrams (after extending their cut edges to infinity). We have $\phi_D = \phi_{D_2}\phi_{D_1}$. That is, vertical composition of SGNTs is vertical concatenation of diagrams.
- If the complement of D contains a simple path extending to vertical infinity in both directions, then its left and right parts D_1 and D_2 are string diagrams and ϕ_D is the horizontal composition of ϕ_{D_2} after ϕ_{D_1} . That is, horizontal composition of SGNTs is horizontal concatenation of diagrams.

Under this interpretation, we have the following correspondence:

- The complement of D in \mathbb{R}^2 consists of finitely many connected open sets, each of which represents a category.
- Each vertical edge is a functor from the category on its left to the one on its right; an upward edge is an f_* , and a downward edge is an f^* (thus, the direction of f is always left-to-right when facing along the edge).
- A natural transformation occurs at any horizontal edge (including those of zero length at the valence-1 vertices). Its direction is from the functor below it to the functor above.

We will say that two diagrams are equivalent if they define equal transformations.

PROOFS OF STRING DIAGRAM IDENTITIES. In this subsection we collect the proofs from [Section 3](#), which are by and large trivial translations of symbolic category-theory notation into string diagrams.

PROOF OF [LEMMA 3.1](#). These diagrams correspond to the fundamental identities [\(1.2\)](#) of the unit and counit that ensure that they define an adjunction. ■

PROOF OF [LEMMA 3.2](#). The left side is the string diagram equivalent of [\(1.4\)](#). The right side follows by Yoneda's lemma from [\(1.5\)](#) and the left side. ■

PROOF OF [LEMMA 3.3](#). The left two diagrams express in string diagram language that \mathbf{comp}_* comes in inverse pairs, which we have assumed by definition. The right two also express this, for \mathbf{comp}^* , which is *not* a definition, but follows immediately from [\(1.5\)](#) and Yoneda's lemma. ■

PROOF OF [LEMMA 3.4](#). The first and third follow from the definition, since \mathbf{triv}_* comes by hypothesis in an inverse pair. The second and fourth express the same for \mathbf{triv}^* and follow from [\(1.7\)](#) by Yoneda's lemma. ■

PROOF OF [LEMMA 3.5](#). The third and fourth express the correspondence [\(1.7\)](#) of \mathbf{triv}^* and \mathbf{triv}_* via Yoneda's lemma. The first and second express the same correspondence of their inverses. ■

PROOF OF LEMMA 3.6. The entire first row of equalities is the string diagram version of (1.6), which we assume to hold. The second row follows from this equation and from (1.5) and (1.7) by Yoneda’s lemma. ■

PROOF OF LEMMA 3.7. These correspond to the assumed compatibility of comp_* with comp^* via adjunction expressed in (1.5). This compatibility entails the equality of the two units, respectively the two counits, of the adjunctions for the two sides of $(fg)^* = g^*f^*$ (resp. of $(fg)_* = f_*g_*$):

(B.2)

If we apply the second row of Lemma 3.3 to each of the four composition shapes in these two diagrams, we get the four diagrams of Lemma 3.7. ■

PROOF OF LEMMA 3.8. We illustrate the upper-left one; the others are formally identical.

(B.3)

using first Lemma 3.1 and then Lemma 3.7. ■

PROOF OF LEMMA 3.9. This interesting diagram is easy to prove from the previous ones. We will omit the labels, since they can be inferred by comparison with the figure. In the equalities below, matching blue and red indicate which portions of the diagram are changed; violet indicates a shape that is both blue and red.

(B.4)

using the second row of Lemma 3.3, then Lemma 3.2, then the first row of Lemma 3.3. The proof of the $*$ version is identical with the appropriate change of notation (or follows by Yoneda’s lemma from (1.5)). ■

PROOF OF LEMMA 3.10. Again using colors to indicate the affected regions, we have

(B.5)

by Lemma 3.9 (twice) and then Lemma 3.3. ■

TOPOLOGICAL INTUITION AND APPLICATION TO OBJECTS. It is a tautology of string diagrams that they are useful because they are visual; our chosen notation, through a combination of design and good fortune, happens to display a remarkable compatibility with topological intuition that has been commented on several times earlier in this paper.

