A Categorical Framework for Coherence Theorems

Nelson Niu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee:

XiaoLin Danny Shi, Chair

John Palmieri

Jian James Zhang

Program Authorized to Offer Degree:
Mathematics

University of Washington

Abstract

A Categorical Framework for Coherence Theorems

Nelson Niu

Chair of the Supervisory Committee:
XiaoLin Danny Shi
Department of Mathematics

Categories with coherently associative, commutative, and distributive products encapsulate higher algebraic structures throughout mathematics: for instance, in homotopy theory, they are the inputs to May's infinite loop space machines, Segal's K-theory, and multifunctorial, multiplicative, and/or equivariant analogues of these by Elmendorf–Mandell, Guillou–May–Merling–Osorno, and Yau. Here we establish a general categorical approach to proving Mac Lane-like coherence theorems versatile enough to incorporate (weak) distributivity laws, module and algebra categories, bicategories, and the higher arity twisted products that appear in equivariant settings. Building on Mac Lane's original proof of his coherence theorem for (symmetric) monoidal categories and Rubin's coherence theorem for his equivariant normed symmetric monoidal categories, we employ tools from combinatorics, logic, and rewriting theory such as Newman's Diamond Lemma to solve categorical normalization problems on the universal parameter categories representing categorical structures of interest. Our approach clarifies the necessary coherence axioms and invariants. We aim to leverage our work to simplify the characterization of bimonoidal categorical input to Yau's multifunctorial equivariant algebraic K-theory.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I need hardly mention that mathematics can be a daunting subject, but it is truly such a joyous endeavor when one has the incredible fortune to be supported by mentors and colleagues who are as brilliant as they are kind. First and foremost, I must thank my advisor, Danny Shi, for his immensely practical wisdom and guidance. It is an honor to be one of his first students in what I am sure will be a long and illustrious list. Danny always urged me to reach out to other mathematicians whose work I respected, and he provided me with so many opportunities at which to meet potential collaborators, opportunities that I can only fully appreciate in hindsight.

Many thanks as well to Jon Rubin, really my unofficial coadvisor, the most generous mathematical collaborator and mentor that I could have asked for. Despite his full-time parental duties, he has never failed to promptly answer my questions or read my writing, caring deeply for my well-being while expecting very little in return. I thank him for his advice and his patience, his dedication and his trust; and I hope this is only the start of our work together.

Thank you to Angélica Osorno, who introduced me to equivariant multiplicative infinite loop space machines and to Jon; and thank you to Kyle Ormsby, who introduced me to equivariant operads in homotopy theory and to Angélica. And thanks to Peter May for encouraging me at some really critical moments.

Thank you to David Spivak and Brendan Fong for being my first category theory teachers and for giving me so many chances to prove myself as a researcher in the subject. Thank you to Valeria de Paiva, who was my first graduate research mentor and never hesitated to offer her support. Thank you to my committee members, John Palmieri and James Zhang, who have taught me well. And thank you to the countless other mathematicians I have had enlightening conversations or collaborations with over the years, who listened when

I spoke about my work and offered their insights, including but not limited to Angeline Aguinaldo, Scott Balchin, Kristine Bauer, Eric Bond, Spencer Breiner, Kris Brown, Matteo Capucci, Eugenia Cheng, Henry Cohn, David DeMark, Joey Dorta, Bruno Gavranović, John Greenlees, Sina Hazratpour, Mike Hill, Tim Hosgood, Samantha Jarvis, Yigal Kamel, Jérémie Koenig, Ian Levitt, Xiaoyan Li, Sophie Libkind, Juxin Liu, Monte Mahlum, Haynes Miller, Morgan Opie, Nate Osgood, Evan Patterson, Piotr Pstrągowski, Julie Rasmusen, Emily Riehl, Brandon Shapiro, Priyaa Srinivasan, Kurt Stoeckl, Danika Van Niel, Jonathan Weinberger, David White, Guoqi Yan, and Jacob Zelko. Each of you has truly inspired me in your own way. Special thanks to my SUBgroups teammates, ever dedicated and selfless: Marissa Loving, Justin Lanier, Katie Waddle, Marquia Williams, Yufei Zhang, Chi Nguyen, and Katie Kruzan.

And thank you so much to all my fellow graduate students at the University of Washington: you were absolutely the best part of this experience. I could not have asked for a better community of peers, colleagues, comrades, friends. Thank you to my fellow graduating topologists, Jackson Morris and Alex Waugh, and the larger topology crew: Collin, Jay, Manyi, Albert, Martin, Clarice, and sometimes Justin and Bryan. Thank you to my fellow occupants of the best office: David, Catherine, Curtiss, Cameron, Leo, Michael Z., and Miles.

Thank you to Natasha Crepeau, my comrade-in-arms; to Alex Wang, my doppelgänger; to Tracy and Yirong, my detectives; to Michael Tang, though we'll finally be apart; to Charlie, Lauren, and Mal, our all-gender restroom champions; to Jessie and Caelan, despite their questionable literary tastes; to board game night hosts Grace and Clare; to budding organizers Isaiah, Jonas, Dan, and Sean; to cohortmates Nathan, Baby Alex, Jack, Garrett, Haoming, Andrew, Yue, and Tony; to Zawad, Joe, Spencer, Patrick, Linhang, Jenny, Josh, Ellie, and anyone else who's played a TTRPG with me over the years.

Every grad student needs a hobby. Thank you to every theater artist I had the pleasure of collaborating with in Seattle, including but not limited to Zanne, Biz, Rose Palmieri (*yes* relation), Garth, Mary, Steven, Matthew, Lesley, Aubrey, David, Kaysy, Ariel, Jean, Daniel, John, Amanda, Angel, Valerie, Kasper, Alex, Divya, McKenna, Sumant, Andrew, Anna, Jack, Kiera, Chloe, Ember, Ronan, Brando, Elliott, Jesse, Steph, Lou, Avery, Sage, Lark, Christyn, Henry, Holly Cow, Athena, Jane, Patrick, Tessio, Zu, Sammy, Messiah, Kae, Xavier, and Izzy. And thanks to Montse, who never fails to awe me.

This thesis is union-made. Thank you to all the members of the Academic Student Employee Unit of UAW Local 4121 for putting your trust in me as your representative the past few years. Thank you to Natalie, Sophie, Alana, Stella, Jake, Joey, Joice, Rachael, Maria, Ella, Mickey, Francesca, Tahiyat, Soohyung, Nolawit, June, Peter, Jayden, Anastasia, Miro, Yuying, Ellen, Shirley, Cecily, Mathieu, Izzy, Avi, Candice, Erin, Iván, Brendy, Lou, Abby, Tricia, Sam, Emily, Alli, and so many other fellow members of our Bargaining Team, Joint Council, Organizing Committee, Political Work Group, and International Solidarity Work Group. Thank you also to Angelina Godoy, not only for serving on my committee, but for her pursuit of justice as the Director of the UW Center for Human Rights. We make a difference.

Thank you to my sweetest cat Caroline, who is always a joy to come home to. Thank you to my family: to Mom and Dad and Sophie, to Jerry and Wency, to Mom and Dad Gillam, to Axel, to all my aunts and uncles, to my loving grandparents. And thank you to my spouse, Eryn Gillam, the most thoughtful and caring partner I could possibly have asked for.

DEDICATION

for Eryn

Chapter 1

INTRODUCTION

In his groundbreaking book [17], May introduced a categorical gadget called an *operad* to parametrize the algebraic structure of both topological spaces and spectra, the fundamental objects of study in stable homotopy theory. These operads, when enriched over a convenient category of spaces, can characterize the way an addition $\oplus$ operation on a space or spectrum, while not necessarily commutative on the nose, nevertheless can satisfy a *homotopy commutativity* condition guaranteeing the existence of a homotopy between maps $(a, b) \mapsto a \oplus b$ and $(a, b) \mapsto b \oplus a$; a higher homotopy between any two such homotopies; and so on. We call a space equipped with such a homotopy commutative operation an *E_∞-space*; it is parametrized by an *E_∞-operad*.

Leveraging the power of abstraction that operads afford, May constructed an *infinite loop space machine*: a functor that sends E_∞-spaces to a broad class of spectra. As spectra are equivalent to generalized cohomology theories, they can often carry an intractable amount of data, but an infinite loop space machine lets us study spectra using the more familiar structure of spaces. The conditions for a space to be an E_∞-space, however, are often just as intractable, requiring the existence of infinitely many families of infinitely many homotopies of arbitrary height between operations of arbitrary arity.

Fortunately, such conditions are familiar to category theorists. It turns out that Mac Lane's coherence theorem for symmetric monoidal categories states that a category with a nullary operation, a binary operation, four families of isomorphisms between operations induced by the original two, and five families of equations satisfied by isomorphisms induced by the original four generate operations of arbitrary arity and guarantee the existence and uniqueness of isomorphisms between operations of the same arity [15, 16]. This is a far more manageable list of conditions to check than those of an E_∞-space, and verifications that a given category equipped with the appropriate operations and natural isomorphisms are standard (that a category with finite products is symmetric monoidal, for instance, is folklore). There is then an analogy between categories with operations, natural isomorphisms between those operations, and coherence equalities between those isomorphisms; and spaces with operations, homotopies between those operations, and higher homotopies between those homotopies. Formalizing and exploiting this analogy by way of the *classifying space functor*, we find that every symmetric monoidal category gives rise to an E_∞-space. May with his operadic infinite loop space machine [18] and Segal with his K-theory functor [25] independently applied this idea to turn symmetric monoidal categories into spectra.

From here there are two directions along which to generalize. Not only would we like to send additive symmetric monoidal categories to spectra, but we would also like to send distributive multiplicative structures on such additive symmetric monoidal categories to *structured ring spectra*, which are spectra that carry a multiplicative E_∞-structure themselves. May achieves this through his *operad pairs* in [19, 20, 21], while Elmendorf and Mandell use a multi-object generalization of operads and operad maps called *multicategories* and *multifunctors* to accomplish this in [3]. Their inputs are categories with two symmetric monoidal structures equipped with a distributivity natural transformation called a *symmetric bimonoidal category*, which satisfies a generalization of Mac Lane's coherence theorem proven by Laplaza [13].

The other direction is to extend to a G-equivariant setting for a finite group G. The theory of N_∞-operads, the G-equivariant analogue of E_∞-operads, is more complicated than that of E_∞-operads, as they parametrize transfer maps in the additive setting and norm maps in the multiplicative setting. Hill, Hopkins, and Ravenel's celebrated solution to the Kervaire invariant one problem in [7] leveraged and renewed interest in these multiplicative norms arising from G-equivariant spectra, leading Blumberg and Hill to introduce N_∞-operads in [2]. Rubin's *normed symmetric monoidal categories* from [24] then give rise to spaces parametrized by an arbitrary (weak equivalence class of) N_∞-operad(s). Yau's G-equivariant machine from [26] further sends multiplicative structures on Guillou, May, Merling, and Osorno's genuine symmetric monoidal G-categories from [6] to analogous multiplicative structures on G-equivariant spectra.

While Rubin's definition of normed symmetric monoidal categories and subsequent coherence theorem in [24] mirrors that of Mac Lane's for symmetric monoidal categories in [15, 16] in the way it generates a large amount of coherence data from a small number of components and axiom families, the same cannot be said of genuine symmetric monoidal G-categories. To address this, we seek a common generalization of Laplaza's coherence theorem for symmetric bimonoidal categories [13] and Rubin's coherence theorem for normed symmetric monoidal categories [24]. This would involve defining and proving a coherence theorem for *normed symmetric bimonoidal categories*. We conjecture that such a definition and a corresponding coherence theorem exists, as do extensions to modules and algebras over these proposed normed symmetric bimonoidal categories.

Indeed, it seems that every kind of infinite loop space machine from structured categories to structured spectra requires a coherence theorem for those structured categories to strictify them and characterize them as (pseudo)algebras over a suitable operad. Mac Lane's coherence theorem for symmetric monoidal categories from [15, 16] is well-known, although its proof is typically glossed over in introductory category courses due to the laborious combinatorial reasoning involved; Laplaza's coherence theorem for symmetric bimonoidal categories from [13] was even more obscure until Yau's account of it in [9, Part I.1].

In this thesis, we develop a generalized categorical framework based on techniques from logic and rewriting theory to generate and prove coherence theorems for the more complicated structures that arise when equivariance, distributivity, and modules and algebras are introduced. There is some work toward this end by Mimram in [22]; we expand it to fully encapsulate Laplaza's coherence theorem, Rubin's coherence theorem, and beyond. In what follows, we shall develop an abstract framework for proving categorical coherence theorems.

In Chapter 2, we will formally encode Mac Lane-like presentations of general algebraic structures on categories as *coherence schemata* that may be interpreted 2-categorically as *universal parameter categories*. We construct such coherence schemata for monoidal and symmetric monoidal categories as well as bicategories. To prove a coherence theorem, we would then only need to work with said universal parameter category. In Chapter 3, we reframe coherence theorems on specific categories as *coherence normalization problems (CNPs)*. We present a wide range of techniques for solving these CNPs and give examples of how to solve them for monoidal and symmetric monoidal categories. Throughout this work, most of the definitions and constructions are joint with Jonathan Rubin, as is the critical Proposition 3.49.

To ground the ideas, we begin by reviewing the example of symmetric monoidal categories. Recall that a *symmetric monoidal category* is a category $\mathscr{C}$ equipped with a bifunctorial *monoidal product*

$$\otimes\colon \mathscr{C} \times \mathscr{C} \to \mathscr{C}$$

and a *unit object* $e \in \mathscr{C}$, such that $\otimes$ is associative, unital, and commutative up to coherent natural isomorphism. More precisely, we have specified natural isomorphisms

$$\alpha\colon (x \otimes y) \otimes z \overset{\sim}{\Longrightarrow} x \otimes (y \otimes z) \qquad \beta\colon x \otimes y \overset{\sim}{\Longrightarrow} y \otimes x$$

$$\lambda\colon e \otimes x \overset{\sim}{\Longrightarrow} x \qquad \rho\colon x \otimes e \overset{\sim}{\Longrightarrow} x$$

such that the five coherence diagrams below commute:

$$\begin{tikzcd}
((w \otimes x) \otimes y) \otimes z \arrow[r, Rightarrow, "\alpha \otimes \mathrm{id}"] \arrow[d, Rightarrow, "\alpha"'] & (w \otimes (x \otimes y)) \otimes z \arrow[r, Rightarrow, "\alpha"] & w \otimes ((x \otimes y) \otimes z) \arrow[d, Rightarrow, "\mathrm{id} \otimes \alpha"] \\
(w \otimes x) \otimes (y \otimes z) \arrow[rr, Rightarrow, "\alpha"] & & w \otimes (x \otimes (y \otimes z)),
\end{tikzcd}$$

$$\begin{tikzcd}
(x \otimes e) \otimes y \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\rho \otimes \mathrm{id}"'] & & x \otimes (e \otimes y) \arrow[dl, Rightarrow, "\mathrm{id} \otimes \lambda"] \\
& x \otimes y, &
\end{tikzcd}$$

$$\begin{tikzcd}
x \otimes y \arrow[rr, equal] \arrow[dr, Rightarrow, "\beta"'] & & x \otimes y \\
& y \otimes x, \arrow[ur, Rightarrow, "\beta"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
x \otimes e \arrow[rr, Rightarrow, "\beta"] \arrow[dr, Rightarrow, "\rho"'] & & e \otimes x \arrow[dl, Rightarrow, "\lambda"] \\
& x, &
\end{tikzcd}$$

$$\begin{tikzcd}
(x \otimes y) \otimes z \arrow[r, Rightarrow, "\alpha"] \arrow[d, Rightarrow, "\beta \otimes \mathrm{id}"'] & x \otimes (y \otimes z) \arrow[r, Rightarrow, "\beta"] & (y \otimes z) \otimes x \arrow[d, Rightarrow, "\alpha"] \\
(y \otimes x) \otimes z \arrow[r, Rightarrow, "\alpha"] & y \otimes (x \otimes z) \arrow[r, Rightarrow, "\mathrm{id} \otimes \beta"] & y \otimes (z \otimes x).
\end{tikzcd}$$

From here, we may use the structure on $\mathscr{C}$ to construct more complicated isomorphisms. For example, we may construct an isomorphism $(x \otimes y) \otimes z \Rightarrow (z \otimes y) \otimes x$ by taking the composite

$$(x \otimes y) \otimes z \xRightarrow{\beta \otimes \mathrm{id}} (y \otimes x) \otimes z \xRightarrow{\beta} z \otimes (y \otimes x) \xRightarrow{\alpha^{-1}} (z \otimes y) \otimes x$$

or by taking the composite

$$(x \otimes y) \otimes z \xRightarrow{\alpha} x \otimes (y \otimes z) \xRightarrow{\beta} (y \otimes z) \otimes x \xRightarrow{\beta \otimes \mathrm{id}} (z \otimes y) \otimes x$$

But in fact, the coherence diagrams above imply that both these isomorphisms $(x \otimes y) \otimes z \Rightarrow (z \otimes y) \otimes x$ are equal, as the third and fifth coherence diagrams imply that

$$\begin{tikzcd}
(x \otimes y) \otimes z \arrow[r, Rightarrow, "\alpha"] \arrow[d, Rightarrow, "\beta \otimes \mathrm{id}"'] & x \otimes (y \otimes z) \arrow[r, Rightarrow, "\beta"] & (y \otimes z) \otimes x \arrow[d, Rightarrow, "\alpha"] \\
(y \otimes x) \otimes z \arrow[r, Rightarrow, "\alpha"] \arrow[d, Rightarrow, "\beta"'] & y \otimes (x \otimes z) \arrow[r, Rightarrow, "\mathrm{id} \otimes \beta"] & y \otimes (z \otimes x) \arrow[d, Rightarrow, "\alpha^{-1}"] \\
z \otimes (y \otimes x) \arrow[r, Rightarrow, "\alpha^{-1}"] & (z \otimes y) \otimes x & (y \otimes z) \otimes x \arrow[l, Rightarrow, "\beta \otimes \mathrm{id}"']
\end{tikzcd}$$

commutes.

Thus far, we have only considered two ways of constructing an isomorphism between $(x \otimes y) \otimes z$ and $(z \otimes y) \otimes x$, but the result we are describing is completely general. In a sense that can be made precise, given n-fold products p_1 and p_2 (each containing arbitrarily many units), there is a unique coherence isomorphism $p_1 \Rightarrow p_2$ that may be constructed from the given isomorphisms α, β, λ, and ρ. This is the coherence theorem for symmetric monoidal categories, originally due to Mac Lane [15] and later simplified by Kelly [10]. A standard reference is [16, VII.2 and XI.1].

That being said, it takes work to make a coherence theorem precise, because there are certain scenarios that need to be excluded. For example, there might be objects $x, y \in \mathscr{C}$ such that $x \otimes y$ and $y \otimes x$ are accidentally equal. Then we could consider the isomorphisms $\beta\colon x \otimes y \Rightarrow y \otimes x$ and $\mathrm{id}\colon x \otimes y \Rightarrow y \otimes x$, and there is no reason that they should be equal. Thus, we should really be considering $x \otimes y$ and $y \otimes x$ as (distinct) formal expressions, rather than as objects of $\mathscr{C}$. But even this perspective is not enough: taking $x = y$, the natural isomorphisms $\beta\colon x \otimes x \Rightarrow x \otimes x$ and $\mathrm{id}\colon x \otimes x \Rightarrow x \otimes x$ are still not necessarily equal, so we must also exclude the case where variables are repeated. Thus, we will borrow a page from first order logic and introduce a formal language that will encode the composite operations that a coherence theorem concerns.

Chapter 2

UNIVERSAL PARAMETER CATEGORIES

In this chapter, we reduce the problem of proving a coherence theorem for a whole family of categories to the problem of proving a coherence theorem for a single category, dubbed a *universal parameter category*, that represents that family. This abstraction allows us to separate the more combinatorial aspect of proving a coherence theorem from the categorical semantics. We draw from techniques in logic to formulate the operations on our categories in purely syntactical terms as *coherence schemata*, before realizing this syntax in the 2-category of categories, functors, and natural transformations (or, more generally, any other 2-category with finite 2-products), recovering Mac Lane-like presentations of a wide range of algebraic structures on categories. Throughout, we use (symmetric) monoidal categories and bicategories as our running examples.

2.1 *Formal languages parametrizing operations*

We begin by laying the groundwork for the purely syntactical aspects of coherence theorems, introducing the formal symbols that we will use to represent the fundamental categorical structures we are concerned with.

Definition 2.1 (Signature)**.** A *signature* $\mathfrak{S}$ consists of:

- a set T of *types*, equipped with a total order $<$;[1]
- a set C of *constant symbols* (or *constants*);
- a set F of *function symbols*;
- a function $\tau\colon C \to T$ called the *type function*;

---

[1]The purpose of this total order will eventually be to ensure that variables in each type are canonically linearly ordered. The order itself may be chosen arbitrarily, such as by the Well-Ordering Theorem.

- a function $\mathrm{dom}\colon F \to \coprod_{n\geq 1} T^n$ called the *domain function*; and

- a function $\mathrm{cod}\colon F \to T$ called the *codomain function*.

Composing the canonical projection $\coprod_{n\geq 1} T^n \to \mathbb{N}_{>0}$ with dom yields a function $\mathrm{ar}\colon F \to \mathbb{N}_{>0}$ called the *arity function*.

For $f \in F$, we write $f\colon (\mathtt{t}_1, \ldots, \mathtt{t}_n) \to \mathtt{t}$ when $\mathrm{dom}(f) = (\mathtt{t}_1, \ldots, \mathtt{t}_n)$ and $\mathrm{cod}(f) = \mathtt{t}$.[2]

Note that our function symbols always have positive arity, while constants may be regarded as nullary.

Example 2.2 (Signature for monoidal categories)**.** Consider the signature $\mathfrak{S}_{\mathrm{M}}$ given by $T \coloneqq \{\mathtt{m}\}$, $C \coloneqq \{e\}$, and $F \coloneqq \{\otimes\}$ with $\mathrm{ar}(\otimes) \coloneqq 2$. Then necessarily $\tau(e) = \mathtt{m}$ and $\otimes\colon (\mathtt{m}, \mathtt{m}) \to \mathtt{m}$. As e and $\otimes$ will respectively represent the unit and product of a monoidal category represented by $\mathtt{m}$, we call $\mathfrak{S}_{\mathrm{M}}$ the *signature for monoidal categories*.

Example 2.3 (Signature for bicategories)**.** Given a (totally ordered) set O, consider the signature $\mathfrak{S}_{\mathrm{B}}(O)$ with $T \coloneqq O \times O$ (with the lexicographic order),

$$C \coloneqq \{\mathbf{1}_a \mid a \in O\}, \quad \text{and} \quad F \coloneqq \{\circ_{a,b,c} \mid a, b, c \in O\},$$

with $\tau(\mathbf{1}_a) \coloneqq (a, a)$ and $\circ_{a,b,c}\colon ((b, c), (a, b)) \to (a, c)$. We call $\mathfrak{S}_{\mathrm{B}}(O)$ the *signature for bicategories with object-set O*:[3] each $\mathbf{1}_a$ will represent an identity 1-morphism, and each $\circ_{a,b,c}$ will represent horizontal composition.

The symbols in a signature will combine to form the terms that will represent the 1-morphisms (e.g. functors) and 2-morphisms (e.g. natural transformations) appearing in our coherence theorems. Throughout this chapter, we fix a signature $\mathfrak{S}$ with types in T, constant and function symbols in C and F, and type, domain, and codomain functions τ, dom, and cod.

---

[2] The domain and codomain functions make $(T, C \sqcup F)$ a *multigraph* in the sense of [14, Section 2.3], where each $c \in C$ may be regarded as a nullary edge $c\colon () \to \tau(c)$.

[3] We work in a sufficiently large Grothendieck universe to avoid size issues, so that the class of objects O of our bicategory of interest is a set. We will use the Well-Ordering Theorem to construct a total order on O.

Definition 2.4 (Alphabet, word, term)**.** The *alphabet* $\mathcal{A}(\mathfrak{S})$ *of* $\mathfrak{S}$ is a set consisting of:

- *variable symbols* (or *variables*) $x_i^{\mathtt{t}}$ for $\mathtt{t} \in T$ and i a positive integer;
- *constant symbols* $c \in C$;
- *function symbols* $f \in F$;
- *punctuation symbols*:
  - the left parenthesis (
  - the right parenthesis)
  - the comma ,

so $\mathcal{A}(\mathfrak{S}) = \{x_i^{\mathtt{t}} \mid i \in \mathbb{N}_{>0}, \mathtt{t} \in T\} \sqcup C \sqcup F \sqcup \{(\} \sqcup \{)\} \sqcup \{,\}$. The variables inherit the lexicographic order on $\mathbb{N}_{>0} \times T$ using the usual order on the integers and the total order $<$ on T, so that $x_i^{\mathtt{t}} < x_j^{\mathtt{u}}$ if either $i < j$ or both $i = j$ and $\mathtt{t} < \mathtt{u}$.

A *word in* $\mathfrak{S}$ is a finite sequence of symbols from $\mathcal{A}(\mathfrak{S})$, written as a string. The *length* of a word w, denoted $\operatorname{len}(w)$, is the number of symbols comprising w.

Define the set $\mathcal{L}(\mathfrak{S})$ of *terms of* $\mathfrak{S}$ as a subset of the words in $\mathfrak{S}$ and an extended *type function* $\tau \colon \mathcal{L}(\mathfrak{S}) \to T$ recursively in parallel as follows:

(1) Every variable $x_i^{\mathtt{t}}$ is a term in $\mathcal{L}(\mathfrak{S})$ of type $\tau(x_i^{\mathtt{t}}) \coloneqq \mathtt{t}$.

(2) Every constant $c \in C$ is a term in $\mathcal{L}(\mathfrak{S})$, and $\tau \colon \mathcal{L}(\mathfrak{S}) \to T$ agrees on constants with the type function $\tau \colon C \to T$ of $\mathfrak{S}$.

(3) Given an n-ary function symbol $f \in F$ and n terms $t_1, \ldots, t_n \in \mathcal{L}(\mathfrak{S})$, if $\operatorname{dom}(f) = (\tau(t_1), \ldots, \tau(t_n))$, then the word $f(t_1, \ldots, t_n)$ is a term in $\mathcal{L}(\mathfrak{S})$ of type $\operatorname{cod}(f)$, the codomain of its leading function symbol.

We call the set $\mathcal{L}(\mathfrak{S})$ the *language of* $\mathfrak{S}$. If a term $t \in \mathcal{L}(\mathfrak{S})$ is a variable or a constant (equivalently, if $\operatorname{len}(t) = 1$), we say that t is *atomic*.

Example 2.5 (Terms for monoidal categories)**.** We let $\mathfrak{S}_{\mathrm{M}}$ be the signature for monoidal categories from Example 2.2. Then its set of terms $\mathcal{L}^{\mathrm{M}} \coloneqq \mathcal{L}(\mathfrak{S}_{\mathrm{M}})$ consists of:

(1) the variable $x_i \coloneqq x_i^{\mathtt{m}}$ for $i > 0$;

(2) the constant e; and

(3) the word $\otimes(s,t)$ for $s, t \in \mathcal{L}^{\mathrm{M}}$.

As $\mathfrak{S}_{\mathrm{M}}$ has only one type, every term has that same type. We call $\mathcal{L}^{\mathrm{M}}$ the *language of monoidal categories.*

Example 2.6 (Terms for bicategories)**.** Let $\mathfrak{S}_{\mathrm{B}}(O)$ be the signature for bicategories with object-set O from Example 2.3. Then its set of terms $\mathcal{L}^{\mathrm{B}}(O) \coloneqq \mathcal{L}(\mathfrak{S}_{\mathrm{B}}(O))$ consists of:

(1) the variable $x_i^{(a,b)}$ for $a, b \in O$ and $i > 0$;

(2) the constant $\mathbf{1}_a$ for $a \in O$ with $\tau(\mathbf{1}_a) = (a,a)$; and

(3) the word $\circ_{a,b,c}(s,t)$ for $a, b, c \in O$ and $s, t \in \mathcal{L}^{\mathrm{B}}(O)$ with $\tau(s) = (b,c)$ and $\tau(t) = (a,b)$, so $\tau(\circ_{a,b,c}(s,t)) = (a,c)$.

We call $\mathcal{L}^{\mathrm{B}}(O)$ the *language of bicategories with object-set O.*

We will often need to consider the set of all variables appearing in our terms. For that, we use the following definition.

Definition 2.7 (Scope, variable-finiteness)**.** For each term $t \in \mathcal{L}(\mathfrak{S})$, a *scope for t* is a finite set of variables V such that every variable that appears in t is an element of V. Here $V = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}$ inherits a total order from the lexicographic order on variables. Furthermore, we let $\mathcal{L}(\mathfrak{S})_V \subset \mathcal{L}(\mathfrak{S})$ be the set of all terms with scope V, we let $V(t)$ be the set of all variables in t, and we let

$$\mathcal{L}(\mathfrak{S})^{\mathrm{sc}} \coloneqq \{(t,V) \mid t \in \mathcal{L}(\mathfrak{S}) \text{ and } V \text{ is a scope for } t\}.$$

We say that a subset of terms $\mathfrak{O} \subset \mathcal{L}(\mathfrak{S})$ is *variable-finite* if the terms share a common scope: that is, only finitely many distinct variables appear among all terms in $\mathfrak{O}$, so $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})_V$ for some finite set of variables V. We write $V(\mathfrak{O})$ to denote the (necessarily finite) set of all variables appearing in $\mathfrak{O}$: it is the intersection of all V for which $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})_V$. Note that $V(\mathcal{L}(\mathfrak{S})_V) = V$.

Example 2.8. Let $V_k \coloneqq \{x_1 < \cdots < x_k\}$ for $k \in \mathbb{N}$. Then we may restrict our language of monoidal categories $\mathcal{L}^{\mathrm{M}}$ from Example 2.5 to terms with scope V_k, i.e. terms whose only variables are those in V_k, yielding $\mathcal{L}_k^{\mathrm{M}} \coloneqq \mathcal{L}(\mathfrak{S}_{\mathrm{M}})_{V_k}$. Then $\mathcal{L}_k^{\mathrm{M}}$ consists of:

(1) the variable x_i for $0 < i \leq k$;

(2) the constant e; and

(3) the word $\otimes(s, t)$ for $s, t \in \mathcal{L}_k^{\mathrm{M}}$.

Given an object-set O and a finite subset $P \subseteq O$, we also restrict our language of bicategories from Example 2.6 by writing $V_k(P) \coloneqq \left\{ x_i^{(a,b)} \;\middle|\; a, b \in P,\ 1 \leq i \leq k \right\}$ for $k \in \mathbb{N}$ and $\mathcal{L}_k^{\mathrm{B}}(P) \coloneqq \mathcal{L}(\mathfrak{S}_{\mathrm{B}}(O))_{V_k(P)}$. Both $\mathcal{L}_k^{\mathrm{M}}$ and $\mathcal{L}_k^{\mathrm{B}}(P)$ are variable-finite.

Our terms will represent the operations we are concerned with in our coherence theorems. But before we give meaning to these terms, we develop the tools for working with them, based on techniques from formal logic.[4] To that end, we will prove the following *Parsing Lemma*, ensuring that when a term is given as a string, we can uniquely determine how it was constructed.

Lemma 2.9 (Parsing Lemma)**.** *Given $t \in \mathcal{L}(\mathfrak{S})$, either t is atomic or t can be written uniquely in the form $f(t_1, \ldots, t_n)$ for a function symbol $f \in F$ and n terms $t_1, \ldots, t_n \in \mathcal{L}(\mathfrak{S})$. Moreover,* $\mathrm{ar}(f) = n$ *and* $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_n))$.

The uniqueness condition of this lemma is more subtle than it appears: for example, we must exclude the possibility that $t = f(s_1, \ldots, s_m)$ for terms $s_1, \ldots, s_m$ for which s_1 is

[4] We could have used trees to represent our syntax visually, as is common in operad theory, computer science, and linguistics. But we choose to use words and terms because they are easier to typeset and to describe using basic mathematical terminology.

a proper prefix of t_1. We therefore introduce terminology for substrings of terms before proving the lemma.

Definition 2.10 (Subword, subterm)**.** Let $w = a_1 \cdots a_n$ be a word (resp. term) comprised of n symbols $a_1, \ldots, a_n$. A *subword* (resp. *subterm*) *of* w is a word (resp. term) of the form $v = a_i \cdots a_j$ for some $1 \leq i \leq j \leq n$ or, by convention (though we will not need it), the empty string. In the nonempty case, we call v the (i,j)*-subword* (resp. (i,j)*-subterm*) *of* w, and we call v a *proper subword* (resp. *proper subterm*) *of* w if $(i,j) \neq (1,n)$. If $i = 1$, we call v an *initial segment of* w, and if further $j \neq n$, we call v a *proper initial segment of* w. Dually, if $j = n$, we call v a *final segment of* w, and if further $i \neq 1$, we call v a *proper final segment of* w. For our purposes, the empty subword is neither an initial nor a final segment.

Example 2.11. Let $\mathfrak{S}_{\mathrm{M}}$ be the signature for monoidal categories from Example 2.2; by Example 2.5, $t = \otimes(\otimes(x_1, x_2), x_3)$ is a term of $\mathfrak{S}_{\mathrm{M}}$. Then (x_1, x_2) is the $(4,8)$-subword of t. The word (x_1, x_2) is not a subterm of t because it is not a term (a term cannot begin with a left parenthesis), but the $(3,8)$-subword $\otimes(x_1, x_2)$ is a subterm of t, because it is itself a term. The words $\otimes(\otimes(x_1$ and $, x_3)$ are proper initial and final segments of t, respectively, but neither is a subterm of t.

We can generalize the final observation of the preceding example by introducing an invariant adapted from [4, Section 2.3] that will distinguish terms from their proper segments.

Definition 2.12 (Charge)**.** Define a function $C \colon \mathcal{A}(\mathfrak{S}) \to \mathbb{Z}$ as follows:

- $C(x_i^{\mathtt{t}}) \coloneqq 1$ for all $i > 0$ and $\mathtt{t} \in T$;
- $C(c) \coloneqq 1$ for all $c \in C$;
- $C(f) \coloneqq -\operatorname{ar}(f)$ for $f \in F$;
- $C(\,(\,) \coloneqq -1$;
- $C(\,,\,) \coloneqq 0$; and

- $C(\,)\,) \coloneqq 2$.

Then for each word $w = a_1 \cdots a_n$ comprised of n symbols $a_i \in \mathcal{A}(\mathfrak{S})$, define $C(w) \coloneqq C(a_1) + \cdots + C(a_n)$. We call $C(w)$ the *charge of* w (by convention, the empty word has charge 0).

Lemma 2.13. *Every term $w \in \mathcal{L}(\mathfrak{S})$ has charge $C(w) = 1$. If u is a proper initial segment of w, then $C(u) < 0$, and if v is a proper final segment of w, then $C(v) > 1$.*

Proof. Induct on w. If w is atomic, its charge is 1, and since a single symbol has no proper initial or final segments, the rest of the claim follows vacuously.

Otherwise, $w = f(t_1, \ldots, t_n)$ for $f \in F$ with $\operatorname{ar}(f) = n > 0$ and $t_1, \ldots, t_n \in \mathcal{L}(\mathfrak{S})$. Suppose the result holds for each $t_i \in \mathcal{L}(\mathfrak{S})$, so in particular $C(t_i) = 1$. As commas contribute zero charge, we have

$$\begin{aligned} C(w) &= C(f) + C(\,(\,) + C(t_1) + \cdots + C(t_n) + C(\,)\,) \\ &= -n - 1 + n + 2 = 1. \end{aligned}$$

Now let u be a proper initial segment of w. If $u = f$ we have $C(u) = -n < 0$. Otherwise, $u = f(t_1, \ldots, t_{k-1}, u_k$ for some $1 \leq k \leq n$, where u_k is either the empty string or an initial segment of t_k. Then $C(u_k) \leq 1$ and

$$\begin{aligned} C(u) &= C(f) + C(\,(\,) + C(t_1) + \cdots + C(t_{k-1}) + C(u_k) \\ &= -n - 1 + k - 1 + C(u_k) \leq -n - 1 + k < 0. \end{aligned}$$

Finally, if v is a proper final segment of w, then $w = uv$ for a proper initial segment u of w, so $C(v) > C(u) + C(v) = C(w) = 1$. $\square$

The following corollary is immediate.

Corollary 2.14. *No proper initial or final segment of a term is a subterm, and no word is both a proper initial segment of a term and a proper final segment of a term.*

We therefore have the following result that will prove our Parsing Lemma.

Corollary 2.15. *Given $m, n > 0$ and terms $s_1, \ldots, s_m, t_1, \ldots, t_n \in \mathcal{L}(\mathfrak{S})$, if we have an equality of words*

$$s_1, \ldots, s_m = t_1, \ldots, t_n$$

then $m = n$ and $s_i = t_i$ for all $1 \leq i \leq n$.

Proof. Induct on n. If $n = 1$, then the term s_1 is an initial segment of the term t_1, so by Corollary 2.14 we must have $s_1 = t_1$ and thus $m = 1 = n$.

If $n > 1$, compare the lengths of s_m and t_n. If $\mathrm{len}(s_m) < \mathrm{len}(t_n)$, then s_m is both a proper final segment of t_n and a term, contradicting Corollary 2.14; and similarly for $\mathrm{len}(s_m) > \mathrm{len}(t_n)$. So $\mathrm{len}(s_m) = \mathrm{len}(t_n)$ and thus both $s_m = t_n$ and $s_1, \ldots, s_{m-1} = t_1, \ldots, t_{n-1}$. By induction $m - 1 = n - 1$ and $s_i = t_i$ for $1 \leq i \leq n - 1$, so we are done. $\square$

The Parsing Lemma follows directly.

Proof of Lemma 2.9. Existence follows by construction, so assume

$$g(s_1, \ldots, s_m) = t = f(t_1, \ldots, t_n)$$

for $g \in F$ and m terms $s_1, \ldots, s_m \in \mathcal{L}(\mathfrak{S})$. The initial symbols of both ends of the equation must coincide, so $g = f$. The $(3, \mathrm{len}(t) - 1)$-subwords of both ends must coincide as well, so $s_1, \ldots, s_m = t_1, \ldots, t_n$. Then the result follows from Corollary 2.15. $\square$

We call the subterm t_i in a non-atomic term $w = f(t_1, \ldots, t_n)$ the i^{th} *entry of w*.

We may also use Corollary 2.14 to characterize where the subterms of a term may appear, which we will often need to do when analyzing terms.

Corollary 2.16. *Say $w = f(t_1, \ldots, t_n) \in \mathcal{L}(\mathfrak{S})$ and v is a proper subterm of w. Then for every instance of v in w, there exists $1 \leq k \leq n$ such that that instance of v is contained in the k^{th} entry t_k of w.*

Proof. Consider the (i, j)-instance of v in w. Then $i \neq 1$, or else v would be a proper initial segment of w, contradicting Corollary 2.14. Moreover, v does not start with a parenthesis or comma, so the index i must correspond to a position in one of the subterms t_k. It follows that the (i, j)-instance of v is contained entirely in t_k, for if not, then intersecting the two yields a subword that is both a proper initial segment of v and a (not necessarily proper) final segment of t_k, contradicting Corollary 2.14. $\square$

Corollary 2.17. *Given $w \in \mathcal{L}(\mathfrak{S})$, suppose v_k is the (i_k, j_k)-subterm of w for $k = 1, 2$. Suppose further that $(i_1, j_1) \neq (i_2, j_2)$. Then either v_1 and v_2 are disjoint or one is contained in the other (with endpoints excluded).*

Proof. If v_1 and v_2 are both atomic, then they are disjoint because $(i_1, j_1) \neq (i_2, j_2)$.

If v_1 is atomic and v_2 is not, then either they are disjoint or v_1 is a symbol in v_2. Since v_2 starts with a function symbol and ends with a parenthesis, v_1 is not at either end of v_2. The case where v_2 is a variable and v_1 is not is analogous.

It remains to consider when neither v_1 nor v_2 are atomic, so $i_k < j_k$. Suppose without loss of generality $(i_1, j_1) < (i_2, j_2)$ in the lexicographic order. If $j_1 < i_2$, then v_1 and v_2 are disjoint. The case $j_1 = i_2$ is impossible because v_1 ends with a right parenthesis and v_2 starts with a function symbol. So it suffices to consider $i_2 < j_1$. By the lexicographic order, we also have $i_1 \leq i_2$, and now we apply Corollary 2.14 repeatedly.

If $i_1 = i_2$, then $j_1 < j_2$ by the lexicographic order, and hence v_1 is a proper initial segment of v_2. This contradicts Corollary 2.14, so $i_1 < i_2$. Now consider the values of j_1 and j_2. If $j_1 = j_2$, then v_2 is a proper final segment of v_1, a contradiction. If $j_1 < j_2$, then $i_1 < i_2 < j_1 < j_2$ and the (i_2, j_1)-subword of w is both a proper final segment of v_1 and a proper initial segment of v_2, a contradiction. Therefore $j_2 < j_1$, so $i_1 < i_2 < j_2 < j_1$. $\square$

Often we will want to restrict the set of terms under consideration but still ensure that we may induct on their constructions. As such, we introduce the following definition.

Definition 2.18 (Subterm-closure)**.** A subset $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})$ of terms is *subterm-closed* if, given $t \in \mathfrak{O}$ and $s \in \mathcal{L}(\mathfrak{S})$ for which s is a subterm of t, we have $s \in \mathfrak{O}$ as well.

Example 2.19. Every language $\mathcal{L}(\mathfrak{S})$ is itself subterm-closed, as is $\mathcal{L}(\mathfrak{S})_V$, the set of all terms whose variables are all in V, for any set of variables V. In particular, the languages of monoidal categories and bicategories with object-set O from Examples 2.5 and 2.6 are subterm-closed, as are their variable-finite subsets $\mathcal{L}_k^{\mathrm{M}}$ and $\mathcal{L}_k^{\mathrm{B}}(P)$ from Example 2.8 for $k \in \mathbb{N}$ and finite $P \subseteq O$.

Note that every nonempty subterm-closed subset of $\mathcal{L}(\mathfrak{S})$ must necessarily contain at least one atomic term, for the positivity of term lengths implies that the subset must contain a term of minimal length, and subterm-closure implies that every term of minimal length is atomic.

Another example of a subterm-closed set of terms is the set of all *regular* terms in a language, which we define here.

Definition 2.20 (Regularity)**.** A term $t \in \mathcal{L}(\mathfrak{S})$ is *regular* if no variable appears in t more than once; that is, each variable $x_i^{\mathtt{t}}$ is a $(j_i^{\mathtt{t}}, j_i^{\mathtt{t}})$-subterm of t for at most one value of $j_i^{\mathtt{t}}$. For $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})$ we let $\mathfrak{O}^{\mathrm{reg}}$ denote the set of all regular terms in $\mathfrak{O}$.

Example 2.21. In the language $\mathcal{L}^{\mathrm{M}}$ of monoidal categories from Example 2.5, the term $\otimes(\otimes(e, x_4), \otimes(x_1, e))$ is regular (a term may have repeated constants and still be regular), but the term $\otimes(\otimes(x_2, x_1), x_2)$ is not. Often we will work with the set of all regular terms $\mathcal{L}^{\mathrm{M,reg}}$ in the language of monoidal categories, or its subset of all regular terms $\mathcal{L}_k^{\mathrm{M,reg}}$ with scope V_k.

Example 2.22. Observe that every subterm of a regular term is still regular. It follows that if $\mathfrak{O}$ is subterm-closed, then $\mathfrak{O}^{\mathrm{reg}}$ is subterm-closed as well. In particular, $\mathcal{L}(\mathfrak{S})^{\mathrm{reg}}$ is subterm-closed, and $\mathcal{L}(\mathfrak{S})_V^{\mathrm{reg}}$ is both variable-finite and subterm-closed when V is finite.

We conclude this section by examining how we may construct new terms from old by substituting terms for variables of the same type. Recall from Definition 2.7 that given a finite set of variables V, a term with scope V is a term whose variables are all in V.

Definition 2.23 (Substitution)**.** Given a term $t \in \mathcal{L}(\mathfrak{S})$ with scope $V = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}$, let $\mathbf{s} = (s_1, \ldots, s_n) \in \mathcal{L}(\mathfrak{S})^n$ be an n-tuple of terms, each of type $\tau(s_k) = \mathtt{t}_k$. We will define $t[V/\mathbf{s}]$ to be the term in $\mathcal{L}(\mathfrak{S})$ with the same type as t obtained by simultaneously replacing every instance of each variable $x_{i_k}^{\mathtt{t}_k}$ in t with s_k. We may define $t[V/\mathbf{s}]$ recursively as follows:

(1) If t is a variable, then $t = x_{i_k}^{\mathtt{t}_k}$ for some k, so define $x_{i_k}^{\mathtt{t}_k}[V/\mathbf{s}] \coloneqq s_k$. We have $\tau(t) = \tau(x_{i_k}^{\mathtt{t}_k}) = \mathtt{t}_k = \tau(s_k) = \tau(t[V/\mathbf{s}])$.

(2) If t is a constant, define $t[V/\mathbf{s}] \coloneqq t$. Again, we have $\tau(t) = \tau(t[V/\mathbf{s}])$.