The general principle of our string diagrams is that topologically identical diagrams are equivalent. We don't know how to prove this in that kind of generality, though a close examination of our many figures will show that they are all topological trivialities (i.e. diagrams equivalent through ambient isotopy are equal as natural transformations), with the exception of Lemmas 3.3 and 3.4, if one takes the double-line notation seriously (this is the reason for its invention). In fact, if in the first line of the latter figure one removes the inner loop formed by one of the sets of doubled lines, the identity is again topologically accurate. However, it is easy to give a simple version that is almost too subtle to remark upon. Indeed, we have never invoked it directly, but it is used constantly to rearrange diagrams to our liking.

B.3. LEMMA. *Let $D \subset \mathbb{R}^2$ be a string diagram and let $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be any continuous map whose restriction to each edge of D is an affine map with positive scale factor in the direction (horizontal or vertical) of that edge. Then $f(D)$ is an equivalent string diagram to D .*

PROOF. We see, first, that $f(D)$ is a string diagram, since each basic shape is transformed by f into a basic shape of the same type, and any edge abuts the same basic shapes or is infinite in both diagrams by continuity and positive affinity of f . Furthermore, f induces a bijection between components of $\mathbb{R}^2 \setminus D$, and any two adjacent components are separated by an edge of the same direction; thus, the interpretation of each edge as a map of schemes remains valid. To see that the diagrams are equivalent, first consider two distinct connected components of D , if they exist. By a simple induction, each of them is equivalent, as a string diagram, to its image under f , and these images are translated with respect to each other as compared to the originals. But translational motion is an equivalence by naturality. ■

Ours is not the only notation for string diagrams; in fact, it appears to differ in some particulars from others commonly used. McCurdy and Melliès [McC12, Mel06], following Cockett and Seely [CS99], use “functorial boxes” rather than line segments to denote functors, but this reflects their different application of the string diagram visualization: their diagrams denote objects of a monoidal category and their morphisms, rather than functors and their natural transformations. In fact, our notation is more general and can easily be extended to display the specific instances of functors and natural transformations obtained by evaluating them at individual objects of the underlying categories. However, we believe that string diagram notation is inherently domain-specific, depending on the natural properties and relationships of the functors and transformations it describes, so it is unlikely that there is one single, universally effective style. That said, the box notation is entirely free of features that display the nature of the functors that appear in it.

As an example of the two diagram styles, we present the definition of a monoidal

functor and a monoidal natural transformation between two monoidal functors. In the notation of Melliès (which is particularly attractive), the monoidality of a functor f^* is expressed in Figure 3. The meaning of the box is that, inside, the f^* is “stripped”

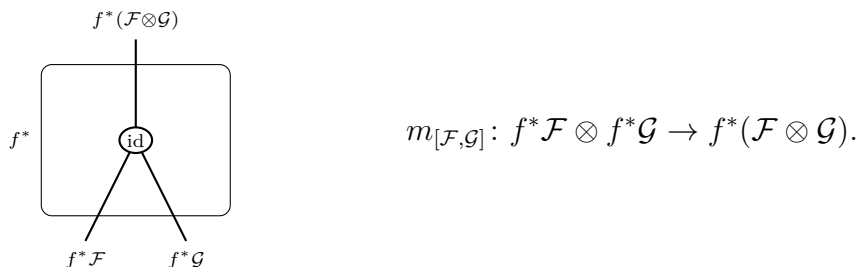


Figure 3: Monoidal functor à la Melliès

from the objects; the name of the map $m_{[F,G]}$ is not mentioned, as it is implicit to the diagram, which is specialized to the needs of monoidal categories. By comparison, the monoidality of f^* may be expressed in our notation as in Figure 4. In each of its diagrams,

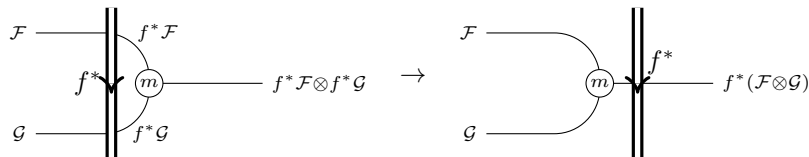


Figure 4: Monoidal functor in our notation

the background layer is a horizontal string diagram with the strings being objects of the category corresponding to the planar region containing them (according to Definition B.2) and their intersections being transformations from left to right (here, marked, as we have no specialized notation for monoidal categories as we do for geofibered ones). Vertical stacking of objects’ strings corresponds to their monoidal product.