(3) Otherwise, $t = f(t_1, \ldots, t_n)$ for n-ary function symbol f and terms $t_1, \ldots, t_n \in \mathcal{L}(\mathfrak{S})$ with shared scope V and $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_n))$. Then with $\tau(t_k[V/\mathbf{s}]) = \tau(t_k)$, we may define

$$f(t_1, \ldots, t_n)[V/\mathbf{s}] \coloneqq f(t_1[V/\mathbf{s}], \ldots, t_n[V/\mathbf{s}])$$

by the Parsing Lemma. We still have $\tau(t) = \mathrm{cod}(f) = \tau(t[V/\mathbf{s}])$.

It turns out that substitutions compose.

Lemma 2.24. *Given a term $t \in \mathcal{L}(\mathfrak{S})$ with scope $V = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}$, let $\mathbf{s} = (s_1, \ldots, s_n) \in \mathcal{L}(\mathfrak{S})^n$ be an n-tuple of terms, each of type $\tau(s_k) = \mathtt{t}_k$. Furthermore, let $W = \{x_{j_1}^{\mathtt{u}_1} < \cdots < x_{j_n}^{\mathtt{u}_m}\}$ be a common scope for $s_1, \ldots, s_n$, so that $t[V/\mathbf{s}]$ also has scope W, and let $\mathbf{r} = (r_1, \ldots, r_m) \in \mathcal{L}(\mathfrak{S})^m$ be an m-tuple of terms, each of type $\tau(r_k) = \mathtt{u}_k$. Then $t[V/\mathbf{s}][W/\mathbf{r}] = t[V/\mathbf{s}']$, where $\mathbf{s}' \coloneqq (s_1[W/\mathbf{r}], \ldots, s_n[W/\mathbf{r}])$.*

Proof. Induct on t. If t is a variable, then $t = x_{i_k}^{\mathtt{t}_k}$ for some k, so $t[V/\mathbf{s}] = s_k$ while $t[V/\mathbf{s}'] = s_k[W/\mathbf{r}]$, and the result follows. If t is a constant, then $t[V/\mathbf{s}][W/\mathbf{r}] = t = t[V/\mathbf{s}']$.

Otherwise, $t = f(t_1, \ldots, t_n)$ for n-ary function symbol f and terms $t_1, \ldots, t_n \in \mathcal{L}(\mathfrak{S})$ with shared scope V and $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_n))$. Then by the Parsing Lemma and induction,

$$\begin{aligned}
t[V/\mathbf{s}][W/\mathbf{r}] &= f(t_1[V/\mathbf{s}], \ldots, t_n[V/\mathbf{s}])[W/\mathbf{r}] \\
&= f(t_1[V/\mathbf{s}][W/\mathbf{r}], \ldots, t_n[V/\mathbf{s}][W/\mathbf{r}]) \\
&= f(t_1[V/\mathbf{s}][W/\mathbf{r}], \ldots, t_n[V/\mathbf{s}][W/\mathbf{r}]) \\
&= f(t_1[V/\mathbf{s}'], \ldots, t_n[V/\mathbf{s}']) \\
&= t[V/\mathbf{s}']
\end{aligned}$$

as desired. □

Example 2.25. Consider the term $t \coloneqq \otimes(x_3, x_1)$ in the language of monoidal categories $\mathcal{L}^{\mathrm{M}}$ from Example 2.5 with scope $V_3 = \{x_1 < x_2 < x_3\}$ from Example 2.8. Let

$$\mathbf{s} \coloneqq (\otimes(x_1, x_2), x_5, \otimes(x_4, e)) \in (\mathcal{L}^{\mathrm{M}})^3.$$

Then $t[V_3/\mathbf{s}]$ is the following term in $\mathcal{L}^{\mathrm{M}}$:

$$\otimes(\otimes(x_4, e), \otimes(x_1, x_2)).$$

2.2 Graphs parametrizing coherence maps

We use terms to parametrize operation functors, but we also seek to parametrize the coherence natural transformations between them. We do so by constructing a directed graph (using category-theoretic graph fundamentals from Section A.1, so in fact all our graphs will be directed) with terms representing operations as vertices and formal transformations representing coherence maps as edges,[5] beginning with a *transformation schema* for the coherence maps initially specified. Note that coherence maps should always be between operations of the same type, as we ensure in the following definition.

Definition 2.26 (Transformation schema)**.** A *transformation schema for* $\mathfrak{S}$ is a set $\mathcal{E}$ of edge labels together with a function

$$\mathfrak{T}\colon \mathcal{E} \to \mathcal{L}(\mathfrak{S}) \times_T \mathcal{L}(\mathfrak{S}) = \left\{(s,t) \in \mathcal{L}(\mathfrak{S}) \times \mathcal{L}(\mathfrak{S}) \,\middle|\, \tau(s) = \tau(t)\right\}.$$

We write $\varepsilon\colon s \Rightarrow t$ for $\mathfrak{T}(\varepsilon) = (s,t)$, as well as $\mathrm{dom}(\varepsilon) \coloneqq s$ and $\mathrm{cod}(\varepsilon) \coloneqq t$, viewing $\mathfrak{T}$ as a (directed) graph (see Definition A.1) with vertices in $\mathcal{L}(\mathfrak{S})$, edges in $\mathcal{E}$, and domains and codomains as specified. We extend the type function τ to $\mathcal{E}$ by setting $\tau(\varepsilon) \coloneqq \tau(s) = \tau(t)$.

Example 2.27 (Transformation schema for monoidal categories)**.** Continuing from Example 2.5 with the signature $\mathfrak{S}_{\mathrm{M}}$ and language $\mathcal{L}^{\mathrm{M}}$ for monoidal categories, we define a set of edge labels $\mathcal{E}_{\mathrm{SM}} \coloneqq \{\alpha, \beta, \lambda, \rho\}$ and transformation schema $\mathfrak{T}_{\mathrm{SM}}\colon \mathcal{E}_{\mathrm{SM}} \to \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$ given by

$$\begin{aligned} \alpha\colon \otimes(\otimes(x_1,x_2),x_3) &\Rightarrow \otimes(x_1,\otimes(x_2,x_3)) \qquad & \beta\colon \otimes(x_1,x_2) &\Rightarrow \otimes(x_2,x_1) \\ \lambda\colon \otimes(e,x_1) &\Rightarrow x_1 \qquad & \rho\colon \otimes(x_1,e) &\Rightarrow x_1 \end{aligned}$$

for symmetric monoidal categories. We can restrict the edge labels to $\mathcal{E}_{\mathrm{M}} \coloneqq \{\alpha, \lambda, \rho\}$ and $\mathfrak{T}_{\mathrm{SM}}$ to $\mathfrak{T}_{\mathrm{M}}\colon \mathcal{E}_{\mathrm{M}} \to \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$ to obtain a transformation schema for monoidal categories.

[5]This means our vertices will in fact represent 1-dimensional functors and our edges will represent 2-dimensional natural transformations; our notation will reflect this.

Example 2.28 (Transformation schema for bicategories)**.** We continue from Example 2.6 using the signature $\mathfrak{S}_{\mathrm{B}}(O)$ and language $\mathcal{L}^{\mathrm{B}}(O)$ for bicategories with object-set O by defining a set of edge labels

$$\mathcal{E}_{\mathrm{B}}(O) \coloneqq \{\alpha^{a,b,c,d} \mid a,b,c,d \in O\} \sqcup \{\lambda^{a,b} \mid a,b \in O\} \sqcup \{\rho^{a,b} \mid a,b \in O\}$$

and transformation schema $\mathfrak{T}_{\mathrm{B}}(O)\colon \mathcal{E}_{\mathrm{B}}(O) \to \mathcal{L}^{\mathrm{B}}(O) \times_{(O\times O)} \mathcal{L}^{\mathrm{B}}(O)$ given by

$$\alpha^{a,b,c,d}\colon \circ_{a,b,d}\left(\circ_{b,c,d}\left(x_1,x_2\right),x_3\right) \Rightarrow \circ_{a,c,d}\left(x_1,\circ_{a,b,c}\left(x_2,x_3\right)\right)$$
$$\lambda^{c,d}\colon \circ_{c,d,d}\left(\mathbf{1}_d,x_1\right) \Rightarrow x_1 \qquad \rho^{c,d}\colon \circ_{c,c,d}\left(x_1,\mathbf{1}_c\right) \Rightarrow x_1$$

for bicategories with object-set O, with

$$x_1 \coloneqq x_1^{(c,d)} \quad , \quad x_2 \coloneqq x_2^{(b,c)} \quad , \quad \text{and} \quad x_3 \coloneqq x_3^{(a,b)}.$$

We may suppress elements of O in the superscripts and subscripts for readability when they are clear from context.

For the remainder of this chapter, we fix a set of edge labels $\mathcal{E}$ and a transformation schema $\mathfrak{T}\colon \mathcal{E} \to \mathcal{L}(\mathfrak{S}) \times_T \mathcal{L}(\mathfrak{S})$ for our signature $\mathfrak{S}$. We now expand the edge-set $\mathcal{E}$ of our directed graph to include arbitrary terms substituted into the variables (as in Definition 2.23) appearing in the endpoints of the edges in our transformation schema.

Definition 2.29 (Atomic transformation)**.** We construct a directed graph $\mathscr{A}t(\mathfrak{S},\mathfrak{T})$ with vertex-set $\mathcal{L}(\mathfrak{S})$ as follows. Let $\varepsilon \in \mathcal{E}$ be an edge label $\varepsilon\colon \mathrm{dom}(\varepsilon) \Rightarrow \mathrm{cod}(\varepsilon)$, and let $V(\varepsilon) = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_k}^{\mathtt{t}_k}\}$ be the set of variables $V(\mathrm{dom}(\varepsilon)) \cup V(\mathrm{cod}(\varepsilon))$ appearing in $\mathrm{dom}(\varepsilon)$ or $\mathrm{cod}(\varepsilon)$. If $s_1,\ldots,s_k \in \mathcal{L}(\mathfrak{S})$ are terms with $\tau(s_j) = \mathtt{t}_j$ for all $1 \le j \le k$, then the $(s_1,\ldots,s_k)$-*instance of* ε is the tuple $(\varepsilon; s_1,\ldots,s_k)$. We shall denote the tuple by $\varepsilon_{\mathbf{s}} = \varepsilon_{s_1,\ldots,s_k}$ and regard it as an edge

$$\varepsilon_{\mathbf{s}}\colon \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}] \Rightarrow \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$$

as in Definition 2.23, so that s_j has been substituted for every instance of $x_{i_j}^{\mathtt{t}_j}$ in both $\mathrm{dom}(\varepsilon)$ and $\mathrm{cod}(\varepsilon)$ for $1 \le j \le k$. We shall call the instances $\varepsilon_{\mathbf{s}}$ of edge labels $\varepsilon \in \mathcal{E}$ *atomic* *transformations*,[6] and we let $\mathrm{At}(\mathfrak{S}, \mathfrak{T})$ denote the set of atomic transformations. Then the directed graph $\mathscr{A}t(\mathfrak{S}, \mathfrak{T})$ has vertex-set $\mathcal{L}(\mathfrak{S})$, edge-set $\mathrm{At}(\mathfrak{S}, \mathfrak{T})$, and

$$\mathrm{dom}(\varepsilon_{\mathbf{s}}) \coloneqq \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}] \quad \text{and} \quad \mathrm{cod}(\varepsilon_{\mathbf{s}}) \coloneqq \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$$

for all $\varepsilon_{\mathbf{s}} \in \mathrm{At}(\mathfrak{S}, \mathfrak{T})$. We set $\tau(\varepsilon_{\mathbf{s}}) \coloneqq \tau(\varepsilon)$ and $V(\varepsilon_{\mathbf{s}}) \coloneqq V(\mathrm{dom}(\varepsilon_{\mathbf{s}})) \cup V(\mathrm{cod}(\varepsilon_{\mathbf{s}}))$, so that $V(\varepsilon_{\mathbf{s}})$ is the set of variables appearing in either the domain or the codomain of $\varepsilon_{\mathbf{s}}$. A *scope for* $\varepsilon_{\mathbf{s}}$ is a finite set of variables that is a scope for both $\mathrm{dom}(\varepsilon_{\mathbf{s}})$ and $\mathrm{cod}(\varepsilon_{\mathbf{s}})$.

For each $\varepsilon \in \mathcal{E}$, we can take the variables in $V(\varepsilon) = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_k}^{\mathtt{t}_k}\}$ and trivially "substitute" each one for itself: setting $\mathbf{x} \coloneqq (x_{i_1}^{\mathtt{t}_1}, \ldots, x_{i_k}^{\mathtt{t}_k})$, we have $\mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{x}] = \mathrm{dom}(\varepsilon)$ and $\mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{x}] = \mathrm{cod}(\varepsilon)$, so we may identify $\varepsilon_{\mathbf{x}}\colon \mathrm{dom}(\varepsilon) \Rightarrow \mathrm{cod}(\varepsilon)$ in $\mathrm{At}(\mathfrak{S}, \mathfrak{T})$ with $\varepsilon\colon \mathrm{dom}(\varepsilon) \Rightarrow \mathrm{cod}(\varepsilon)$ in $\mathcal{E}$, yielding a canonical inclusion $\mathcal{E} \hookrightarrow \mathrm{At}(\mathfrak{S}, \mathfrak{T})$.

More generally, recall that the domain and codomain of an edge label $\varepsilon \in \mathcal{E}$ share a type, and substitution does not change type, so the domain and codomain of an atomic transformation $\varepsilon_{\mathbf{s}}$ share a type as well, namely $\tau(\varepsilon)$. So we have extended our type function τ to atomic transformations in such a way that the type of an atomic transformation matches the type of its domain and codomain.

Example 2.30 (Atomic transformations for monoidal categories)**.** We continue from Example 2.27. The atomic transformations in $\mathrm{At}^{\mathrm{SM}} \coloneqq \mathrm{At}(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{SM}})$ are the edges

$$\alpha_{r,s,t}\colon \otimes(\otimes(r,s),t) \Rightarrow \otimes(r,\otimes(s,t)) \qquad \beta_{s,t}\colon \otimes(s,t) \Rightarrow \otimes(t,s)$$
$$\lambda_t\colon \otimes(e,t) \Rightarrow t \qquad \rho_t\colon \otimes(t,e) \Rightarrow t$$

for terms $r, s, t \in \mathcal{L}^{\mathrm{M}}$. These edges form the graph $\mathscr{A}t^{\mathrm{SM}}$ with terms in $\mathcal{L}^{\mathrm{M}}$ as vertices. The atomic transformations in $\mathrm{At}^{\mathrm{M}} \coloneqq \mathrm{At}(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}})$ are those in $\mathrm{At}^{\mathrm{SM}}$ not of the form $\beta_{s,t}$, comprising the edges of a subgraph $\mathscr{A}t^{\mathrm{M}}$ of $\mathscr{A}t^{\mathrm{SM}}$ that still has all the same vertices. We also let $\mathscr{A}t_k^{\mathrm{SM}}$ (resp. $\mathscr{A}t_k^{\mathrm{M}}$) be the full subgraph of $\mathscr{A}t^{\mathrm{SM}}$ (resp. $\mathscr{A}t^{\mathrm{M}}$) whose vertices are in $\mathcal{L}_k^{\mathrm{M}}$ from Example 2.8, with edge-set $\mathrm{At}_k^{\mathrm{SM}}$ (resp. $\mathrm{At}_k^{\mathrm{M}}$).

[6] Note that the *atomic transformations* for a transformation schema $(\mathfrak{S}, \mathfrak{T})$ are not the same as the *atomic terms* for the signature $\mathfrak{S}$, although we use the same adjective for both. In fact, in the directed graph that we define shortly, atomic terms are found among the vertices, while atomic transformations comprise the edges.

Example 2.31 (Atomic transformations for bicategories)**.** Continuing from Example 2.28, the atomic transformations in $\mathrm{At}^{\mathrm{B}}(O) \coloneqq \mathrm{At}(\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O))$ consist of the following edges:

$$\alpha_{r,s,t} \colon \circ(\circ(r,s),t) \Rightarrow \circ(r, \circ(s,t))$$
$$\lambda_t \colon \circ(\mathbf{1}_b, t) \Rightarrow t \qquad \rho_t \colon \circ(t, \mathbf{1}_a) \Rightarrow t$$

for $a, b, c, d \in O$ and terms $r, s, t \in \mathcal{L}^{\mathrm{B}}(O)$, with subscripts suppressed for clarity, such that

$$\tau(r) = (c,d), \quad \tau(s) = (b,c), \quad \text{and} \quad \tau(t) = (a,b).$$

These edges form the graph $\mathscr{A}t^{\mathrm{B}}(O)$ with terms in $\mathcal{L}^{\mathrm{B}}(O)$ as vertices. Let $\mathscr{A}t^{\mathrm{B}}_k(P)$ be the full subgraph of $\mathscr{A}t^{\mathrm{B}}(O)$ with vertex-set $\mathcal{L}^{\mathrm{B}}_k(P)$ from Example 2.8, with edge-set $\mathrm{At}^{\mathrm{B}}_k(P)$.

Next, we introduce formal identity transformations and apply function symbols alongside these to each atomic transformation to obtain what we call *basic transformations*, following Mac Lane in [16, p. 166]. Each basic transformation φ will contain exactly one atomic transformation $\varepsilon_{\mathbf{s}}$; the domain and codomain of $\varepsilon_{\mathbf{s}}$ will be subterms of the domain and codomain of φ. While φ may have a larger domain than $\varepsilon_{\mathbf{s}}$, only *identity transformations* will "act" on any portion of $\mathrm{dom}(\varphi)$ outside of its subterm $\mathrm{dom}(\varepsilon_{\mathbf{s}})$, so that $\mathrm{cod}(\varphi)$ is obtained by replacing the subterm $\mathrm{dom}(\varepsilon_{\mathbf{s}})$ of $\mathrm{dom}(\varphi)$ with $\mathrm{cod}(\varepsilon_{\mathbf{s}})$ and changing nothing else. We formalize this as follows, building on the constructions from Definition 2.29.

Definition 2.32 (Identity transformation, basic transformation)**.** Define a signature with the same sets of types T and function symbols F as $\mathfrak{S}$ but with new constant symbols

$$\{\mathrm{id}_t \mid t \in \mathcal{L}(\mathfrak{S})\} \sqcup \mathrm{At}(\mathfrak{S}, \mathfrak{T})$$

with $\tau(\mathrm{id}_t) \coloneqq \tau(t)$ and $\tau(\varepsilon_{\mathbf{s}}) \coloneqq \tau(\varepsilon)$ for $\varepsilon_{\mathbf{s}} \in \mathrm{At}(\mathfrak{S}, \mathfrak{T})$ (consistent with the definition of τ on $\mathrm{At}(\mathfrak{S}, \mathfrak{T})$ from Definition 2.29).[7] So for $t \in \mathcal{L}(\mathfrak{S})$, there is a term id_t of this new signature of type $\tau(t)$ called the *identity transformation for t*.

We seek to define a set of terms $\mathrm{Bs} \coloneqq \mathrm{Bs}(\mathfrak{S}, \mathfrak{T})$ called *basic transformations* that we regard as edges via functions $\mathrm{dom}, \mathrm{cod} \colon \mathrm{Bs}(\mathfrak{S}, \mathfrak{T}) \rightrightarrows \mathcal{L}(\mathfrak{S})$ to form a directed graph whose

[7] In this new signature, distinct from $\mathfrak{S}$, the atomic transformations are constants and are thus atomic terms for the new signature, justifying their name. But to avoid confusion, henceforth when we refer to constant and atomic terms, we will always mean the constant and atomic terms *for the signature* $\mathfrak{S}$, distinct from the id_t's and the atomic transformations that are constants for this new signature.

vertices are terms and whose edges are the basic transformations between them. Although we will *not* include identity transformations among our basic transformations or in our directed graph, it may aid intuition to think of an identity transformation id_t as an edge $\mathrm{id}_t \colon t \Rightarrow t$. Below, we recursively define our basic transformations, along with their types, domains, and codomains.

(1) Every atomic transformation $\varepsilon_{\mathbf{s}} \in \mathrm{At}(\mathfrak{S}, \mathfrak{T})$ is also a basic transformation, with $\mathrm{dom}(\varepsilon_{\mathbf{s}}) \coloneqq \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ and $\mathrm{cod}(\varepsilon_{\mathbf{s}}) \coloneqq \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ as in Definition 2.29. So $\mathrm{At}(\mathfrak{S}, \mathfrak{T}) \subseteq \mathrm{Bs}(\mathfrak{S}, \mathfrak{T})$. Again $\tau(\varepsilon_{\mathbf{s}}) = \tau(\varepsilon)$, so $\varepsilon_{\mathbf{s}}$ shares a type with its domain and codomain.

(2) Given a function symbol $f \in F$ with arity n, a basic transformation $\varphi \in \mathrm{Bs}(\mathfrak{S}, \mathfrak{T})$, and terms $t_1, \ldots, t_{k-1}, t_{k+1}, \ldots, t_n \in \mathcal{L}(\mathfrak{S})$ for $1 \leq k \leq n$ with

$$\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_{k-1}), \tau(\varphi), \tau(t_{k+1}), \ldots, \tau(t_n)),$$

we can construct the term

$$\varphi' \coloneqq f(\mathrm{id}_{t_1}, \ldots, \mathrm{id}_{t_{k-1}}, \varphi, \mathrm{id}_{t_{k+1}}, \ldots, \mathrm{id}_{t_n})$$

of type $\mathrm{cod}(f)$ and let it be a basic transformation as well. We set

$$\begin{aligned} \mathrm{dom}(\varphi') &\coloneqq f(t_1, \ldots, t_{k-1}, \mathrm{dom}(\varphi), t_{k+1}, \ldots, t_n) \\ \text{and} \quad \mathrm{cod}(\varphi') &\coloneqq f(t_1, \ldots, t_{k-1}, \mathrm{cod}(\varphi), t_{k+1}, \ldots, t_n) \end{aligned}$$

as these are well-defined by the Parsing Lemma, so that the basic transformation φ' also shares a type with its domain and codomain.

We let $\mathscr{B} \coloneqq \mathscr{B}(\mathfrak{S}, \mathfrak{T})$ be the directed graph given by $\mathrm{dom}, \mathrm{cod} \colon \mathrm{Bs} \rightrightarrows \mathcal{L}(\mathfrak{S})$ as defined above, where each edge shares a type with its domain and codomain.