Unlike in the purely functorial diagrams, each individual diagram does not depict a morphism m ; rather, it denotes the endpoints of such a morphism (the rightmost labeled object in each) and the “location” of m . The morphism itself is obtained by “sweeping” the functorial diagram over the objects from left to right, as shown above. Whenever it crosses a marked morphism, the motion corresponds to the introduction of that map between the “endpoint” on the left and the one on the right. (Obviously, we must forbid endpoint diagrams that would allow degenerate or ambiguous configurations.)

Thus, the diagrams display the history of an object and the *potential* natural morphisms, but it is the topological relationships between them that give the actual maps.

As a further example of this, consider the monoidality of the natural transformation $\theta = \text{comp}^*(g, f)$. In Melliès’ notation, it is the equation of diagrams in Figure 5. The

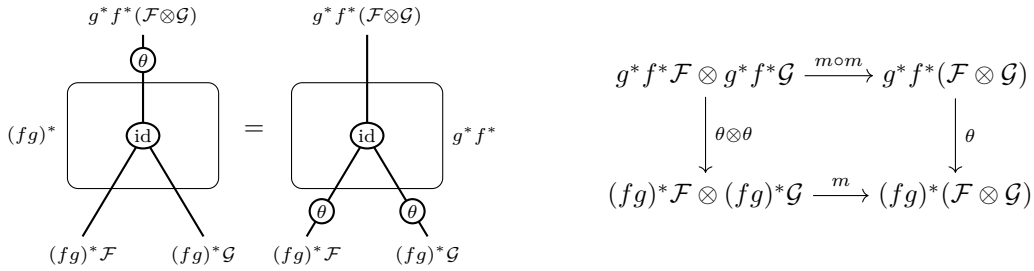


Figure 5: Monoidal transformation à la Mellies

interpretation is straightforward; vertical juxtaposition of morphisms is composition and horizontal is monoidal product. In our notation, we express it in Figure 6 as a closed loop of translations. Moving the functorial “foreground” from left to right, as before,

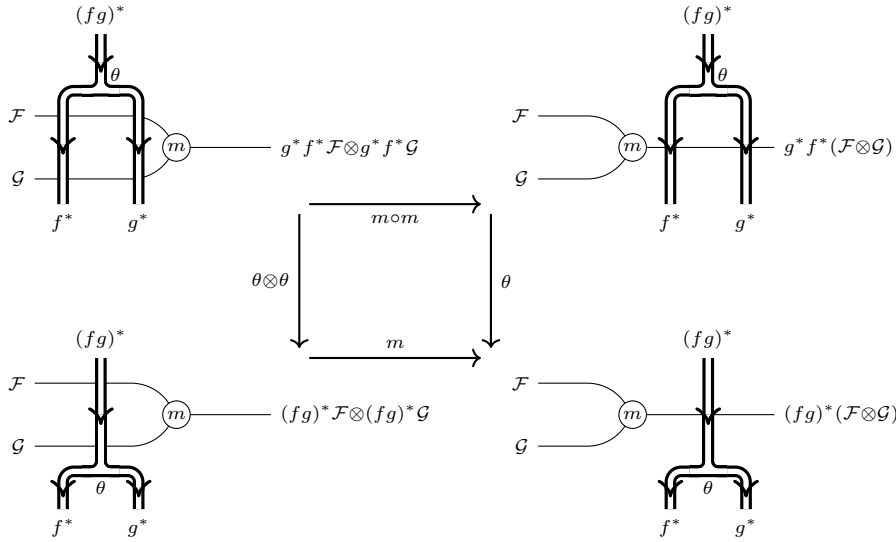


Figure 6: Monoidal transformation in our notation

produces one or two iterations of m , depending on how many of its strings cross that vertex. Moving the objects’ “background” from bottom to top produces an instance of the natural transformation $\text{comp}^*(f, g)$ (the one on the object $\mathcal{F} \otimes \mathcal{G}$ for the θ edge; for each factor of $\theta \otimes \theta$, we get the instances on \mathcal{F} or \mathcal{G} individually). Since the translational motions commute, so does the diagram of maps. We cannot help but feel that this is an instance of the Eckmann–Hilton argument for the commutativity of higher homotopy groups (as depicted in Hatcher’s book [Hat02, p. 340]).