For a subset of terms $\mathfrak{D} \subseteq \mathcal{L}(\mathfrak{S})$, we write $\mathrm{Bs}_{\mathfrak{D}} \subseteq \mathrm{Bs}$ for the subset of basic transformations whose domain and codomain are both in $\mathfrak{D}$, and we write $\mathscr{B}_{\mathfrak{D}}$ for the full subgraph of $\mathscr{B}$ spanned by $\mathfrak{D}$: its vertices are given by $\mathfrak{D}$ and its edges are given by $\mathrm{Bs}_{\mathfrak{D}}$. For a set of variables V for $\mathfrak{S}$, we write $\mathrm{Bs}_V \coloneqq \mathrm{Bs}_{\mathcal{L}(\mathfrak{S})_V}$ for the subset of basic transformations whose domain and codomain have scopes contained in V, and we write $\mathscr{B}_V \coloneqq \mathscr{B}_{\mathcal{L}(\mathfrak{S})_V}$ for the full subgraph of $\mathscr{B}$ spanned by $\mathcal{L}(\mathfrak{S})_V$: its vertices are given by $\mathcal{L}(\mathfrak{S})_V$ and its edges are given by Bs_V. For finite V, we say that V is a *scope for* φ if $\varphi \in \mathrm{Bs}_V$. We write $V(\varphi) \coloneqq V(\mathrm{dom}(\varphi)) \cup V(\mathrm{cod}(\varphi))$ so that for $\varphi' = f(\mathrm{id}_{t_1}, \ldots, \mathrm{id}_{t_{k-1}}, \varphi, \mathrm{id}_{t_{k+1}}, \ldots, \mathrm{id}_{t_n})$,

$$V(\varphi') = V(t_1) \cup \cdots \cup V(t_{k-1}) \cup V(\varphi) \cup V(t_{k+1}) \cup \cdots \cup V(t_n).$$

Example 2.33 (Basic transformations for monoidal categories)**.** We continue from Example 2.30. The basic transformations in $\mathrm{Bs}^{\mathrm{SM}} \coloneqq \mathrm{Bs}(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{SM}})$ include atomic transformations

$$\alpha_{r,s,t} \colon \otimes(\otimes(r,s),t) \Rightarrow \otimes(r, \otimes(s,t))$$

as well as $\beta_{u,v}, \lambda_u$, and ρ_u for $r,s,t,u,v \in \mathcal{L}^{\mathrm{M}}$, as well as longer terms like

$$\begin{aligned} &\otimes(\mathrm{id}_u, \alpha_{r,s,t}) \colon \otimes(u, \otimes(\otimes(r,s),t)) \Rightarrow \otimes(u, \otimes(r, \otimes(s,t))) \\ \text{and} \quad &\otimes(\otimes(\mathrm{id}_u, \alpha_{r,s,t}), \mathrm{id}_v) \colon \otimes(\otimes(u, \otimes(\otimes(r,s),t)), v) \Rightarrow \otimes(\otimes(u, \otimes(r, \otimes(s,t))), v). \end{aligned}$$

These are the edges of the graph $\mathscr{B}^{\mathrm{SM}}$ with terms in $\mathcal{L}^{\mathrm{M}}$ as vertices. If we delete every edge containing a term of the form $\beta_{u,v}$ from $\mathscr{B}^{\mathrm{SM}}$, we obtain a graph $\mathscr{B}^{\mathrm{M}}$, still with terms in $\mathcal{L}^{\mathrm{M}}$ vertices and with basic transformations in $\mathrm{Bs}^{\mathrm{M}} \coloneqq \mathrm{Bs}(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}})$ as edges. We also let $\mathscr{B}^{\mathrm{SM}}_k$ (resp. $\mathscr{B}^{\mathrm{M}}_k$) be the full subgraph of $\mathscr{B}^{\mathrm{SM}}$ (resp. $\mathscr{B}^{\mathrm{M}}$) whose vertices are in $\mathcal{L}^{\mathrm{M}}_k$, with edge-set $\mathrm{Bs}^{\mathrm{SM}}_k$ (resp. $\mathrm{Bs}^{\mathrm{M}}_k$).

Example 2.34 (Basic transformations for bicategories)**.** We continue from Example 2.31, suppressing subscripts for readability. Let $a,b,c \in O$ and fix terms s of type (b,c) and t of type (a,b). The basic transformations in $\mathrm{Bs}^{\mathrm{B}}(O) \coloneqq \mathrm{Bs}(\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O))$ include the atomic transformations $\varepsilon_{\mathbf{s}}$ for $\varepsilon_{\mathbf{s}} \in \mathrm{At}^{\mathrm{B}}(O)$, such as

$$\lambda \colon \circ(e,t) \Rightarrow t$$

of type (a,b), as well as longer terms like

$$\begin{aligned} &\circ(\mathrm{id}_s, \lambda) \colon \circ(s, \circ(e,t)) \Rightarrow \circ(s,t) \\ \text{and} \quad &\circ(\circ(\mathrm{id}_s, \lambda), \mathrm{id}_e) \colon \circ(\circ(s, \circ(e,t)), e) \Rightarrow \circ(\circ(s,t), e) \end{aligned}$$

both of type (a, c). These are the edges of the graph $\mathscr{B}^{\mathrm{B}}(O)$ with terms in $\mathcal{L}^{\mathrm{B}}(O)$ as vertices. We also let $\mathscr{B}^{\mathrm{B}}_k(P)$ be the full subgraph of $\mathscr{B}^{\mathrm{B}}(O)$ whose vertices are in $\mathcal{L}^{\mathrm{B}}_k(P)$, with edge-set $\mathrm{Bs}^{\mathrm{B}}_k(P)$; for instance, the preceding basic transformation is in $\mathrm{Bs}^{\mathrm{B}}_1(\{a, b, c\})$.

We define a number of useful auxiliary functions on basic transformations. The Parsing Lemma ensures that these constructions, which depend on the entries of each basic transformation, are all well-defined.

Definition 2.35. Fix $\varepsilon_{\mathbf{s}} \in \mathrm{At}(\mathfrak{S}, \mathfrak{T})$, $f \in F$, $\varphi \in \mathrm{Bs}(\mathfrak{S}, \mathfrak{T})$, and $t_1, \ldots, t_{k-1}, t_{k+1}, \ldots, t_n \in \mathcal{L}(\mathfrak{S})$ for $1 \leq k \leq n$ with $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_{k-1}), \tau(\varphi), \tau(t_{k+1}), \ldots, \tau(t_n))$, so that we have a basic transformation

$$\varphi' := f(\mathrm{id}_{t_1}, \ldots, \mathrm{id}_{t_{k-1}}, \varphi, \mathrm{id}_{t_{k+1}}, \ldots, \mathrm{id}_{t_n}).$$

(1) The *atom* of a basic transformation is the atomic transformation given by the function $A\colon \mathrm{Bs}(\mathfrak{S}, \mathfrak{T}) \to \mathrm{At}(\mathfrak{S}, \mathfrak{T})$ defined recursively by

$$A(\varepsilon_{\mathbf{s}}) := \varepsilon_{\mathbf{s}} \quad \text{and} \quad A(\varphi') := A(\varphi).$$

It is the unique atomic transformation appearing in the basic transformation. By induction, $V(A(\varphi)) \subseteq V(\varphi)$, so a scope for φ is always a scope for $A(\varphi)$.

(2) The *label* of a basic transformation is the edge label given by $L\colon \mathrm{Bs}(\mathfrak{S}, \mathfrak{T}) \to \mathcal{E}$ defined recursively by

$$L(\varepsilon_{\mathbf{s}}) := \varepsilon \quad \text{and} \quad L(\varphi') := L(\varphi).$$

Note that the label function factors through A, and in fact $L = L \circ A$. We say that φ is an $L(\varphi)$-*basic transformation.*

(3) The *span* of a basic transformation φ is an ordered pair of positive integers $S(\varphi) := (S_1(\varphi), S_2(\varphi))$ given recursively by

$$\begin{aligned} S(\varepsilon_{\mathbf{s}}) &:= (1, \mathrm{len}(\mathrm{dom}(\varepsilon_{\mathbf{s}}))) \\ \text{and} \quad S(\varphi') &:= S(\varphi) + \mathrm{len}(t_1) + \cdots + \mathrm{len}(t_{k-1}) + k + 1, \end{aligned}$$

where $(a, b) + c \coloneqq (a + c, b + c)$. Despite the notation, we typically think of the span as a closed interval $[S_1(\varphi), S_2(\varphi)]$.

The span is defined so that the domain of the atom of φ is the $S(\varphi)$-subterm of the domain of φ, which we can show inductively: it is true for the base case of $\varepsilon_{\mathbf{s}}$, and it is true by induction for $\varphi' = f(\mathrm{id}_{t_1}, \dots, \mathrm{id}_{t_{k-1}}, \varphi, \mathrm{id}_{t_{k+1}}, \dots, \mathrm{id}_{t_n})$ because the domain $f(t_1, \dots, t_{k-1}, \mathrm{dom}(\varphi), t_{k+1}, \dots, t_n)$ of φ' has $\mathrm{len}(t_1) + \cdots + \mathrm{len}(t_{k-1}) + k + 1$ total symbols appearing before $\mathrm{dom}(\varphi)$ does. We can further deduce the identity

$$S_2(\varphi) - S_1(\varphi) = \mathrm{len}(\mathrm{dom}(A(\varphi))) - 1.$$

(4) The *difference* of φ is $\Delta(\varphi) \coloneqq \mathrm{len}(\mathrm{cod}(\varphi)) - \mathrm{len}(\mathrm{dom}(\varphi))$. By induction, the difference function factors through the atom function: $\Delta = \Delta \circ A$. Depending on the transformation schema, Δ might further factor through L (see the following example). Also note that the codomain of the atom of φ is the $(S_1(\varphi), S_2(\varphi) + \Delta(\varphi))$-subterm of the codomain of φ.

Example 2.36. Consider the basic transformations in $\mathrm{Bs}^{\mathrm{SM}}$, the edges of $\mathscr{B}^{\mathrm{SM}}$, from Example 2.33. The atom function A lands in $\mathrm{At}^{\mathrm{SM}}$, and we have for instance

$$\begin{aligned} &A(\otimes(\mathrm{id}_{\otimes(x_2,x_3)}, \alpha_{x_1,x_3,\otimes(x_4,x_2)})) = \alpha_{x_1,x_3,\otimes(x_4,x_2)} \\ \text{and} \quad &A(\otimes(\otimes(\lambda_{x_3}, \mathrm{id}_{x_3}), \mathrm{id}_{x_1})) = \lambda_{x_3}, \end{aligned}$$

while the closely related label function L lands in $\mathcal{E}_{\mathrm{SM}} = \{\alpha, \beta, \lambda, \rho\}$, with

$$\begin{aligned} &L(\otimes(\mathrm{id}_{\otimes(x_2,x_3)}, \alpha_{x_1,x_3,\otimes(x_4,x_2)})) = \alpha \\ \text{and} \quad &L(\otimes(\otimes(\lambda_{x_3}, \mathrm{id}_{x_3}), \mathrm{id}_{x_1})) = \lambda. \end{aligned}$$

Meanwhile the span function S satisfies

$$\begin{aligned} S(\otimes(\mathrm{id}_{\otimes(x_2,x_3)}, \alpha_{x_1,x_3,\otimes(x_4,x_2)})) &= S(\alpha_{x_1,x_3,\otimes(x_4,x_2)}) + \mathrm{len}(\otimes(x_2, x_3)) + 3 \\ &= (1, \mathrm{len}(\otimes(\otimes(x_1, x_3), \otimes(x_4, x_2)))) + 9 \\ &= (10, 25), \end{aligned}$$

with the $(10, 25)$-subterm of

$$\mathrm{dom}(\otimes(\mathrm{id}_{\otimes(x_2,x_3)}, \alpha_{x_1,x_3,\otimes(x_4,x_2)})) = \otimes(\otimes(x_2, x_3), \otimes(\otimes(x_1, x_3), \otimes(x_4, x_2)))$$

given by $\otimes(\otimes(x_1, x_3), \otimes(x_4, x_2)) = \mathrm{dom}(\alpha_{x_1,x_3,\otimes(x_4,x_2)})$. We also have

$$\begin{aligned} S(\otimes(\otimes(\lambda_{x_3}, \mathrm{id}_{x_3}), \mathrm{id}_{x_1})) &= S(\otimes(\lambda_{x_3}, \mathrm{id}_{x_3})) + 2 \\ &= S(\lambda_{x_3}) + 4 \\ &= (1, \mathrm{len}(\otimes(e, x_3))) + 4 \\ &= (5, 10), \end{aligned}$$

with the $(5, 10)$-subterm of

$$\mathrm{dom}(\otimes(\otimes(\lambda_{x_3}, \mathrm{id}_{x_3}), \mathrm{id}_{x_1})) = \otimes(\otimes(\otimes(e, x_3), x_3), x_1)$$

given by $\otimes(e, x_3) = \mathrm{dom}(\lambda_{x_3})$.

Finally, the difference function always factors through A, and in this case can further be factored through L: for $r, s, t \in \mathcal{L}^{\mathrm{M}}$, we have

$$\begin{aligned} \Delta(\alpha_{r,s,t}) &= \mathrm{len}(\otimes(r, \otimes(s, t))) - \mathrm{len}(\otimes(\otimes(r, s), t)) = 0, \\ \Delta(\beta_{s,t}) &= \mathrm{len}(\otimes(t, s)) - \mathrm{len}(\otimes(s, t)) = 0, \\ \Delta(\lambda_{s,t}) &= \mathrm{len}(t) - \mathrm{len}(\otimes(e, t)) = -5, \\ \text{and} \quad \Delta(\rho_{s,t}) &= \mathrm{len}(t) - \mathrm{len}(\otimes(t, e)) = -5, \end{aligned}$$

which we can generalize to deduce that all α- and β-basic transformations have difference 0 and all λ- and ρ-basic transformations have difference -5.

We can invoke the Parsing Lemma to uniquely determine a basic transformation given its atom, domain, and span.

Lemma 2.37. *Given an atomic transformation $\varepsilon_{\mathbf{s}} \in \mathrm{At}(\mathfrak{S}, \mathfrak{T})$ and a term $t \in \mathcal{L}(\mathfrak{S})$ whose (i, j)-subterm is the domain of $\varepsilon_{\mathbf{s}}$, there is a unique basic transformation φ with domain t, atom $\varepsilon_{\mathbf{s}}$, and span (i, j). Moreover φ has scope $V(\varphi) = V(t) \cup V(\mathrm{cod}(\varepsilon_{\mathbf{s}}))$.*

Proof. We begin with the case where $i = 1$ and $j = \mathrm{len}(t)$, making $t = \mathrm{dom}(\varepsilon_{\mathbf{s}})$. Then $\varepsilon_{\mathbf{s}}$ is a basic transformation with domain t, atom $\varepsilon_{\mathbf{s}}$, and span $(1, \mathrm{len}(t))$, with scope $V(\varepsilon_{\mathbf{s}}) = V(\mathrm{dom}(\varepsilon_{\mathbf{s}})) \cup V(\mathrm{cod}(\varepsilon_{\mathbf{s}}))$. If φ is another basic transformation with span $(1, \mathrm{len}(t))$, then φ is necessarily an atomic transformation, so $A(\varphi) = \varepsilon_{\mathbf{s}}$ implies $\varphi = \varepsilon_{\mathbf{s}}$, proving uniqueness.

We proceed by induction on t. If t is an atomic term, then $i = 1 = \mathrm{len}(t) = j$, and we are in the case above. Now given a term $t = f(t_1, \ldots, t_n)$ with n-ary function symbol f and terms $t_1, \ldots, t_n$ with $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_n))$, assume $\mathrm{dom}(\varepsilon_{\mathbf{s}})$ is its (i, j)-subterm. Again, we are done if $(i, j) = (1, \mathrm{len}(t))$. Otherwise, by Corollary 2.16, the (i, j)-subterm $\mathrm{dom}(\varepsilon_{\mathbf{s}})$ of t is contained entirely in some t_k. More precisely, if the k^{th} entry t_k is the (i_k, j_k)-subterm of t, then $3 \leq i_k \leq i \leq j \leq j_k$, with $\mathrm{dom}(\varepsilon_{\mathbf{s}})$ the $(i - i_k + 1, j - i_k + 1)$-subterm of t_k. By induction, there is a unique basic transformation φ with domain t_k, atom $\varepsilon_{\mathbf{s}}$, and span $(i - i_k + 1, j - i_k + 1)$, with scope $V(\varphi) = V(t_k) \cup V(\mathrm{cod}(\varepsilon_{\mathbf{s}}))$. Then the basic transformation

$$\psi = f(\mathrm{id}_{t_1}, \ldots, \mathrm{id}_{t_{k-1}}, \varphi, \mathrm{id}_{t_{k+1}}, \ldots, \mathrm{id}_{t_n})$$

has domain t, atom $\varepsilon_{\mathbf{s}}$, and span (i, j), with scope

$$\begin{aligned} V(\psi) &= V(t_1) \cup \cdots \cup V(t_{k-1}) \cup V(\varphi) \cup V(t_{k+1}) \cup \cdots \cup V(t_n) \\ &= V(t_1) \cup \cdots \cup V(t_{k-1}) \cup V(t_k) \cup V(\mathrm{cod}(\varepsilon_{\mathbf{s}})) \cup V(t_{k+1}) \cup \cdots \cup V(t_n) \\ &= V(t) \cup V(\mathrm{cod}(\varepsilon_{\mathbf{s}})). \end{aligned}$$

To prove uniqueness, let ψ' be another basic transformation with domain t, atom $\varepsilon_{\mathbf{s}}$, and span (i, j). As $i > 1$, ψ' is not atomic, so

$$\psi' = g(\mathrm{id}_{s_1}, \ldots, \mathrm{id}_{s_{\ell-1}}, \varphi', \mathrm{id}_{s_{\ell+1}}, \ldots, \mathrm{id}_{s_m})$$

for m-ary function symbol g, terms $s_1, \ldots, s_{\ell-1}, s_{\ell+1}, \ldots, s_m$, and basic transformation φ' with $\mathrm{dom}(g) = (\tau(s_1), \ldots, \tau(s_{\ell-1}), \tau(\varphi'), \tau(s_{\ell+1}), \ldots, \tau(s_m))$, such that $A(\varphi') = \varepsilon_{\mathbf{s}}$ and the (i, j)-subterm of $\mathrm{dom}(\psi)$ appears in the ℓ^{th} entry. Then $t = \mathrm{dom}(\psi')$ implies

$$f(t_1, \ldots, t_n) = g(s_1, \ldots, s_{\ell-1}, \mathrm{dom}(\varphi'), s_{\ell+1}, \ldots, s_m);$$

so by the Parsing Lemma $n = m$, $f = g$, and $t_p = s_p$ for $1 \leq p \leq n$ if we let $s_\ell \coloneqq \mathrm{dom}(\varphi')$. The fact that ψ' has span (i, j) with the (i, j)-subterm of the left hand side appearing in the

k^{th} entry forces $\mathrm{dom}(\varphi')$ to also be the k^{th} entry on the right hand side, so $k = \ell$. It follows that φ' is a basic transformation with domain $s_\ell = t_k$, atom $\varepsilon_{\mathbf{s}}$, and span $(i-i_k+1, j-i_k+1)$, so by uniqueness from the inductive hypothesis, $\varphi = \varphi'$ and thus $\psi = \psi'$. $\square$

Just as we may substitute into terms, we shall also perform substitutions into basic transformations.

Definition 2.38 (Substitution for basic transformations)**.** Let $V = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_m}^{\mathtt{t}_m}\}$ be a finite set of variables for $\mathfrak{S}$ and let $\varphi \in \mathrm{Bs}_V$ be a basic transformation with scope V, so that both the domain and codomain of φ have scope V. Let $\mathbf{r} = (r_1, \ldots, r_m) \in \mathcal{L}(\mathfrak{S})^m$ with $\tau(r_k) = \mathtt{t}_k$. Then $\varphi[V/\mathbf{r}]$ denotes the basic transformation in $\mathrm{Bs}(\mathfrak{S}, \mathfrak{T})$ obtained by simultaneously replacing every instance of each variable $x_{i_k}^{\mathtt{t}_k}$ in φ with r_k. More precisely, define substitution recursively by

$$\varepsilon_{s_1,\ldots,s_n}[V/\mathbf{r}] \coloneqq \varepsilon_{s_1[V/\mathbf{r}],\ldots,s_n[V/\mathbf{r}]}$$
$$f(\mathrm{id}_{t_1}, \ldots, \varphi, \ldots, \mathrm{id}_{t_p})[V/\mathbf{r}] \coloneqq f(\mathrm{id}_{t_1[V/\mathbf{r}]}, \ldots, \varphi[V/\mathbf{r}], \ldots, \mathrm{id}_{t_p[V/\mathbf{r}]})$$

By inducting on the definition above and on our definition of term substitution in Definition 2.23, and using the fact that substitutions compose from Lemma 2.24, we find that $\mathrm{dom}(\varphi[V/\mathbf{r}]) = \mathrm{dom}(\varphi)[V/\mathbf{r}]$ and $\mathrm{cod}(\varphi[V/\mathbf{r}]) = \mathrm{cod}(\varphi)[V/\mathbf{r}]$. If $r_1, \ldots, r_m$ have common scope W (i.e. $r_1, \ldots, r_m \in \mathcal{L}(\mathfrak{S})_W$), then $\varphi[V/\mathbf{r}]$ has scope W as well (i.e. $\varphi[V/\mathbf{r}] \in \mathrm{Bs}_W$).

We may wish to designate a subset of our basic transformations as invertible by *marking* them in our graph and formally inverting them, as in the following definitions.

Definition 2.39 (Marked graph)**.** A *marked (directed) graph* is a pair $(\mathscr{G}, \mathscr{I})$ consisting of a (directed) graph $\mathscr{G} = [E \rightrightarrows V]$ and a subset $I \subseteq E$ of *marked edges*. Then the directed graph $\mathscr{G}[I^{-1}]$ has the same vertex-set V as $\mathscr{G}$ and edge-set $E \sqcup I^{-1} = E \sqcup \{i^{-1} \mid i \in I\}$, with domains and codomains on I^{-1} given by $\mathrm{dom}(i^{-1}) \coloneqq \mathrm{cod}(i)$ and $\mathrm{cod}(i^{-1}) \coloneqq \mathrm{dom}(i)$.

Note that an isomorphism $\varphi\colon (\mathscr{G}_1, I_1) \to (\mathscr{G}_2, I_2)$ of marked directed graphs (i.e. a graph isomorphism $\varphi\colon \mathscr{G}_1 \to \mathscr{G}_2$ as in Definition A.2 that restricts to a bijection $I_1 \to I_2$ on marked edges) extends to a graph isomorphism $\varphi\colon \mathscr{G}_1[I_1^{-1}] \to \mathscr{G}_2[I_2^{-1}]$ that sends i^{-1} to $\varphi(i)^{-1}$ for each $i \in I$. Furthermore, given a graph morphism $\psi\colon \mathscr{G} \to \mathscr{C}$ (see Definition A.2) from a marked directed graph $(\mathscr{G}, I)$ to the underlying graph of a category $\mathscr{C}$ (see Definition A.3) that sends every marked edge in I to an isomorphism in $\mathscr{C}$, we can extend ψ to a graph morphism $\psi\colon \mathscr{G}[I^{-1}] \to \mathscr{C}$ that sends i^{-1} to $\psi(i)^{-1}$ (note that without the latter condition, there is not necessarily a unique graph morphism $\mathscr{G}[I^{-1}] \to \mathscr{C}$).

We can therefore mark our transformation schemata.

Definition 2.40 (Marked transformation schema)**.** A *marked transformation schema for* $\mathfrak{S}$ is a transformation schema $\mathfrak{T}\colon \mathcal{E} \to \mathcal{L}(\mathfrak{S}) \times_T \mathcal{L}(\mathfrak{S})$ along with a set of *marked labels* $\mathfrak{I} \subseteq \mathcal{E}$. The $\mathfrak{I}$-*marked basic transformations* are the elements of the set $L^{-1}(\mathfrak{I})$, i.e.

$$\mathrm{Bs}_{\mathfrak{I}} \coloneqq \mathrm{Bs}_{\mathfrak{I}}(\mathfrak{S}, \mathfrak{T}) \coloneqq \{\varphi \in \mathrm{Bs}(\mathfrak{S}, \mathfrak{T}) \mid L(\varphi) \in \mathfrak{I}\},$$

yielding marked directed graph $\mathscr{B}_{\mathfrak{I}} \coloneqq \mathscr{B}_{\mathfrak{I}}(\mathfrak{S}, \mathfrak{T}) \coloneqq (\mathscr{B}, \mathrm{Bs}_{\mathfrak{I}})$ and directed graph $\mathscr{B}[\mathrm{Bs}_{\mathfrak{I}}^{-1}]$. Restricting our terms to $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})$, so that $\mathrm{Bs}_{\mathfrak{O}}$ is the set of basic transformations between terms in $\mathfrak{O}$, and $\mathscr{B}_{\mathfrak{O}}$ is the graph whose vertices are the terms in $\mathfrak{O}$ and whose edges are the basic transformations in $\mathrm{Bs}_{\mathfrak{O}}$ between them, we let $\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}} \coloneqq \mathrm{Bs}_{\mathfrak{O}} \cap \mathrm{Bs}_{\mathfrak{I}}$, the $\mathfrak{I}$-marked basic transformations between terms in $\mathfrak{O}$, yielding marked directed graph $\mathscr{B}_{\mathfrak{O},\mathfrak{I}} \coloneqq (\mathscr{B}_{\mathfrak{O}}, \mathrm{Bs}_{\mathfrak{O},\mathfrak{I}})$ and directed graph $\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}^{-1}]$. Restricting our terms to $\mathcal{L}(\mathfrak{S})_V$ for a set of variables V for $\mathfrak{S}$, we let $\mathrm{Bs}_{V,\mathfrak{I}} \coloneqq \mathrm{Bs}_{\mathcal{L}(\mathfrak{S})_V,\mathfrak{I}}$, yielding marked directed graph $\mathscr{B}_{V,\mathfrak{I}} \coloneqq (\mathscr{B}_V, \mathrm{Bs}_{V,\mathfrak{I}})$ and directed graph $\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}]$.

In the rest of this chapter, we fix a set of marked labels $\mathfrak{I} \subseteq \mathcal{E}$ to obtain a marked transformation schema $(\mathfrak{T}, \mathfrak{I})$ for $\mathfrak{S}$, a set of $\mathfrak{I}$-marked basic transformations $\mathrm{Bs}_{\mathfrak{I}}$, a marked directed graph $\mathscr{B}_{\mathfrak{I}}$, and an associated directed graph $\mathscr{B}[\mathrm{Bs}_{\mathfrak{I}}^{-1}]$ with terms in $\mathcal{L}(\mathfrak{S})$ as vertices and all our basic transformations and inverses for the marked ones as edges.

Example 2.41 (Marking all edges)**.** Sometimes we may wish to mark every edge label in $\mathcal{E}$ as invertible, in which case the set $\mathrm{Bs}_{\mathcal{E}}$ of $\mathcal{E}$-marked basic transformations is the entirety of $\mathrm{Bs} \coloneqq \mathrm{Bs}(\mathfrak{S}, \mathfrak{T})$, forming the marked directed graph $\mathscr{B}_{\mathcal{E}} = (\mathscr{B}, \mathrm{Bs})$ and associated directed graph $\mathscr{B}[\mathrm{Bs}^{-1}]$.

For instance, continuing from Example 2.33, we mark every edge label in $\mathcal{E}_{\mathrm{M}}$ to obtain the marked transformation schema for monoidal categories, forming the marked directed graph

$\mathscr{B}^{\mathrm{M}}_{\mathcal{E}_{\mathrm{M}}} = (\mathscr{B}^{\mathrm{M}}, \mathrm{Bs}^{\mathrm{M}})$ and directed graph $\mathscr{B}^{\mathrm{M}}[(\mathrm{Bs}^{\mathrm{M}})^{-1}]$. Restricting to the vertex-set $\mathcal{L}^{\mathrm{M}}_k$, we have a marked directed graph $\mathscr{B}^{\mathrm{M}}_{\mathcal{E}_{\mathrm{M}},k} \coloneqq (\mathscr{B}^{\mathrm{M}}_k, \mathrm{Bs}^{\mathrm{M}}_k)$ and directed graph $\mathscr{B}^{\mathrm{M}}_k[(\mathrm{Bs}^{\mathrm{M}}_k)^{-1}]$.

For the marked transformation schema for symmetric monoidal categories, it turns out we will only need to mark the edge labels in $\mathcal{E}_{\mathrm{M}} = \{\alpha, \lambda, \rho\} \subset \{\alpha, \beta, \lambda, \rho\} = \mathcal{E}_{\mathrm{SM}}$, marks just the basic transformations $\mathrm{Bs}^{\mathrm{M}} \subset \mathrm{Bs}^{\mathrm{SM}}$. This yields the marked directed graph $\mathscr{B}^{\mathrm{SM}}_{\mathcal{E}_{\mathrm{M}}} = (\mathscr{B}^{\mathrm{SM}}, \mathrm{Bs}^{\mathrm{M}})$ and directed graph $\mathscr{B}^{\mathrm{SM}}[(\mathrm{Bs}^{\mathrm{M}})^{-1}]$. Restricting to the vertex-set $\mathcal{L}^{\mathrm{M}}_k$, we have a marked directed graph $\mathscr{B}^{\mathrm{SM}}_{\mathcal{E}_{\mathrm{M}},k} \coloneqq (\mathscr{B}^{\mathrm{SM}}_k, \mathrm{Bs}^{\mathrm{M}}_k)$ and directed graph $\mathscr{B}^{\mathrm{SM}}_k[(\mathrm{Bs}^{\mathrm{M}}_k)^{-1}]$.

Similarly, continuing from Example 2.34, we mark every edge label in $\mathcal{E}_{\mathrm{B}}(O)$ to obtain the marked transformation schema for bicategories with object-set O, forming the marked directed graph $\mathscr{B}^{\mathrm{B}}(O)_{\mathcal{E}_{\mathrm{B}}(O)} = (\mathscr{B}^{\mathrm{B}}(O), \mathrm{Bs}^{\mathrm{B}}(O))$ and directed graph $\mathscr{B}^{\mathrm{B}}(O)[\mathrm{Bs}^{\mathrm{B}}(O)^{-1}]$. Restricting to vertex-set $\mathcal{L}^{\mathrm{B}}_k(P)$ yields marked directed graph $\mathscr{B}^{\mathrm{B}}_{\mathcal{E}_{\mathrm{B}}(O),k} \coloneqq (\mathscr{B}^{\mathrm{B}}_k(P), \mathrm{Bs}^{\mathrm{B}}_k(P))$ and directed graph $\mathscr{B}^{\mathrm{B}}_k(P)[\mathrm{Bs}^{\mathrm{B}}_k(P)^{-1}]$.

In summary, our marked transformation schema $(\mathfrak{T}, \mathfrak{I})$ for our signature $\mathfrak{S}$ gives rise to a directed graph $\mathscr{B}[\mathrm{Bs}^{-1}_{\mathfrak{I}}]$ with terms of $\mathfrak{S}$ as vertices along with basic transformations and formal inverses for $\mathfrak{I}$-marked basic transformations as edges. In order to be able to compose these basic transformations and their inverses, we form the free category $F(\mathscr{B}[\mathrm{Bs}^{-1}_{\mathfrak{I}}])$, which we will use in the next section to parametrize our algebraic structures on categories. We conclude this section with some notational remarks regarding this free category and, more generally, to its subcategories $F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}^{-1}_{\mathfrak{D},\mathfrak{I}}])$ for $\mathfrak{D} \subseteq \mathcal{L}(\mathfrak{S})$.

Notation 2.42. Following Definition A.4, let $F(\mathscr{B}[\mathrm{Bs}^{-1}_{\mathfrak{I}}])$ be the free category on the directed graph $\mathscr{B}[\mathrm{Bs}^{-1}_{\mathfrak{I}}]$ and $F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}^{-1}_{\mathfrak{D},\mathfrak{I}}])$ be the free category on the full subgraph $\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}^{-1}_{\mathfrak{D},\mathfrak{I}}]$ spanned by some $\mathfrak{D} \subseteq \mathcal{L}(\mathfrak{S})$. We use $\bullet$ to denote (vertical) composition in $F(\mathscr{B}[\mathrm{Bs}^{-1}_{\mathfrak{I}}])$, and we write parentheses around its morphisms for clarity. An arbitrary morphism of $F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}^{-1}_{\mathfrak{D},\mathfrak{I}}])$ may then be written as

$$\nu \coloneqq (\varphi^{\pm}_m \bullet \cdots \bullet \varphi^{\pm}_1)$$

for $m \in \mathbb{N}$, for which each $\varphi^{\pm}_i \colon t_{i-1} \Rightarrow t_i$ is an edge in $\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}^{-1}_{\mathfrak{D},\mathfrak{I}}]$. That is, each $\varphi^{\pm}_i$ either bears a positive exponent, in which case it is a basic transformation $\varphi_i \colon t_{i-1} \Rightarrow t_i$ between terms in $\mathfrak{D}$, or a negative exponent, in which case it is φ^{-1}_i for an $\mathfrak{I}$-marked basic transformation $\varphi_i \colon t_i \Rightarrow t_{i-1}$ between terms in $\mathfrak{D}$. As $F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}_{\mathfrak{D},\mathfrak{I}}^{-1}])$ is free, despite the notation φ_i^{-1} here is *not* the categorical inverse of φ_i; later we will quotient out by a congruence that enforces an inverse law.

Recall from Definition 2.7 that $V(t)$ is the set of all variables in a term $t \in \mathcal{L}(\mathfrak{S})$ and from Definition 2.32 that $V(\varphi)$ is the set of all variables in the domain and codomain of a basic transformation $\varphi \in \mathrm{Bs}$. Given ν as above, we may further define

$$V(\nu) \coloneqq V(\varphi_1) \cup \cdots \cup V(\varphi_m) = V(t_0) \cup V(t_1) \cup \cdots \cup V(t_{m-1}) \cup V(t_m).$$

When ν is viewed as a path, $V(\nu)$ is the set of variables appearing in any vertex along ν.

2.3 Congruences parametrizing commutativity

We parametrize coherence axioms as local binary relations (see Definition A.5) on our free category $F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{I}}^{-1}])$.

Definition 2.43 (Commutativity schema, (restricted) coherence schema)**.** A *commutativity schema* for $(\mathfrak{S}, \mathfrak{T}, \mathfrak{I})$ is a local binary relation $\mathfrak{R}$ on the free category $F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{I}}^{-1}])$. We call the quadruple $\mathfrak{C} \coloneqq (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R})$ a *coherence schema*.

A triple $(\mathfrak{C}, V, \mathfrak{D})$ is a *restricted coherence schema* if $\mathfrak{C} = (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R})$ is a coherence schema, V is a finite set of variables for $\mathfrak{S}$, and $\mathfrak{D} \subseteq \mathcal{L}(\mathfrak{S})_V \subset \mathcal{L}(\mathfrak{S})$ is a (necessarily variable-finite; see Definition 2.7) subset of terms that is also subterm-closed (see Definition 2.18). Often we will denote a restricted coherence schema by just the pair $(\mathfrak{C}, \mathfrak{D})$, in which case V is understood to be $V(\mathfrak{D})$, the finite set variables appearing in $\mathfrak{D}$.

Example 2.44 (Coherence schema for monoidal categories)**.** Building off of Examples 2.33 and 2.41, we define a local binary relation $\mathfrak{R}^{\mathrm{M}}$ on the free category $\mathscr{F}^{\mathrm{LM}} \coloneqq F(\mathscr{B}^{\mathrm{M}})$, with no edges marked for inversion; here 'LM' stands for *lax monoidal*. Since $\mathscr{F}^{\mathrm{LM}}$ is a (wide) subcategory of the free category $\mathscr{F}^{\mathrm{M}} \coloneqq F(\mathscr{B}^{\mathrm{M}}[(\mathrm{Bs}^{\mathrm{M}})^{-1}])$, our local binary relation $\mathfrak{R}^{\mathrm{M}}$ will also be a local binary relation on $\mathscr{F}^{\mathrm{M}}$. The local binary relation will consist of exactly two relations, one corresponding to each coherence axiom for monoidal categories. In particular, $\mathfrak{R}^{\mathrm{M}}$ relates either path around each of the following diagrams in $\mathscr{F}^{\mathrm{LM}} \subset \mathscr{F}^{\mathrm{M}}$ (we suppress subscripts and empty pairs of parentheses in the basic transformations for readability):

$$\begin{tikzcd}
\otimes(\otimes(\otimes(x_1,x_2),x_3),x_4) \arrow[r, Rightarrow, "\otimes(\alpha,\mathrm{id})"] \arrow[dd, Rightarrow, "\alpha"'] & \otimes(\otimes(x_1,\otimes(x_2,x_3)),x_4) \arrow[d, Rightarrow, "\alpha"] \\
& \otimes(x_1,\otimes(\otimes(x_2,x_3),x_4)) \arrow[d, Rightarrow, "\otimes(\mathrm{id},\alpha)"] \\
\otimes(\otimes(x_1,x_2),\otimes(x_3,x_4)) \arrow[r, Rightarrow, "\alpha"] & \otimes(x_1,\otimes(x_2,\otimes(x_3,x_4)));
\end{tikzcd}$$

$$\begin{tikzcd}
\otimes(\otimes(x_1,e),x_2) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\otimes(\rho,\mathrm{id})"'] & & \otimes(x_1,\otimes(e,x_2)) \arrow[dl, Rightarrow, "\otimes(\mathrm{id},\lambda)"] \\
& \otimes(x_1,x_2). &
\end{tikzcd}$$

Then $\mathfrak{R}^{\mathrm{M}}$ on $\mathscr{F}^{\mathrm{M}}$ is the commutativity schema for monoidal categories, yielding the coherence schema $\mathfrak{C}_{\mathrm{M}} \coloneqq (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{M}})$ capturing the axiomatic definition of a monoidal category.

When no edges are marked, we expand $\mathfrak{R}^{\mathrm{M}}$ on $\mathscr{F}^{\mathrm{LM}}$ by adding three more relations to form $\mathfrak{R}^{\mathrm{LM}}$. This relation relates either path around the two diagrams above and the three diagrams below:

$$\begin{tikzcd}
\otimes(\otimes(e,x_1),x_2) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\otimes(\lambda,\mathrm{id})"'] & & \otimes(e,\otimes(x_1,x_2)) \arrow[dl, Rightarrow, "\lambda"] \\
& \otimes(x_1,x_2); &
\end{tikzcd}$$

$$\begin{tikzcd}
\otimes(\otimes(x_1,x_2),e) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\rho"'] & & \otimes(x_1,\otimes(x_2,e)) \arrow[dl, Rightarrow, "\otimes(\mathrm{id},\rho)"] \\
& \otimes(x_1,x_2); &
\end{tikzcd}$$

$$\begin{tikzcd}
\otimes(e,e) \arrow[r, Rightarrow, bend left, "\lambda"] \arrow[r, Rightarrow, bend right, "\rho"'] & e.
\end{tikzcd}$$

We thus have the coherence schema $\mathfrak{C}_{\mathrm{LM}} \coloneqq (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing, \mathfrak{R}^{\mathrm{LM}})$ capturing the axiomatic definition of what is sometimes called a *lax monoidal category*, where the associator and unitors need not be invertible, but we then need these additional axioms to ensure coherence.

For symmetric monoidal categories, we expand $\mathfrak{R}^{\mathrm{M}}$ differently to a local binary relation $\mathfrak{R}^{\mathrm{SM}}$ on the free category $\mathscr{F}^{\mathrm{SM}} \coloneqq F(\mathscr{B}^{\mathrm{SM}}[(\mathrm{Bs}^{\mathrm{M}})^{-1}])$ (of which $\mathscr{F}^{\mathrm{M}}$ is a subcategory) consisting of five relations, one for each coherence axiom for symmetric monoidal categories. So $\mathfrak{R}^{\mathrm{SM}}$ relates either path around the two preceding diagrams in $\mathscr{F}^{\mathrm{SM}}$ as well as the three following:

$$\begin{tikzcd} \otimes(x_1,x_2) \arrow[rr, equal, "\mathrm{id}"] \arrow[dr, Rightarrow, "\beta"'] & & \otimes(x_1,x_2) \\ & \otimes(x_2,x_1); \arrow[ur, Rightarrow, "\beta"'] & \end{tikzcd} \qquad \begin{tikzcd} \otimes(x_1,e) \arrow[rr, Rightarrow, "\beta"] \arrow[dr, Rightarrow, "\rho"'] & & \otimes(e,x_1) \arrow[dl, Rightarrow, "\lambda"] \\ & x_1; & \end{tikzcd}$$

$$\begin{tikzcd} \otimes(\otimes(x_1,x_2),x_3) \arrow[r, Rightarrow, "\alpha"] \arrow[d, Rightarrow, "\otimes(\beta,\mathrm{id})"'] & \otimes(x_1,\otimes(x_2,x_3)) \arrow[r, Rightarrow, "\beta"] & \otimes(\otimes(x_2,x_3),x_1) \arrow[d, Rightarrow, "\alpha"] \\ \otimes(\otimes(x_2,x_1),x_3) \arrow[r, Rightarrow, "\alpha"] & \otimes(x_2,\otimes(x_1,x_3)) \arrow[r, Rightarrow, "\otimes(\mathrm{id},\beta)"] & \otimes(x_2,\otimes(x_3,x_1)). \end{tikzcd}$$

Here the horizontal morphism labeled id in the upper left diagram is not a basic transformation but an identity morphism in the free category $\mathscr{F}^{\mathrm{SM}}$.

Then $\mathfrak{R}^{\mathrm{SM}}$ is the commutativity schema for symmetric monoidal categories, yielding the coherence schema $\mathfrak{C}_{\mathrm{SM}} \coloneqq (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{SM}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{SM}})$ capturing the axiomatic definition of a symmetric monoidal category (we do not mark $\beta \in \mathcal{E}_{\mathrm{SM}}$ because the upper left triangle already ensures β is invertible).

For each $k \in \mathbb{N}$, recall from Example 2.21 that $\mathcal{L}_k^{\mathrm{M,reg}} \subseteq \mathcal{L}_k^{\mathrm{M}} \subset \mathcal{L}^{\mathrm{M}}$ is the set of terms with scope $V_k = \{x_1 < \cdots < x_k\}$ that are regular, meaning they have no repeated variables. We will focus on the terms in $\mathcal{L}_k^{\mathrm{M,reg}}$ in the coherence theorem for symmetric monoidal categories. Then $(\mathfrak{C}_{\mathrm{SM}}, V_k, \mathcal{L}_k^{\mathrm{M,reg}})$ (or simply $(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}_k^{\mathrm{M,reg}})$, as $V(\mathcal{L}_k^{\mathrm{M,reg}}) = V_k$) is our restricted *regular* coherence schema for symmetric monoidal categories in k variables. We also have our restricted coherence schema $(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}})$ (resp. $(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$) for monoidal (resp. lax monoidal) categories in k variables.

Example 2.45 (Coherence schema for bicategories)**.** Continuing again from Example 2.41, we define a local binary relation $\mathfrak{R}^{\mathrm{B}}(O)$ on the free category $\mathscr{F}^{\mathrm{LB}}(O) \coloneqq F(\mathscr{B}^{\mathrm{B}}(O))$, with no edges marked for inversion; here 'LB' stands for *lax bicategory*. Since $\mathscr{F}^{\mathrm{LB}}(O)$ is a (wide) subcategory of the free category $\mathscr{F}^{\mathrm{B}}(O) \coloneqq F(\mathscr{B}^{\mathrm{B}}(O)[\mathrm{Bs}^{\mathrm{B}}(O)^{-1}])$, our local binary relation $\mathfrak{R}^{\mathrm{B}}(O)$ will also be a local binary relation on $\mathscr{F}^{\mathrm{B}}(O)$. The local binary relation $\mathfrak{R}^{\mathrm{B}}(O)$ will consist of two families of relations, one corresponding to each coherence axiom for bicategories. So $\mathfrak{R}^{\mathrm{B}}(O)$ relates either path around each of the following diagrams in $\mathscr{F}^{\mathrm{LB}}(O) \subset \mathscr{F}^{\mathrm{B}}(O)$ for $a, b, c, d, e \in O$:

$$\begin{tikzcd}
\circ\left(\circ\left(\circ(x_1, x_2), x_3\right), x_4\right) \arrow[r, Rightarrow, "\circ(\alpha,\mathrm{id})"] \arrow[dd, Rightarrow, "\alpha"'] & \circ\left(\circ\left(x_1, \circ(x_2, x_3)\right), x_4\right) \arrow[d, Rightarrow, "\alpha"] \\
& \circ\left(x_1, \circ\left(\circ(x_2, x_3), x_4\right)\right) \arrow[d, Rightarrow, "\circ(\mathrm{id},\alpha)"] \\
\circ\left(\circ(x_1, x_2), \circ\left(x_3, x_4\right)\right) \arrow[r, Rightarrow, "\alpha"] & \circ\left(x_1, \circ\left(x_2, \circ\left(x_3, x_4\right)\right)\right);
\end{tikzcd}$$

$$\begin{tikzcd}
\circ\left(\circ(x_1, \mathbf{1}_d), x_2\right) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\circ(\rho,\mathrm{id})"'] & & \circ\left(x_1, \circ\left(\mathbf{1}_d, x_2\right)\right) \arrow[dl, Rightarrow, "\circ(\mathrm{id},\lambda)"] \\
& \circ\left(x_1, x_2\right) &
\end{tikzcd}$$

with

$$x_1 \coloneqq x_1^{(d,e)}, \quad x_2 \coloneqq x_2^{(c,d)}, \quad x_3 \coloneqq x_3^{(b,c)}, \quad \text{and} \quad x_4 \coloneqq x_4^{(a,b)}.$$

Then $\mathfrak{R}^{\mathrm{B}}(O)$ is the commutativity schema for bicategories with object-set O, yielding the coherence schema $\mathfrak{C}_{\mathrm{B}}(O) \coloneqq (\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \mathcal{E}_{\mathrm{B}}(O), \mathfrak{R}^{\mathrm{B}}(O))$ capturing the axiomatic definition of a bicategory with object-set O.