The fact that we have expressed a single natural morphism, as in Figure 4, in terms of an equivalence of several diagrams, would seem to be at odds with our intention of

using individual diagrams to express natural transformations. In fact, this morphism could be expressed in a purely functorial notation with some additional convention as to the appearance of the “monoidal product string”, which would involve the complication of expressing its bivariant nature. However, it is actually *this* approach that is at odds with the philosophy of our diagrams, since the morphism of Figure 4 is not a “structural” morphism but rather a “compatibility” between two independent structures. Moreover, the structures are not of an equivalent nature: the monoidal product is “internal” and the pullback functor (or any monoidal functor) is “external”. So it is entirely appropriate that their compatibility should be expressed through superimposed string diagrams of which one represents objects in the categories cut out by the other.

In summary: in this paper we have used “string diagrams” representing functors and their natural transformations, but we can “specialize” them to the corresponding values and their natural maps when applied to specific objects by superimposing such a diagram on a horizontal diagram of objects and morphisms in the categories cut out by the functorial diagram. Then any continuous motion (avoiding some degeneracies) “instantiates” the morphisms when the strings of one diagram cross the vertices of the other, and any closed loop should represent a commutative diagram of these maps.

We feel that this topological intuition is a valuable asset in a graphical algebraic notation, since it elevates the picture from simply a two-dimensional symbolic calculus to a genuinely graphical reasoning tool.

References

- [SGA IV₃] Michael Artin, *Théorie des topos et cohomologie étale des schémas, 1963–1964* **305** (1973), 64–78. Exposés xii–xvi.
- [Cat] The Catsters, *String diagrams, adjunctions, and monads* (2007), <http://www.youtube.com/playlist?list=PL50ABC4792BD0A086>. Accessed 2014-04-06. Youtube video series.
- [CS99] J. R. B. Cockett and R. A. G. Seely, *Linearly distributive functors*, J. Pure Appl. Algebra **143** (1999), no. 1-3, 155–203. Special volume on the occasion of the 60th birthday of Professor Michael Barr (Montreal, QC, 1997).
- [Con00] Brian Conrad, *Grothendieck duality and base change*, Lecture Notes in Mathematics, vol. 1750, Springer-Verlag, Berlin, 2000.
- [FHM03] H. Fausk, P. Hu, and J. P. May, *Isomorphisms between left and right adjoints*, Theory Appl. Categ. **11** (2003), No. 4, 107–131. MR1988072 (2004d:18006)
- [Hat02] Allen Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002. Available at <http://www.math.cornell.edu/~hatcher/AT/ATpage.html>.
- [JS91] André Joyal and Ross Street, *The geometry of tensor calculus. I*, Adv. Math. **88** (1991), no. 1, 55–112.

- [LZ12] Yifeng Liu and Weizhe Zhang, *Enhanced six operations and base change theorem for sheaves on Artin stacks* (2012), <http://arxiv.org/abs/1211.5948>. Accessed 2014-04-06.
- [Lur05] Jacob Lurie, *Tannaka Duality for Geometric Stacks* (2005), <http://arxiv.org/abs/math/0412266>. Accessed 2013-10-11. AIM 2005-6.
- [ML63] Saunders Mac Lane, *Natural associativity and commutativity*, Rice Univ. Studies **49** (1963), no. 4, 28–46.
- [McC12] Micah Blake McCurdy, *Graphical methods for Tannaka duality of weak bialgebras and weak Hopf algebras*, Theory Appl. Categ. **26** (2012), No. 9, 233–280.
- [Mel06] Paul-André Mellies, *Functorial boxes in string diagrams*, Computer science logic, Lecture Notes in Comput. Sci., vol. 4207, Springer, Berlin, 2006, pp. 1–30.
- [Mil80] James S. Milne, *Étale cohomology*, Princeton Mathematical Series, vol. 33, Princeton University Press, Princeton, N.J., 1980.
- [Pen71] Roger Penrose, *Applications of negative dimensional tensors*, Combinatorial Mathematics and its Applications (Proc. Conf., Oxford, 1969), Academic Press, London, 1971, pp. 221–244.
- [PR87] Roger Penrose and Wolfgang Rindler, *Spinors and space-time. Vol. 1*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, Cambridge, 1987. Two-spinor calculus and relativistic fields.
- [Stacks] The Stacks Project Authors, *Stacks Project* (2014), <http://stacks.math.columbia.edu>. Accessed 2014-04-06.
- [Vis05] Angelo Vistoli, *Grothendieck topologies, fibered categories and descent theory*, Fundamental algebraic geometry, Math. Surveys Monogr., vol. 123, Amer. Math. Soc., Providence, RI, 2005, pp. 1–104. Available at <http://arxiv.org/abs/math/0412512>.