When no edges are marked, we expand $\mathfrak{R}^{\mathrm{B}}(O)$ on $\mathscr{F}^{\mathrm{LB}}(O)(O)$ by adding three more families of relations to form $\mathfrak{R}^{\mathrm{LB}}(O)$, relating either path around each of the following diagrams for $a, b, c \in O$:

$$\begin{tikzcd}
\circ(\circ(\mathbf{1}_c, x_1), x_2) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\circ(\lambda,\mathrm{id})"'] & & \circ(\mathbf{1}_c, \circ(x_1, x_2)) \arrow[dl, Rightarrow, "\lambda"] \\
& \circ(x_1, x_2); &
\end{tikzcd}$$

$$\begin{tikzcd}
\circ(\circ(x_1, x_2), \mathbf{1}_a) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\rho"'] & & \circ(x_1, \circ(x_2, \mathbf{1}_a)) \arrow[dl, Rightarrow, "\circ(\mathrm{id},\rho)"] \\
& \circ(x_1, x_2); &
\end{tikzcd}$$

$$\begin{tikzcd}
\circ(\mathbf{1}_a, \mathbf{1}_a) \arrow[r, Rightarrow, bend left, "\lambda"] \arrow[r, Rightarrow, bend right, "\rho"'] & \mathbf{1}_a.
\end{tikzcd}$$

with

$$x_1 \coloneqq x_1^{(b,c)} \quad \text{and} \quad x_2 \coloneqq x_2^{(a,b)}.$$

We thus have the coherence schema $\mathfrak{C}_{\mathrm{LB}}(O) \coloneqq (\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \varnothing, \mathfrak{R}^{\mathrm{LB}}(O))$ capturing the axiomatic definition of what we shall call a *lax bicategory* with object-set O, where the associator and unitors need not be invertible, but we then need these additional axioms to ensure coherence.

For each $k \in \mathbb{N}$ and finite subset $P \subseteq O$, we have restricted coherence schemas given by $(\mathfrak{C}_{\mathrm{LB}}(O), \mathcal{L}_k^{\mathrm{B}}(P))$ and $(\mathfrak{C}_{\mathrm{B}}(O), \mathcal{L}_k^{\mathrm{B}}(P))$ (see Example 2.8), with the latter being our restricted coherence schema in k variables for bicategories with objects in $P \subseteq O$. (Note here that $V(\mathcal{L}_k^{\mathrm{B}}(P)) = V_k(P)$.)

For the rest of this chapter, we fix a commutativity schema $\mathfrak{R}$ for $(\mathfrak{S}, \mathfrak{T}, \mathfrak{I})$, yielding a coherence schema $\mathfrak{C} = (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R})$.

The relations in $\mathfrak{R}$ parametrize the axioms asserting commutativity of certain diagrams of coherence maps. By forming horizontal composites and appealing to inverse laws, functoriality, and naturality, we can deduce additional commutativity conditions between the composites of basic transformations that comprise the morphisms of $F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{I}}^{-1}])$. We formalize this process as follows, restricting to a finite set of variables V when we need to fix the scope.

Definition 2.46. Given a set of (not necessarily finite[8]) variables V for $\mathfrak{S}$, construct a local binary relation $R(\mathfrak{C})_V$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ as follows. As in Notation 2.42, we write $\bullet$ for (vertical) composition in $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$, and we use parentheses for clarity. We first construct four local binary relations $R^{\mathrm{inv}} \coloneqq R^{\mathrm{inv}}(\mathfrak{C})_V, R^{\mathrm{fun}} \coloneqq R^{\mathrm{fun}}(\mathfrak{C})_V, R^{\mathrm{nat}} \coloneqq R^{\mathrm{nat}}(\mathfrak{C})_V$, and $R^{\mathrm{com}} \coloneqq R^{\mathrm{com}}(\mathfrak{C})_V$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ like so. (See Example 2.47 for concrete examples of each relation.)

(a) The local binary relation R^{inv} consists of the relations

$$(\varphi^{-1} \bullet \varphi) \;\; R_{s,s}^{\mathrm{inv}} \;\; \mathrm{id}_s \quad \text{and} \quad (\varphi \bullet \varphi^{-1}) \;\; R_{t,t}^{\mathrm{inv}} \;\; \mathrm{id}_t$$

[8] We will always take V to be either finite or the set of all variables for $\mathfrak{S}$.

for all marked basic transformations $\varphi\colon s \Rightarrow t$ in $\mathrm{Bs}_{V,\mathfrak{J}}$ (i.e. all marked basic transformations $\varphi\colon s \Rightarrow t$ in $\mathrm{Bs}_{\mathfrak{J}}$ for which $s, t \in \mathcal{L}(\mathfrak{S})_V$).[9]

(b) The local binary relation R^{fun} is defined as follows. Recall from Definition 2.35 that the *atom* $A(\varphi)$ of a basic transformation φ is the unique atomic transformation appearing in φ; that the *span* $S(\varphi) = (S_1(\varphi), S_2(\varphi))$ of φ is defined so that $\mathrm{dom}(A(\varphi))$ is the $S(\varphi)$-subterm of $\mathrm{dom}(\varphi)$; and that the *difference* of φ is given by the formula $\Delta(\varphi) = \mathrm{len}(\mathrm{cod}(\varphi)) - \mathrm{len}(\mathrm{dom}(\varphi))$. So for any two basic transformations $\varphi\colon s \Rightarrow t_1$ and $\psi\colon s \Rightarrow t_2$ in Bs_V with $S_2(\varphi) < S_1(\psi)$, we have that s has $\mathrm{dom}(A(\varphi))$ as its $S(\varphi)$-subterm and $\mathrm{dom}(A(\psi))$ as its $S(\psi)$-subterm. Then:

(i) The term t_1 is given by replacing the $S(\varphi)$-subterm of s with $\mathrm{cod}(A(\varphi))$ without changing the rest of the term, so t_1 has $\mathrm{dom}(A(\psi))$ as its $(S(\psi)+\Delta(\varphi))$-subterm. So we can let $\psi'\colon t_1 \Rightarrow u$ be the unique basic transformation with domain t_1, atom $A(\psi)$, and span $S(\psi) + \Delta(\varphi)$ guaranteed by Lemma 2.37.

(ii) The term t_2 is given by replacing the $S(\psi)$-subterm of s with $\mathrm{cod}(A(\psi))$, so t_2 still has $\mathrm{dom}(A(\varphi))$ as its $S(\varphi)$-subterm. So let $\varphi'\colon t_2 \Rightarrow u'$ be the unique basic transformation with domain t_2, atom $A(\varphi)$, and span $S(\varphi)$ from Lemma 2.37.

Then necessarily $u = u'$. By Lemma 2.37, since $\varphi, \psi \in \mathrm{Bs}_V$, we have

$$V(\psi') = V(t_1) \cup V(\mathrm{cod}(A(\psi))) \subseteq V(\varphi) \cup V(\psi) \subseteq V,$$

so $u = \mathrm{cod}(\psi') \in \mathcal{L}(\mathfrak{S})_V$ and thus $\varphi', \psi' \in \mathrm{Bs}_V$. So we declare

$$(\psi' \bullet \varphi) \;\; R^{\mathrm{fun}}_{s,u} \;\; (\varphi' \bullet \psi).$$

(c) The local binary relation R^{nat} is defined as follows. Given an atomic transformation $\varepsilon_{\mathbf{s}}\colon t_0 \Rightarrow t_0'$ in $\mathscr{B}_V$ with $\mathbf{s} = (s_1, \ldots, s_k) \in \mathcal{L}(\mathfrak{S})_V^k$, so $t_0 = \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}] \in \mathcal{L}(\mathfrak{S})_V$ and $t_0' = \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}] \in \mathcal{L}(\mathfrak{S})_V$, fix some $1 \le \ell \le k$ and a basic transformation $\psi\colon s_\ell \Rightarrow s_\ell'$ in $\mathscr{B}_V$. Then:

---

[9] Recall that when forming $\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}_{\mathfrak{D},\mathfrak{J}}^{-1}]$, we only take formal inverses of entire basic transformations, never their constituent parts. Syntactically, this means that the symbol $^{-1}$ can only ever appear at the very end of a string denoting an edge of $\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}_{\mathfrak{D},\mathfrak{J}}^{-1}]$. We will interpret these inverses in (2.73).

(i) Let the pairs $(a_1, b_1), \ldots, (a_m, b_m)$ denote the positions of all the instances of s_ℓ in t_0, ordered right to left (so $a_m < b_m < \cdots < a_1 < b_1$), arising from substitutions of s_ℓ for the ℓ^{th} variable of $V(\varepsilon)$ in $\operatorname{dom}(\varepsilon)$. For $1 \le i \le m$, we recursively define $\psi_i \colon t_{i-1} \Rightarrow t_i$ in Bs_V to be the unique basic transformation with domain t_{i-1}, atom $A(\psi)$, and span $S(\psi) + a_i - 1$ given by Lemma 2.37.

(ii) Similarly, let the pairs $(a'_1, b'_1), \ldots, (a'_n, b'_n)$ denote the positions of the instances of s_ℓ in t'_0 ordered right to left ($a_n < b_n < \cdots < a_1 < b_1$) arising from substitutions of s_ℓ for the ℓ^{th} variable of $V(\varepsilon)$ in $\operatorname{cod}(\varepsilon)$. For $1 \le j \le n$, recursively define $\psi'_j \colon t'_{j-1} \Rightarrow t'_j$ in Bs_V to be the unique basic transformation with domain t'_{j-1}, atom $A(\psi)$, and span $S(\psi) + a'_j - 1$ given by Lemma 2.37.

It follows that there is an atomic transformation $\varepsilon_{\mathbf{s}'} \colon t_m \Rightarrow t'_n$ if we define $\mathbf{s}' = (s_1, \ldots, s_{\ell-1}, s'_\ell, s_{\ell+1}, \ldots, s_k)$. So we declare

$$(\psi'_n \bullet \cdots \bullet \psi'_1 \bullet \varepsilon_{\mathbf{s}}) \; R^{\text{nat}}_{t_0, t'_n} \; (\varepsilon_{\mathbf{s}'} \bullet \psi_m \bullet \cdots \bullet \psi_1).$$

(d) The local relation R^{com} is defined as follows. For each $\mathfrak{R}$-relation

$$\nu = (\varphi_m^{\pm 1} \bullet \cdots \bullet \varphi_1^{\pm 1}) \; \mathfrak{R} \; (\psi_n^{\pm 1} \bullet \cdots \bullet \psi_1^{\pm 1}) = \nu',$$

where the φ_i's and ψ_j's are basic transformations, define the finite set of variables

$$W \coloneqq V(\nu) \cup V(\nu') = V(\varphi_1) \cup \cdots \cup V(\varphi_m) \cup V(\psi_1) \cup \cdots \cup V(\psi_n)$$

appearing in either ν or ν' as in Notation 2.42. Order and label the variables in W so that $W = \{x_{\ell_1}^{\mathtt{t}_1} < \cdots < x_{\ell_k}^{\mathtt{t}_k}\}$. If $\mathbf{r} = (r_1, \ldots, r_k) \in \mathcal{L}(\mathfrak{S})_V^k$ is a k-tuple of terms such that $\tau(r_p) = \mathtt{t}_p$ for all $1 \le p \le k$, then

$$(\varphi_m[W/\mathbf{r}]^{\pm 1} \bullet \cdots \bullet \varphi_1[W/\mathbf{r}]^{\pm 1}) \; R^{\text{com}} \; (\psi_n[W/\mathbf{r}]^{\pm 1} \bullet \cdots \bullet \psi_1[W/\mathbf{r}]^{\pm 1}).$$

Finally, define the local binary relation $R(\mathfrak{C})_V$ recursively as follows. First, we include $R^{\text{inv}}, R^{\text{fun}}, R^{\text{nat}}$, and R^{com} in $R(\mathfrak{C})_V$. Then suppose

$$(\varphi_m^{\pm 1} \bullet \cdots \bullet \varphi_1^{\pm 1}) \; R(\mathfrak{C})_V \; (\psi_n^{\pm 1} \bullet \cdots \bullet \psi_1^{\pm 1})$$

where the φ_i's and ψ_j's are basic transformations in Bs_V, and let their common type be $\mathtt{t}$. Take a p-ary function symbol $f \in F$ and $t_1, \ldots, t_{k-1}, t_{k+1}, \ldots, t_p \in \mathcal{L}(\mathfrak{S})_V$ so that $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_{k-1}), \mathtt{t}, \tau(t_{k+1}), \ldots, \tau(t_p))$. For each basic transformation φ with $\tau(\varphi) = \mathtt{t}$, define the basic transformation $\mathcal{F}(\varphi) \coloneqq f(\mathrm{id}_{t_1}, \ldots, \mathrm{id}_{t_{k-1}}, \varphi, \mathrm{id}_{t_{k+1}}, \ldots, \mathrm{id}_{t_p})$. Then

$$(\mathcal{F}(\varphi_m)^{\pm 1} \bullet \cdots \bullet \mathcal{F}(\varphi_1)^{\pm 1}) \;\; R(\mathfrak{C})_V \;\; (\mathcal{F}(\psi_n)^{\pm 1} \bullet \cdots \bullet \mathcal{F}(\psi_1)^{\pm 1}).$$

This concludes the construction of our local binary relation $R(\mathfrak{C})_V$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$, which in turn generates a congruence $\langle R(\mathfrak{C})_V \rangle$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$ as in Example A.7 (4). When V is the set of all variables for $\mathfrak{S}$, we simply write $R(\mathfrak{C}) \coloneqq R(\mathfrak{C})_V$; it is a local binary relation on $F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{J}}^{-1}])$.

Example 2.47. We give examples of each of the four constituent local binary relations defined in Definition 2.46 on the subcategory $\mathscr{F}_4^{\mathrm{SM}} \coloneqq F(\mathscr{B}_{V_4}^{\mathrm{SM}}[(\mathrm{Bs}_{V_4}^{\mathrm{SM}})^{-1}])$ of the free category $\mathscr{F}^{\mathrm{SM}}$ from Example 2.44 (with $V_4 = \{x_1 < x_2 < x_3 < x_4\}$) that generate the local binary relation $R \coloneqq R(\mathfrak{C}_{\mathrm{SM}})$ on $\mathscr{F}_4^{\mathrm{SM}}$ for our coherence schema for symmetric monoidal categories.

(a) The relation R^{inv} on $\mathscr{F}_4^{\mathrm{SM}}$ relates either path around each of the following diagrams:

$$\begin{tikzcd} \otimes(\otimes(x_2,x_3),x_1) \arrow[r, Rightarrow, "\otimes(\beta_{x_2,x_3},\mathrm{id}_{x_1})"] \arrow[dr, Rightarrow, "\mathrm{id}_{\otimes(\otimes(x_2,x_3),x_1)}"'] & \otimes(\otimes(x_3,x_2),x_1) \arrow[d, Rightarrow, "\otimes(\beta_{x_2,x_3},\mathrm{id}_{x_1})^{-1}"] \\ & \otimes(\otimes(x_2,x_3),x_1); \end{tikzcd}$$

$$\begin{tikzcd} \otimes(\otimes(x_3,x_2),x_1) \arrow[r, Rightarrow, "\otimes(\beta_{x_2,x_3},\mathrm{id}_{x_1})^{-1}"] \arrow[dr, Rightarrow, "\mathrm{id}_{\otimes(\otimes(x_3,x_2),x_1)}"'] & \otimes(\otimes(x_2,x_3),x_1) \arrow[d, Rightarrow, "\otimes(\beta_{x_2,x_3},\mathrm{id}_{x_1})"] \\ & \otimes(\otimes(x_3,x_2),x_1). \end{tikzcd}$$

(b) The relation R^{fun} on $\mathscr{F}_4^{\mathrm{SM}}$ relates either path around the following diagram:

$$\begin{tikzcd} \otimes(\otimes(x_2,x_3),\otimes(e,x_1)) \arrow[r, Rightarrow, "\otimes\left(\beta_{x_2,x_3},\mathrm{id}_{\otimes(e,x_1)}\right)"] \arrow[d, Rightarrow, "\otimes\left(\mathrm{id}_{\otimes(x_2,x_3)},\lambda_{x_1}\right)"'] & \otimes(\otimes(x_3,x_2),\otimes(e,x_1)) \arrow[d, Rightarrow, "\otimes\left(\mathrm{id}_{\otimes(x_3,x_2)},\lambda_{x_1}\right)"] \\ \otimes(\otimes(x_2,x_3),x_1) \arrow[r, Rightarrow, "\otimes(\beta_{x_2,x_3},\mathrm{id}_{x_1})"] & \otimes(\otimes(x_3,x_2),x_1). \end{tikzcd}$$

(c) The relation R^{nat} on $\mathscr{F}_4^{\mathrm{SM}}$ relates either path around the following diagram:

$$\begin{tikzcd}
\otimes(\otimes(x_1,x_2),\otimes(x_3,e)) \arrow[r, Rightarrow, "\alpha_{x_1,x_2,\otimes(x_3,e)}"] \arrow[d, Rightarrow, "\otimes\left(\mathrm{id}_{\otimes(x_1,x_2)},\rho_{x_3}\right)"'] & \otimes(x_1,\otimes(x_2,\otimes(x_3,e))) \arrow[d, Rightarrow, "\otimes(\mathrm{id}_{x_1},\otimes(\mathrm{id}_{x_2},\rho_{x_3}))"] \\
\otimes(\otimes(x_1,x_2),x_3) \arrow[r, Rightarrow, "\alpha_{x_1,x_2,x_3}"] & \otimes(x_1,\otimes(x_2,x_3)).
\end{tikzcd}$$

For a more elaborate example, say we had an additional binary function symbol $\oplus$ in our signature and an additional edge

$$\delta\colon \otimes(\oplus(x_1,x_2),x_3) \Rightarrow \oplus(\otimes(x_1,x_3),\otimes(x_2,x_3))$$

in our transformation schema. Then the relation R^{nat} would also relate either path around the following diagram, with some subscripts suppressed for clarity:

$$\begin{tikzcd}
\otimes(\oplus(x_1,x_2),\otimes(x_3,e)) \arrow[r, Rightarrow, "\delta_{x_1,x_2,\otimes(x_3,e)}"] \arrow[dd, Rightarrow, "\otimes(\mathrm{id},\rho)"'] & \oplus(\otimes(x_1,\otimes(x_3,e)),\otimes(x_2,\otimes(x_3,e))) \arrow[d, Rightarrow, "\oplus(\mathrm{id},\otimes(\mathrm{id},\rho))"] \\
& \oplus(\otimes(x_1,\otimes(x_3,e)),\otimes(x_2,x_3)) \arrow[d, Rightarrow, "\oplus(\otimes(\mathrm{id},\rho),\mathrm{id})"] \\
\otimes(\oplus(x_1,x_2),x_3) \arrow[r, Rightarrow, "\delta_{x_1,x_2,x_3}"] & \oplus(\otimes(x_1,x_3),\otimes(x_2,x_3)).
\end{tikzcd}$$

(d) The relation R^{com} on $\mathscr{F}_4^{\mathrm{SM}}$ relates either path around the following diagram:

$$\begin{tikzcd}
\otimes(\otimes(e,e),\otimes(x_1,x_3)) \arrow[r, Rightarrow, "\otimes\left(\rho_e,\mathrm{id}_{\otimes(x_1,x_3)}\right)"] \arrow[d, Rightarrow, "\alpha_{e,e,\otimes(x_1,x_3)}"'] & \otimes(e,\otimes(x_1,x_3)) \\
\otimes(e,\otimes(e,\otimes(x_1,x_3))). \arrow[ur, Rightarrow, "\otimes\left(\mathrm{id}_e,\lambda_{\otimes(x_1,x_3)}\right)"'] &
\end{tikzcd}$$

Finally, the recursive step applied to the preceding diagram ensures that $R(\mathfrak{C})$ relates either path around each of the following diagrams (with some subscripts suppressed for clarity). The first diagram is obtained by applying $\otimes(-,\mathrm{id}_{x_2})$ to each morphism in the diagram above, while the second diagram is obtained by applying $\otimes(\mathrm{id}_{\otimes(x_4,e)},-)$ to each morphism in the first.

$$\begin{tikzcd}
\otimes(\otimes(\otimes(e,e),\otimes(x_1,x_3)),x_2) \arrow[r, Rightarrow, "\otimes(\otimes(\rho,\mathrm{id}),\mathrm{id})"] \arrow[d, Rightarrow, "\otimes(\alpha,\mathrm{id})"'] & \otimes(\otimes(e,\otimes(x_1,x_3)),x_2) \\
\otimes(\otimes(e,\otimes(e,\otimes(x_1,x_3))),x_2); \arrow[ur, Rightarrow, "\otimes(\otimes(\mathrm{id},\lambda),\mathrm{id})"'] &
\end{tikzcd}$$

$$\begin{tikzcd}
\otimes(\otimes(x_4,e),\otimes(\otimes(\otimes(e,e),\otimes(x_1,x_3)),x_2)) \arrow[r, Rightarrow, "{\otimes(\mathrm{id},\otimes(\otimes(\rho,\mathrm{id}),\mathrm{id}))}"] \arrow[d, Rightarrow, "{\otimes(\mathrm{id},\otimes(\alpha,\mathrm{id}))}"'] & \otimes(\otimes(x_4,e),\otimes(\otimes(e,\otimes(x_1,x_3)),x_2)) \\
\otimes(\otimes(x_4,e),\otimes(\otimes(e,\otimes(e,\otimes(x_1,x_3))),x_2)). \arrow[ur, Rightarrow, "{\otimes(\mathrm{id},\otimes(\otimes(\mathrm{id},\lambda),\mathrm{id}))}"'] &
\end{tikzcd}$$

Remark 2.48. Given two sets of variables $W \subset V$ for $\mathfrak{S}$, we have $\mathcal{L}(\mathfrak{S})_W \subset \mathcal{L}(\mathfrak{S})_V$, with $\mathscr{B}_W$ the full subgraph of $\mathscr{B}_V$ spanned by $\mathcal{L}(\mathfrak{S})_W$ and $\mathrm{Bs}_{W,\mathfrak{I}}$ precisely the edges in $\mathrm{Bs}_{V,\mathfrak{I}}$ that appear in $\mathscr{B}_W$. Then $F(\mathscr{B}_W[\mathrm{Bs}_{W,\mathfrak{I}}^{-1}])$ is a subcategory of $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$, although it is not necessarily full: there may be a path in $\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}]$ between two vertices (terms) in $\mathcal{L}(\mathfrak{S})_W$ that passes through vertices that are not in $\mathcal{L}(\mathfrak{S})_W$, and such a path would be a morphism in $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ that is not in $F(\mathscr{B}_W[\mathrm{Bs}_{W,\mathfrak{I}}^{-1}])$ even though its domain and codomain are.

Then by Definition 2.46, a pair of morphisms in $F(\mathscr{B}_W[\mathrm{Bs}_{W,\mathfrak{I}}^{-1}])$ that are $R(\mathfrak{C})_W$-related are necessarily also $R(\mathfrak{C})_V$-related in $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$, inducing a functor between quotients

$$\frac{F(\mathscr{B}_W[\mathrm{Bs}_{W,\mathfrak{I}}^{-1}])}{\langle R(\mathfrak{C})_W\rangle} \to \frac{F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])}{\langle R(\mathfrak{C})_V\rangle}.$$

We prove the following property of the congruence $\langle R(\mathfrak{C})_V\rangle$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ from Definition 2.46.

Lemma 2.49. *Suppose*

$$(\varphi_m^{\pm 1} \bullet \cdots \bullet \varphi_1^{\pm 1}) \ \langle R(\mathfrak{C})_V\rangle \ (\psi_n^{\pm 1} \bullet \cdots \bullet \psi_1^{\pm 1})$$

where the φ_i's and ψ_j's are basic transformations in Bs_V*, and let their common type be* $\mathtt{t}$*. Take a p-ary function symbol $f \in F$ and $t_1, \ldots, t_{k-1}, t_{k+1}, \ldots, t_p \in \mathcal{L}(\mathfrak{S})_V$ so that* $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_{k-1}), \mathtt{t}, \tau(t_{k+1}), \ldots, \tau(t_p))$*. For each basic transformation φ with $\tau(\varphi) = \mathtt{t}$, define the basic transformation $\mathcal{F}(\varphi) \coloneqq f(\mathrm{id}_{t_1}, \ldots, \mathrm{id}_{t_{k-1}}, \varphi, \mathrm{id}_{t_{k+1}}, \ldots, \mathrm{id}_{t_p})$. Then applying $\mathcal{F}$ preserves $\langle R(\mathfrak{C})_V\rangle$-relations; that is,*

$$(\mathcal{F}(\varphi_m)^{\pm 1} \bullet \cdots \bullet \mathcal{F}(\varphi_1)^{\pm 1}) \ \langle R(\mathfrak{C})_V\rangle \ (\mathcal{F}(\psi_n)^{\pm 1} \bullet \cdots \bullet \mathcal{F}(\psi_1)^{\pm 1}).$$

Proof. By construction, applying $\mathcal{F}$ preserves $R(\mathfrak{C})_V$-relations. Following Example A.7 (4), we construct $\langle R(\mathfrak{C})_V\rangle$ by defining a local binary relation $\overline{R(\mathfrak{C})_V}$ by $(k \bullet f \bullet h)\,\overline{R(\mathfrak{C})_V}\,(k \bullet f' \bullet h)$ if $f \; R(\mathfrak{C})_V \; f'$, so that $\langle R(\mathfrak{C})_V\rangle$ is the reflexive, symmetric, and transitive closure of $\overline{R(\mathfrak{C})_V}$.

First we show that applying $\mathcal{F}$ preserves $\overline{R(\mathfrak{C})_V}$-relations. Each $\overline{R(\mathfrak{C})_V}$-relation can be written in the form

$$\begin{aligned} &(\kappa_q^{\pm 1} \bullet \cdots \bullet \kappa_1^{\pm 1} \bullet \varphi_m^{\pm 1} \bullet \ldots \bullet \varphi_1^{\pm 1} \bullet \eta_p^{\pm 1} \bullet \cdots \bullet \eta_1^{\pm 1}) \\ \overline{R(\mathfrak{C})_V}\ &(\kappa_q^{\pm 1} \bullet \cdots \bullet \kappa_1^{\pm 1} \bullet \psi_n^{\pm 1} \bullet \ldots \bullet \psi_1^{\pm 1} \bullet \eta_p^{\pm 1} \bullet \cdots \bullet \eta_1^{\pm 1}) \end{aligned}$$

for $(\varphi_m^{\pm 1} \bullet \cdots \bullet \varphi_1^{\pm 1})\ R(\mathfrak{C})_V\ (\psi_n^{\pm 1} \bullet \cdots \bullet \psi_1^{\pm 1})$. It follows $(\mathcal{F}(\varphi_m)^{\pm 1} \bullet \cdots \bullet \mathcal{F}(\varphi_1)^{\pm 1})$ is $R(\mathfrak{C})_V$-related to $(\mathcal{F}(\psi_n)^{\pm 1} \bullet \cdots \bullet \mathcal{F}(\psi_1)^{\pm 1})$, so

$$\begin{aligned} &(\mathcal{F}(\kappa_q^{\pm 1}) \bullet \cdots \bullet \mathcal{F}(\kappa_1)^{\pm 1} \bullet F(\varphi_m)^{\pm 1} \bullet \ldots \bullet F(\varphi_1)^{\pm 1} \bullet F(\eta_p)^{\pm 1} \bullet \cdots \bullet F(\eta_1)^{\pm 1}) \\ \overline{R(\mathfrak{C})_V}\ &(\mathcal{F}(\kappa_q^{\pm 1}) \bullet \cdots \bullet \mathcal{F}(\kappa_1)^{\pm 1} \bullet F(\psi_n)^{\pm 1} \bullet \ldots \bullet F(\psi_1)^{\pm 1} \bullet F(\eta_p)^{\pm 1} \bullet \cdots \bullet F(\eta_1)^{\pm 1}). \end{aligned}$$

Finally, if an operation preserves a relation, then it preserves its reflexive, symmetric, and transitive closure as well, so the result follows. ☐

Sometimes our commutativity schema $\mathfrak{R}$ may contain redundant axioms: that is, there exists a proper subrelation $\mathfrak{R}' \subsetneq \mathfrak{R}$ defining a modified coherence schema $\mathfrak{C}' = (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R}')$ for which the congruence $\langle R(\mathfrak{C}')_V \rangle$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ as constructed in Definition 2.46 with $\mathfrak{C}'$ in place of $\mathfrak{C}$ is in fact equal to the congruence $\langle R(\mathfrak{C})_V \rangle$ generated from the larger commutativity schema $\mathfrak{R}$. We give a sufficient condition for when this holds as follows.

Lemma 2.50. *Given a coherence schema $\mathfrak{C} = (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R})$ and a set of variables V with a congruence $\langle R(\mathfrak{C})_V \rangle$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ as in Definition 2.46, let $\mathfrak{R}' \subset \mathfrak{R}$ be a subrelation. Construct a new coherence schema $\mathfrak{C}' = (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R}')$ with $\mathfrak{R}'$ replacing $\mathfrak{R}$ as the commutativity schema, giving rise to a congruence $\langle R(\mathfrak{C}')_V \rangle$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ by Definition 2.46. If the subrelation $R^{\mathrm{com}}(\mathfrak{C})_V$ of $R(\mathfrak{C})_V$ is contained in $\langle R(\mathfrak{C}') \rangle_V$, then $\langle R(\mathfrak{C}')_V \rangle = \langle R(\mathfrak{C})_V \rangle$.*

Proof. With $\mathfrak{R}' \subset \mathfrak{R}$, we have $R(\mathfrak{C}')_V \subset R(\mathfrak{C})_V$, so $\langle R(\mathfrak{C}')_V \rangle \subset \langle R(\mathfrak{C})_V \rangle$. For the opposite direction, every relation in $R^{\mathrm{inv}}(\mathfrak{C})_V, R^{\mathrm{fun}}(\mathfrak{C})_V$, or $R^{\mathrm{nat}}(\mathfrak{C})_V$ is already also in $R(\mathfrak{C}')_V$, so because $R^{\mathrm{com}}(\mathfrak{C})_V \subset \langle R(\mathfrak{C}') \rangle_V$ and because applying $\mathcal{F}$ preserves $\langle R(\mathfrak{C}')_V \rangle$-relations, it follows that $R(\mathfrak{C})_V \subset \langle R(\mathfrak{C}')_V \rangle$. Hence $\langle R(\mathfrak{C})_V \rangle \subset \langle R(\mathfrak{C}')_V \rangle$. ☐

We are now ready to form our universal parameter categories using $R(\mathfrak{C})_V$.

Definition 2.51 (Universal parameter categories presented by coherence schemata)**.** Let V be a finite set of variables for $\mathfrak{S}$. Recall that $\mathcal{L}(\mathfrak{S})_V \subset \mathcal{L}(\mathfrak{S})$ is the set of all terms with scope V; that $\mathscr{B}_V$ is the full subgraph of $\mathscr{B}$ spanned by $\mathcal{L}(\mathfrak{S})_V$, whose edges comprise the set Bs_V; that $\mathrm{Bs}_{V,\mathfrak{J}} = \mathrm{Bs}_{\mathfrak{J}} \cap \mathrm{Bs}_V$; and that $R(\mathfrak{C})_V$ is a local binary relation on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$. The *(universal parameter) category presented by the restricted coherence schema* $(\mathfrak{C}, \mathcal{L}(\mathfrak{S})_V)$ is then the quotient category (see Definition A.6)

$$\mathscr{P}_V(\mathfrak{C}) \coloneqq \frac{F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])}{\langle R(\mathfrak{C})_V \rangle}.$$

As we saw in Remark 2.48, for every $W \subseteq V$, there is an induced functor $\mathscr{P}_W(\mathfrak{C}) \to \mathscr{P}_V(\mathfrak{C})$.

In addition, the relation $R(\mathfrak{C})_V$ generates a left-cancelable congruence $\langle R(\mathfrak{C})_V \rangle^{\mathrm{L}}$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$ as in (A.12). Then the *monomorphic (universal parameter) category presented by the restricted coherence schema* $(\mathfrak{C}, \mathcal{L}(\mathfrak{S})_V)$ is the quotient category

$$\mathscr{P}_V^{\mathrm{L}}(\mathfrak{C}) \coloneqq \frac{F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])}{\langle R(\mathfrak{C})_V \rangle^{\mathrm{L}}},$$

and there is a quotient functor $\mathscr{P}_V(\mathfrak{C}) \to \mathscr{P}_V^{\mathrm{L}}(\mathfrak{C})$ and an induced functor $\mathscr{P}_W^{\mathrm{L}}(\mathfrak{C}) \to \mathscr{P}_V^{\mathrm{L}}(\mathfrak{C})$ for $W \subseteq V$.

More generally, fix a variable-finite, subterm-closed subset $\mathfrak{D} \subseteq \mathcal{L}(\mathfrak{S})_V$. So $\mathscr{B}_{\mathfrak{D}}$ is the full subgraph of $\mathscr{B}_V$ spanned by $\mathfrak{D}$, whose edges comprise the set $\mathrm{Bs}_{\mathfrak{D}} \subseteq \mathrm{Bs}_V$, while $\mathrm{Bs}_{\mathfrak{D},\mathfrak{J}} = \mathrm{Bs}_{\mathfrak{J}} \cap \mathrm{Bs}_{\mathfrak{D}} \subseteq \mathrm{Bs}_{V,\mathfrak{J}}$. Then the free category $F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}_{\mathfrak{D},\mathfrak{J}}^{-1}])$ embeds as a (not necessarily full[10]) subcategory of $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$. Abusing notation, we let $\langle R(\mathfrak{C})_V \rangle$ also denote the restriction of the congruence $\langle R(\mathfrak{C})_V \rangle$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$ to the subcategory $F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}_{\mathfrak{D},\mathfrak{J}}^{-1}])$, as in Example A.7 (5)(i). The *(universal parameter) category presented by the restricted coherence schema* $(\mathfrak{C}, V, \mathfrak{D})$ is then the quotient category

$$\mathscr{P}(\mathfrak{C}, V, \mathfrak{D}) \coloneqq \frac{F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}_{\mathfrak{D},\mathfrak{J}}^{-1}])}{\langle R(\mathfrak{C})_V \rangle},$$

a subcategory of $\mathscr{P}_V(\mathfrak{C})$. When $V = V(\mathfrak{D})$, we denote this quotient category simply by $\mathscr{P}(\mathfrak{C}, \mathfrak{D})$. If $W \subseteq V$ with $\mathfrak{D} \subseteq \mathcal{L}(\mathfrak{S})_W$, there is an inclusion $R(\mathfrak{C})_W \subseteq R(\mathfrak{C})_V$ (as explicated in Remark 2.48), so there is a quotient functor $\mathscr{P}(\mathfrak{C}, W, \mathfrak{D}) \to \mathscr{P}(\mathfrak{C}, V, \mathfrak{D})$.

[10] As in the argument from Remark 2.48, it is possible that a path in the graph $\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}]$ between two vertices (terms) in $\mathfrak{D}$ passes through vertices that are not in $\mathfrak{D}$. Such a path would be a morphism of $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$ but not a morphism of $F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}_{\mathfrak{D},\mathfrak{J}}^{-1}])$, even though its domain and codomain are in $\mathfrak{D}$.

Restricting $\langle R(\mathfrak{C})_V\rangle^{\mathrm{L}}$ to $F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{J}}^{-1}])$ likewise yields a congruence that is also left-cancelable by Lemma A.14. Then the *monomorphic (universal parameter) category presented by the restricted coherence schema* $(\mathfrak{C}, V, \mathfrak{O})$ is the quotient category

$$\mathscr{P}^{\mathrm{L}}(\mathfrak{C}, V, \mathfrak{O}) \coloneqq \frac{F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{J}}^{-1}])}{\langle R(\mathfrak{C})_V\rangle^{\mathrm{L}}},$$

a subcategory of $\mathscr{P}_V^{\mathrm{L}}(\mathfrak{C})$ and a quotient category of $\mathscr{P}(\mathfrak{C}, V, \mathfrak{O})$. When $V = V(\mathfrak{O})$, we denote this quotient category simply by $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$. For $W \subseteq V$ such that $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})_V$ there is an induced functor $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}, W, \mathfrak{O}) \to \mathscr{P}^{\mathrm{L}}(\mathfrak{C}, V, \mathfrak{O})$.

Finally, to consider all variables for $\mathfrak{S}$, define

$$\mathscr{P}(\mathfrak{C}) \coloneqq \frac{F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{J}}^{-1}])}{\langle R(\mathfrak{C})\rangle} \quad \text{and} \quad \mathscr{P}^{\mathrm{L}}(\mathfrak{C}) \coloneqq \frac{F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{J}}^{-1}])}{\langle R(\mathfrak{C})\rangle^{\mathrm{L}}},$$

the *(universal parameter) category presented by the coherence schema* $\mathfrak{C}$ and the *monomorphic (universal parameter) category presented by the coherence schema* $\mathfrak{C}$, respectively, with a quotient functor $\mathscr{P}(\mathfrak{C}) \to \mathscr{P}^{\mathrm{L}}(\mathfrak{C})$. Since $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$ embeds as a (not necessarily full) subcategory of $F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{J}}^{-1}])$, and $R(\mathfrak{C})_V$ is contained in the restriction of $R(\mathfrak{C})$ to $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{J}}^{-1}])$, there are canonical functors $\mathscr{P}_V(\mathfrak{C}) \to \mathscr{P}(\mathfrak{C})$ and $\mathscr{P}_V^{\mathrm{L}}(\mathfrak{C}) \to \mathscr{P}^{\mathrm{L}}(\mathfrak{C})$.

2.4 Models of coherence schemata

Now that we have formally parametrized our algebraic categories, we may construct models of them by interpreting the formal symbols in these parametrizations. These models will, of course, be categories or families of categories: objects or families of objects in the strict 2-category $\underline{\mathbf{Cat}}$. But our work will generalize to any strict 2-category with finite 2-products.

Definition 2.52 (Cartesian monoidal 2-category)**.** A *cartesian monoidal* 2*-category* is a strict 2-category (i.e. strictly **Cat**-enriched category) $\underline{\mathbf{K}}$ equipped with all finite 2-products. That is, for any $n \geq 0$ and objects $\mathscr{C}_1, \ldots, \mathscr{C}_n \in \underline{\mathbf{K}}$, there is an object $\mathscr{C}_1 \times \cdots \times \mathscr{C}_n \in \underline{\mathbf{K}}$, together with projection 1-morphisms $\pi_i \colon \mathscr{C}_1 \times \cdots \times \mathscr{C}_n \to \mathscr{C}_i$ in $\underline{\mathbf{K}}$ for $1 \leq i \leq n$, such that for any object $\mathscr{D} \in \underline{\mathbf{K}}$, horizontal composition with the projections induces an isomorphism

$$\langle (\pi_1)_*, \ldots, (\pi_n)_* \rangle \colon \underline{\mathbf{K}}(\mathscr{D}, \mathscr{C}_1 \times \cdots \times \mathscr{C}_n) \xrightarrow{\cong} \underline{\mathbf{K}}(\mathscr{D}, \mathscr{C}_1) \times \cdots \times \underline{\mathbf{K}}(\mathscr{D}, \mathscr{C}_n)$$

of hom 1-categories. When $n \coloneqq 0$, we denote the empty product by $*$, and we have an isomorphism $\underline{\mathbf{K}}(\mathscr{D}, *) \cong *$ between $\underline{\mathbf{K}}(\mathscr{D}, *)$ and the terminal category, also denoted by $*$.

Throughout the remainder of this chapter, we fix a cartesian monoidal 2-category $\underline{\mathbf{K}}$. Typically we set $\underline{\mathbf{K}} := \underline{\mathbf{Cat}}$, the strict 2-category of (small) categories, functors, and natural transformations, equipped with the usual cartesian product. So when we give semantics to the signature $\mathfrak{S}$ of our coherence schema, we want to interpret its types in T as small categories, its constants in C as objects in those small categories with compatible types as specified by τ, and its function symbols in F as functors between these small categories with compatible domains and codomains as specified by dom and cod. We generalize this to $\underline{\mathbf{K}}$ below.

Definition 2.53 (Model of signature)**.** A *model of* $\mathfrak{S}$ *in* $\underline{\mathbf{K}}$ consists of the following data:

(1) an object $|\mathtt{t}| \in \underline{\mathbf{K}}$ for each type $\mathtt{t} \in T$ (i.e. an assignment $|\cdot|\colon T \to \mathrm{Ob}(\underline{\mathbf{K}})$);

(2) a 1-morphism $|c|\colon * \to |\tau(c)|$ in $\underline{\mathbf{K}}$ for each constant $c \in C$;[11]

(3) a 1-morphism $|f|\colon |\mathtt{s}_1| \times \cdots \times |\mathtt{s}_n| \to |\mathtt{t}|$ in $\underline{\mathbf{K}}$ for each n-ary function symbol $f \in F$ with $\mathrm{dom}(f) = (\mathtt{s}_1, \ldots, \mathtt{s}_n)$ and $\mathrm{cod}(f) = \mathtt{t}$. For ease of notation, we let $|(\mathtt{s}_1, \ldots, \mathtt{s}_n)| := |\mathtt{s}_1| \times \cdots \times |\mathtt{s}_n|$, so that $|\mathrm{dom}(f)| = \mathrm{dom}(|f|)$.

We write $|\cdot|\colon \mathfrak{S} \to \underline{\mathbf{K}}$ to denote a model of $\mathfrak{S}$ in $\underline{\mathbf{K}}$.

Example 2.54 (Model of signature for monoidal categories)**.** A model $(\mathscr{C}, e^{\mathscr{C}}, \otimes^{\mathscr{C}})$ of the signature $\mathfrak{S}_{\mathrm{M}}$ for monoidal categories from Example 2.2 in $\underline{\mathbf{Cat}}$ consists of:

(1) a (small) category $\mathscr{C} := |\mathtt{m}| \in \underline{\mathbf{Cat}}$;

(2) a distinguished object $e^{\mathscr{C}} := |e| \in \mathscr{C}$ called the *monoidal unit*; and

(3) a functor $\otimes^{\mathscr{C}} := |\otimes|\colon \mathscr{C} \times \mathscr{C} \to \mathscr{C}$, which we write using infix notation, called the *monoidal product*.

Example 2.55 (Model of signature for bicategories)**.** A model $\mathscr{B}$ of the signature $\mathfrak{S}_{\mathrm{B}}(O)$ for bicategories with object-set O from Example 2.3 in $\underline{\mathbf{Cat}}$ consists of:

[11] When $\underline{\mathbf{K}} := \underline{\mathbf{Cat}}$, this is equivalent to an object in the category $|\tau(c)|$.

(1) a (small) category $\mathscr{B}(a,b) \coloneqq |(a,b)| \in \underline{\mathbf{Cat}}$ for all $a,b \in O$, whose objects are 1-*cells* $a \to b$ and whose morphisms from $f\colon a \to b$ to $g\colon a \to b$ are 2-*cells* $f \Rightarrow g$;[12]

(2) a distinguished 1-cell $\mathbf{1}_a^{\mathscr{B}} \coloneqq |\mathbf{1}_a| \in \mathscr{B}(a,a)$ for $a \in O$ called the *identity* 1-*cell on* a;

(3) a functor $\circ \coloneqq |\circ_{a,b,c}|\colon \mathscr{B}(b,c) \times \mathscr{B}(a,b) \to \mathscr{B}(a,c)$ for $a,b,c \in O$, which we write using infix notation, called *horizontal composition*.

For the rest of the chapter, fix a model $|\cdot|\colon \mathfrak{S} \to \underline{\mathbf{K}}$. The model lets us interpret the constant and function symbols of $\mathfrak{S}$ as 1-morphisms in $\underline{\mathbf{K}}$. We would like to extend this interpretation to any term in $\mathcal{L}(\mathfrak{S})$ that is built from these symbols. Crucially, we use scopes to fix domains: without specifying a scope for a term, the interpretation of its domain would not be well-defined.