Department of Mathematics, University of Michigan
530 Church Street
Ann Arbor, MI 48103
Email: ryancr@umich.edu

This article may be accessed at <http://www.tac.mta.ca/tac/>

THEORY AND APPLICATIONS OF CATEGORIES (ISSN 1201-561X) will disseminate articles that significantly advance the study of categorical algebra or methods, or that make significant new contributions to mathematical science using categorical methods. The scope of the journal includes: all areas of pure category theory, including higher dimensional categories; applications of category theory to algebra, geometry and topology and other areas of mathematics; applications of category theory to computer science, physics and other mathematical sciences; contributions to scientific knowledge that make use of categorical methods.

Articles appearing in the journal have been carefully and critically refereed under the responsibility of members of the Editorial Board. Only papers judged to be both significant and excellent are accepted for publication.

Full text of the journal is freely available in .dvi, Postscript and PDF from the journal's server at <http://www.tac.mta.ca/tac/> and by ftp. It is archived electronically and in printed paper format.

SUBSCRIPTION INFORMATION Individual subscribers receive abstracts of articles by e-mail as they are published. To subscribe, send e-mail to tac@mta.ca including a full name and postal address. For institutional subscription, send enquiries to the Managing Editor, Robert Rosebrugh, rrosebrugh@mta.ca.

INFORMATION FOR AUTHORS The typesetting language of the journal is $\text{T}_\text{E}\text{X}$, and $\text{L}^{\text{A}}\text{T}_\text{E}\text{X}2\text{e}$ strongly encouraged. Articles should be submitted by e-mail directly to a Transmitting Editor. Please obtain detailed information on submission format and style files at <http://www.tac.mta.ca/tac/>.

MANAGING EDITOR. Robert Rosebrugh, Mount Allison University: rrosebrugh@mta.ca

$\text{T}_\text{E}\text{X}$ TECHNICAL EDITOR. Michael Barr, McGill University: barr@math.mcgill.ca

ASSISTANT $\text{T}_\text{E}\text{X}$ EDITOR. Gavin Seal, Ecole Polytechnique Fédérale de Lausanne: gavin_seal@fastmail.fm

TRANSMITTING EDITORS.

Clemens Berger, Université de Nice-Sophia Antipolis: cberger@math.unice.fr

Richard Blute, Université d' Ottawa: rblute@uottawa.ca

Lawrence Breen, Université de Paris 13: breen@math.univ-paris13.fr

Ronald Brown, University of North Wales: [ronnie.profbrown\(at\)btinternet.com](mailto:ronnie.profbrown(at)btinternet.com)

Valeria de Paiva: valeria.depaiva@gmail.com

Ezra Getzler, Northwestern University: [getzler\(at\)northwestern\(dot\)edu](mailto:getzler(at)northwestern(dot)edu)

Kathryn Hess, Ecole Polytechnique Fédérale de Lausanne: kathryn.hess@epfl.ch

Martin Hyland, University of Cambridge: M.Hyland@dpmms.cam.ac.uk

Anders Kock, University of Aarhus: kock@imf.au.dk

Stephen Lack, Macquarie University: steve.lack@mq.edu.au

F. William Lawvere, State University of New York at Buffalo: wlawvere@buffalo.edu

Tom Leinster, University of Edinburgh: Tom.Leinster@ed.ac.uk

Ieke Moerdijk, Radboud University Nijmegen: i.moerdijk@math.ru.nl

Susan Niefield, Union College: niefiels@union.edu

Robert Paré, Dalhousie University: pare@mathstat.dal.ca

Jiri Rosicky, Masaryk University: rosicky@math.muni.cz

Giuseppe Rosolini, Università di Genova: rosolini@disi.unige.it

Alex Simpson, University of Edinburgh: Alex.Simpson@ed.ac.uk

James Stasheff, University of North Carolina: jds@math.upenn.edu

Ross Street, Macquarie University: street@math.mq.edu.au

Walter Tholen, York University: tholen@mathstat.yorku.ca

Myles Tierney, Rutgers University: tierney@math.rutgers.edu

Robert F. C. Walters, University of Insubria: robert.walters@uninsubria.it

R. J. Wood, Dalhousie University: rjwood@mathstat.dal.ca