Definition 2.56 (Model of language, operation)**.** We shall extend $|\cdot|\colon \mathfrak{S} \to \underline{\mathbf{K}}$ to $\mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$ from Definition 2.7 to a function

$$|\cdot|\colon \mathcal{L}(\mathfrak{S})^{\mathrm{sc}} \to \coprod_{n\in\mathbb{N}} \coprod_{(\mathtt{t}_0,\dots,\mathtt{t}_n)\in T^{n+1}} \underline{\mathbf{K}}(|\mathtt{t}_1| \times \cdots \times |\mathtt{t}_n|, |\mathtt{t}_0|)$$

The function takes as input some $(t,V) \in \mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$, so that every variable appearing in t is in V. Then $|(t,V)|$ is a 1-morphism in $\underline{\mathbf{K}}$. The domain of $|(t,V)|$ is a product of objects in $\underline{\mathbf{K}}$ of the form $|\mathtt{t}|$, as given by the model, for some $t \in T$; while the codomain of $|(t,V)|$ will always be $|\tau(t)|$. That is, given a fixed $t \in \mathcal{L}(\mathfrak{S})$, we have a commutative square

$$\begin{tikzcd}
\{t\} \times \{V \mid V \text{ is a scope for } t\} \arrow[r] \arrow[d, hook] & \displaystyle\coprod_{n\in\mathbb{N}} \coprod_{(\mathtt{t}_1,\dots,\mathtt{t}_n)\in T^{n}} \underline{\mathbf{K}}(|\mathtt{t}_1| \times \cdots \times |\mathtt{t}_n|, |\tau(t)|) \arrow[d, hook] \\
\mathcal{L}(\mathfrak{S})^{\mathrm{sc}} \arrow[r, "|\cdot|"] & \displaystyle\coprod_{n\in\mathbb{N}} \coprod_{(\mathtt{t}_0,\dots,\mathtt{t}_n)\in T^{n+1}} \underline{\mathbf{K}}(|\mathtt{t}_1| \times \cdots \times |\mathtt{t}_n|, |\mathtt{t}_0|).
\end{tikzcd}$$

We now construct the domain of $|(t,V)|$ before defining $|(t,V)|$ itself. Given a finite set of variables $V = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}$ for $\mathfrak{S}$, define

$$|V| = \left|\{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}\right| \coloneqq |(\mathtt{t}_1,\dots,\mathtt{t}_n)| \coloneqq |\mathtt{t}_1| \times \cdots \times |\mathtt{t}_n| \in \underline{\mathbf{K}}. \tag{2.57}$$

[12]In keeping with the pattern, we sometimes call the objects in O the *0-cells* of $\mathscr{B}$.

to be the domain of $|(t,V)|$, so that we have a commutative diagram

$$\begin{tikzcd}
\{t\} \arrow[rr] \arrow[d, hook] & & \underline{\mathbf{K}}(|V|, |\tau(t)|) \arrow[d, hook] \\
\mathcal{L}(\mathfrak{S})_V = \{s \in \mathcal{L}(\mathfrak{S}) \mid (s,V) \in \mathcal{L}(\mathfrak{S})^{\mathrm{sc}}\} \arrow[rr, "|\cdot|_V"] \arrow[d, hook, "(\cdot, V)"] & & \coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|) \arrow[d, hook] \\
\mathcal{L}(\mathfrak{S})^{\mathrm{sc}} \arrow[rr, "|\cdot|"'] & & \coprod_{n\in\mathbb{N}} \coprod_{(\mathtt{t}_0,\dots,\mathtt{t}_n)\in T^{n+1}} \underline{\mathbf{K}}(|\mathtt{t}_1| \times \cdots \times |\mathtt{t}_n|, |\mathtt{t}_0|)
\end{tikzcd}$$

where we name the middle horizontal arrow $|\cdot|_V$.

We now define $|\cdot|$ recursively on elements of $\mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$ with second coordinate V to be 1-morphisms in $\underline{\mathbf{K}}$ with domain $|V|$ as follows.

(1) For each variable $x_{i_k}^{\mathtt{t}_k} \in V$, let $\left|(x_{i_k}^{\mathtt{t}_k}, V)\right| : |V| \to \left|\tau(x_{i_k}^{\mathtt{t}_k})\right| = |\mathtt{t}_k|$ be the projection onto the k^{th} factor of $|V|$.

(2) For each constant $c \in C$, let $|(c,V)| : |V| \to |\tau(c)|$ be the constant 1-morphism given by the composite

$$|V| \xrightarrow{!} * \xrightarrow{|c|} |\tau(c)|.$$

(3) For each function symbol $f \in F$ and terms $s_1, \dots, s_n \in \mathcal{L}(\mathfrak{S})$ with $\operatorname{dom}(f) = (\tau(s_1), \dots, \tau(s_n))$, let $|(f(s_1,\dots,s_n),V)| : |V| \to |\operatorname{cod}(f)|$ be the composite

$$|V| \xrightarrow{\langle |(s_1,V)|, \dots, |(s_n,V)| \rangle} |\tau(s_1)| \times \cdots \times |\tau(s_n)| \xrightarrow{|f|} |\operatorname{cod}(f)|.$$

As shorthand, for $(t,V) \in \mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$ and $(s_1,V), \dots, (s_n,V) \in \mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$ with $\mathbf{s} = (s_1, \dots, s_n)$, we write $|t|_V := |(t,V)|$ and $|\mathbf{s}|_V := \langle |(s_1,V)|, \dots, |(s_n,V)| \rangle$, so we may write the previously displayed composite as $|f(s_1,\dots,s_n)|_V = |f| \circ |\mathbf{s}|_V$. We have now constructed a function

$$|\cdot|_V : \mathcal{L}(\mathfrak{S})_V \to \coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|), \tag{2.58}$$

where the domain is the set of terms containing only variables from V, as in Definition 2.7, and the codomain is the underlying set of 1-morphisms, satisfying $\operatorname{cod}(|t|_V) = |\tau(t)|$.

For all $(t, V) \in \mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$, we call the 1-morphism $|t|_V \colon |V| \Rightarrow |\mathtt{t}|$ of $\underline{\mathbf{K}}$ an *operation* for the model $|\cdot| \colon \mathfrak{S} \to \underline{\mathbf{K}}$.

Finally, if $t \in \mathcal{L}(\mathfrak{S})$ and no scope is specified, define $|t| \coloneqq |t|_{V(t)}$.[13]

Example 2.59. Continuing from Example 2.54, take $V_k = \{x_1 < \cdots < x_k\}$ from Example 2.8, so that $|V_k| = \mathscr{C}^k$. Then V_k is a scope for the term $t \coloneqq \otimes(x_3, \otimes(x_1, e))$ whenever $k \geq 3$, with the operation $|t|_{V_k}$ equal to the functor $\mathscr{C}^k \to \mathscr{C}$ sending $(c_1, c_2, c_3, \ldots, c_k) \mapsto c_3 \otimes^{\mathscr{C}} (c_1 \otimes^{\mathscr{C}} e^{\mathscr{C}})$.

Example 2.60. We similarly build on Example 2.55, suppressing some subscripts for clarity. Fix $a, b \in O$ and let $x_1 \coloneqq x_1^{(b,a)}$ and $x_2 \coloneqq x_2^{(a,b)}$. By the order on variables, $x_1 < x_2$, so

$$|\circ(x_1, \circ(x_2, \mathbf{1}_a))| \tag{2.61}$$

is a functor with domain

$$|\{x_1 < x_2\}| = \mathscr{B}(b, a) \times \mathscr{B}(a, b)$$

and codomain $\mathscr{B}(a, a)$ sending $(f, g) \mapsto f \circ (g \circ \mathbf{1}_a^{\mathscr{B}})$.

We may characterize the construction of (2.58) in Definition 2.56 as follows.

Proposition 2.62. *Let $|\cdot| \colon \mathfrak{S} \to \underline{\mathbf{K}}$ be a model of the signature $\mathfrak{S}$ in the cartesian monoidal 2-category $\underline{\mathbf{K}}$, as defined in Definition 2.53, and let V be a finite set of variables for $\mathfrak{S}$, with $|V|$ defined as in (2.57). Then there is a unique function*

$$|\cdot|_V \colon \mathcal{L}(\mathfrak{S})_V \to \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|), \tag{2.63}$$

where the codomain is the underlying set of 1-morphisms, such that

(1) $\left|x_k^{\mathtt{t}}\right|_V$ is the projection $|V| \to |\mathtt{t}|$ to the $x_k^{\mathtt{t}}$-factor, with $|\mathtt{t}|$ given by the model, for each variable $x_k^{\mathtt{t}} \in V$;

(2) $|c|_V$ is the constant morphism $|V| \to * \xrightarrow{|c|} |\tau(c)|$, with $*$ the terminal object of $\underline{\mathbf{K}}$ and $|c| \colon * \to |\tau(c)|$ given by the model, for each constant $c \in C$; and

[13] Note that this notation is compatible with our previous definition of $|\cdot|$ on constants $c \in C$, as $V(c) = \varnothing$ and $|\varnothing| = *$.

(3) $|f(s_1,\ldots,s_n)|_V = |f| \circ |\mathbf{s}|_V$ *for an n-ary function symbol $f \in F$ and n-tuple of terms $\mathbf{s} = (s_1,\ldots,s_n)$ in $\mathcal{L}(\mathfrak{S})_V$ for which* $\mathrm{dom}(f) = (\tau(s_1),\ldots,\tau(s_n))$*, with $|f|$ as given by the model and $|\mathbf{s}|_V = \langle |s_1|_V,\ldots,|s_n|_V\rangle$.*

We also have $\mathrm{cod}(|t|_V) = |\tau(t)|$ *for all $t \in \mathcal{L}(\mathfrak{S})$.*

More generally, given a variable-finite, subterm-closed subset $\mathfrak{D} \subseteq \mathcal{L}(\mathfrak{S})_V$, there is also a unique function

$$|\cdot|_V \colon \mathfrak{D} \to \coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|) \tag{2.64}$$

satisfying (1) for every variable in $\mathfrak{D}$, (2) for every constant in $\mathfrak{D}$, and (3) for every other term in $\mathfrak{D}$. It also satisfies $\mathrm{cod}(|t|_V) = |\tau(t)|$ *for all $t \in \mathfrak{D}$.*

Proof. The function (2.63) satisfying the given conditions is precisely our construction (2.58) from Definition 2.56; well-definedness and uniqueness follow from induction and the Parsing Lemma. Then the function (2.64) is obtained by restricting the domain of (2.63) to $\mathfrak{D}$. Here uniqueness is given by induction on the terms in $\mathfrak{D}$, with subterm-closure ensuring the inductive step works. □

With this definition of term interpretation, substitutions are interpreted as horizontal composition, as the following lemma demonstrates (see also Lemma 2.80).

Lemma 2.65. *Fix a term $t \in \mathcal{L}(\mathfrak{S})$ with scope $W = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}$ and terms $s_1,\ldots,s_n \in \mathcal{L}(\mathfrak{S})$ with common scope V such that $\tau(s_j) = \mathtt{t}_j$ for all $1 \le j \le n$. Let $\mathbf{s} \coloneqq (s_1,\ldots,s_n)$ so that $t[W/\mathbf{s}]$ is the term obtained by substituting s_j for each $x_{i_j}^{\mathtt{t}_j}$ in t for $1 \le j \le n$. Then $|t[W/\mathbf{s}]|_V$ is the composite 1-morphism*

$$|V| \xrightarrow{\langle |s_1|_V,\ldots,|s_n|_V\rangle} |(\mathtt{t}_1,\ldots,\mathtt{t}_n)| = |W| \xrightarrow{|t|_W} |\tau(t)|, \tag{2.66}$$

denoted more compactly by $|t|_W \circ |\mathbf{s}|_V$.

Proof. Induct on t. If t is a constant c, substitution does nothing, and terminality of $*$ means $|c|_V = |c| \circ {!} \colon |V| \to * \to |\tau(c)|$ is equal to the composite of $|c|_W = |c| \circ {!} \colon |W| \to * \to |\tau(c)|$ with any 1-morphism $|V| \to |W|$.

If t is a variable $x_{i_j}^{\mathtt{t}_j} \in W$, then s_j is the term with scope V and type $\tau(s_j) = \mathtt{t}_j$ that we substitute for $x_{i_j}^{\mathtt{t}_j}$ to obtain $x_{i_j}^{\mathtt{t}_j}[W/\mathbf{s}] = s_j$. Since $\left|x_{i_j}^{\mathtt{t}_j}\right|_W$ is the j^{th} projection $\pi_j \colon |W| \to |\mathtt{t}_j|$, the composite (2.66) when $t = x_{i_j}^{\mathtt{t}_j}$ given by $\left|x_{i_j}^{\mathtt{t}_j}\right|_W \circ |\mathbf{s}|_V = \pi_j \circ \langle |s_1|_V, \ldots, |s_n|_V \rangle$ is $|s_j|_V$.

Finally, we let $t = f(t_1, \ldots, t_m)$ for $f \in F$ and $t_1, \ldots, t_m \in \mathcal{L}(\mathfrak{S})$ with $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_m))$. Since W is a scope for t, it is also a scope for each t_i. Moreover the first symbol of $t[W/\mathbf{s}]$ is also f, and by the Parsing Lemma $t[W/\mathbf{s}] = f(t_1[W/\mathbf{s}], \ldots, t_m[W/\mathbf{s}])$. Then $|t[W/\mathbf{s}]|_V$ is the composite operation

$$|V| \xrightarrow{\langle |t_1[W/\mathbf{s}]|_V, \ldots, |t_m[W/\mathbf{s}]|_V \rangle} |\tau(t_1)| \times \cdots \times |\tau(t_m)| \xrightarrow{|f|} |\mathrm{cod}(f)|,$$

and by induction each $|t_i[W/\mathbf{s}]|_V$ is given by the composite operation

$$|V| \xrightarrow{\langle |s_1|_V, \ldots, |s_n|_V \rangle} |\mathtt{t}_1| \times \cdots \times |\mathtt{t}_n| = |W| \xrightarrow{|t_i|_W} |\tau(t_i)|.$$

Hence $|t[W/\mathbf{s}]|_V$ factors through $|W|$ as

$$|V| \xrightarrow{\langle |s_1|_V, \ldots, |s_n|_V \rangle} |W| \xrightarrow{\langle |t_1|_W \cdots, |t_m|_W \rangle} |\tau(t_1)| \times \cdots \times |\tau(t_m)| \xrightarrow{|f|} |\mathrm{cod}(f)|.$$

The middle and right arrows compose to $|t|_W$, so we are done. $\square$

It follows that expanding the scope behaves as expected.

Corollary 2.67. *If $(t, V), (t, W) \in \mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$ such that $W \subseteq V$, then $|t|_V = |t|_W \circ \pi$, where $\pi \colon |V| \to |W|$ is the canonical projection.*

Proof. Write $W = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}$. Now the result follows from Lemma 2.65 with $s_j := x_{i_j}^{\mathtt{t}_j}$ for all $1 \leq j \leq n$, since substituting each $x_{i_j}^{\mathtt{t}_j}$ for itself yields $t[W/\mathbf{s}] = t$ itself, while $|\mathbf{s}|_V = \langle \left|x_{i_1}^{\mathtt{t}_1}\right|_V, \ldots, \left|x_{i_n}^{\mathtt{t}_n}\right|_V \rangle$ is precisely the projection $|V| \to |W|$. $\square$

So for $t \in \mathcal{L}(\mathfrak{S})$ whose constituent variables comprise the set $V(t)$, if $(t, V) \in \mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$, we may always write $|t|_V$ in terms of $|t| = |t|_{V(t)}$: we have $|t|_V = |t| \circ \pi$ for the canonical projection $\pi \colon |V| \to |V(t)|$. In other words, if $V(\mathfrak{D})$ is the set of all variables appearing in a variable-finite, subterm-closed subset $\mathfrak{D} \subset \mathcal{L}(\mathfrak{S})$, then for all finite sets of variables

$V \supseteq V(\mathfrak{O})$, the functions $|\cdot|_V$ and $|\cdot|_{V(\mathfrak{O})}$ from (2.64) assemble into a commutative triangle

$$\begin{tikzcd} \mathfrak{O} \arrow[r, "|\cdot|_{V(\mathfrak{O})}"] \arrow[dr, "|\cdot|_V"'] & \coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V(\mathfrak{O})|, |\mathtt{t}|) \arrow[d] \\ & \coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|), \end{tikzcd} \tag{2.68}$$

where the vertical map is given by precomposition with the projection $\pi\colon |V| \to |V(\mathfrak{O})|$.

We may now extend our model $|\cdot|$ from Definition 2.56 of the language of our signature $\mathfrak{S}$ to our transformation schema $\mathfrak{T}\colon \mathcal{E} \to \mathcal{L}(\mathfrak{S}) \times_T \mathcal{L}(\mathfrak{S})$ with marked edges $\mathfrak{I} \subseteq \mathcal{E}$.

Definition 2.69 (Model of marked transformation schema)**.** A *model of* $(\mathfrak{S}, \mathfrak{T}, \mathfrak{I})$ *in* $\underline{\mathbf{K}}$ is a model $|\cdot|\colon \mathfrak{S} \to \underline{\mathbf{K}}$ of the signature, extended to $\mathcal{L}(\mathfrak{S})^{\mathrm{sc}}$ as in Definition 2.56, together with

(4) a 2-morphism $|\varepsilon|\colon |\mathrm{dom}(\varepsilon)|_{V(\varepsilon)} \Rightarrow |\mathrm{cod}(\varepsilon)|_{V(\varepsilon)}$ for every $\varepsilon \in \mathcal{E}$,

such that $|\varepsilon|$ is a vertical 2-isomorphism whenever $\varepsilon \in \mathfrak{I}$.

To ensure this definition is not vacuous, we verify that $|\mathrm{dom}(\varepsilon)|_{V(\varepsilon)}$ and $|\mathrm{cod}(\varepsilon)|_{V(\varepsilon)}$ have the same domain and codomain. Indeed, they have the same scope $V(\varepsilon)$ and therefore the same domain $|V(\varepsilon)|$; and they have the same codomain $|\tau(\mathrm{dom}(\varepsilon))| = |\tau(\varepsilon)| = |\tau(\mathrm{cod}(\varepsilon))|$.

Example 2.70 (Model of transformation schemata for monoidal categories)**.** We extend our model $(\mathscr{C}, e^{\mathscr{C}}, \otimes^{\mathscr{C}})$ from Example 2.54 of the signature $\mathfrak{S}_{\mathrm{M}}$ for monoidal categories in $\underline{\mathbf{Cat}}$ to our transformation schemata $\mathfrak{T}_{\mathrm{M}}\colon \{\alpha, \lambda, \rho\} \to \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$ for monoidal categories and $\mathfrak{T}_{\mathrm{SM}}\colon \mathcal{E}_{\mathrm{SM}} = \{\alpha, \beta, \lambda, \rho\} \to \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$ for symmetric monoidal categories from Example 2.27, potentially with all edges marked as in Example 2.41.

A model of $(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing)$ in $\underline{\mathbf{Cat}}$ consists of $(\mathscr{C}, e^{\mathscr{C}}, \otimes^{\mathscr{C}})$ from Example 2.54 along with the following three natural transformations, one for each edge label in $\mathcal{E}_{\mathrm{M}} = \{\alpha, \lambda, \rho\}$:

- a natural transformation $\alpha^{\mathscr{C}} \coloneqq |\alpha|\colon |\otimes(\otimes(x_1, x_2), x_3)|_{V_3} \Rightarrow |\otimes(x_1, \otimes(x_2, x_3))|_{V_3}$, i.e. one between functors $\mathscr{C}^3 \to \mathscr{C}$ sending

$$(c_1, c_2, c_3) \mapsto (c_1 \otimes^{\mathscr{C}} c_2) \otimes^{\mathscr{C}} c_3 \quad \text{and} \quad (c_1, c_2, c_3) \mapsto c_1 \otimes^{\mathscr{C}} (c_2 \otimes^{\mathscr{C}} c_3),$$

  that we call the *associator*;

- a natural transformation $\lambda^{\mathscr{C}} \coloneqq |\lambda|\colon |\otimes(e, x_1)|_{V_1} \Rightarrow |x_1|_{V_1}$, i.e. one between functors $\mathscr{C} \to \mathscr{C}$ sending

$$c_1 \mapsto e^{\mathscr{C}} \otimes^{\mathscr{C}} c_1 \quad \text{and} \quad c_1 \mapsto c_1,$$

  that we call the *left unitor*; and

- a natural transformation $\rho^{\mathscr{C}} \coloneqq |\rho|\colon |\otimes(x_1, e)|_{V_1} \Rightarrow |x_1|_{V_1}$, i.e. one between functors $\mathscr{C} \to \mathscr{C}$ sending

$$c_1 \mapsto c_1 \otimes^{\mathscr{C}} e^{\mathscr{C}} \quad \text{and} \quad c_1 \mapsto c_1,$$

  that we call the *right unitor*.

A model of $(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \mathcal{E}_{\mathrm{M}})$ in $\underline{\mathbf{Cat}}$ is then a model of $(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing)$ in $\underline{\mathbf{Cat}}$ for which the associator and both unitors are required to be natural isomorphisms (i.e. invertible).

Finally, a model of $(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{SM}}, \mathcal{E}_{\mathrm{M}})$ in $\underline{\mathbf{Cat}}$ is a model of $(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \mathcal{E}_{\mathrm{M}})$ in $\underline{\mathbf{Cat}}$ equipped with one more natural transformation for the additional edge label β in $\mathcal{E}_{\mathrm{SM}} = \{\alpha, \lambda, \rho, \beta\}$:

- a natural transformation $\beta^{\mathscr{C}} \coloneqq |\beta|\colon |\otimes(x_1, x_2)|_{V_2} \Rightarrow |\otimes(x_2, x_1)|_{V_2}$, i.e. one between functors $\mathscr{C}^2 \to \mathscr{C}$ sending

$$(c_1, c_2) \mapsto c_1 \otimes^{\mathscr{C}} c_2 \quad \text{and} \quad (c_1, c_2) \mapsto c_2 \otimes^{\mathscr{C}} c_1.$$

  that we call the *braiding*.

Example 2.71 (Model of transformation schema for bicategories)**.** Similarly, we may extend our model $\mathscr{B}$ from Example 2.55 of the signature $\mathfrak{S}_{\mathrm{B}}(O)$ for bicategories in $\underline{\mathbf{Cat}}$ to our transformation schema $\mathfrak{T}_{\mathrm{B}}(O)\colon \mathcal{E}_{\mathrm{B}}(O) \to \mathcal{L}^{\mathrm{B}}(O) \times_{(O \times O)} \mathcal{L}^{\mathrm{B}}(O)$ for bicategories with object-set O from Example 2.28, potentially with all edges marked as in Example 2.41.

A model of $(\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \varnothing)$ in $\underline{\mathbf{Cat}}$ generalizes a model of $(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing)$ from Example 2.70, with families of associator and unitor natural transformations consisting of one transformation for each edge label in $\mathcal{E}_{\mathrm{B}}(O)$. A model of $(\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \mathcal{E}_{\mathrm{B}}(O))$ in $\underline{\mathbf{Cat}}$ is then a model of $(\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \varnothing)$ in $\underline{\mathbf{Cat}}$ for which the associators and unitors are required to be natural isomorphisms.

Henceforth we fix a model $|\cdot|\colon (\mathfrak{S},\mathfrak{T},\mathfrak{I}) \to \underline{\mathbf{K}}$ of our marked transformation schema. To extend this model to a model of our coherence schema, we must first extend $|\cdot|$ to atomic and basic transformations. Again, we must take care to ensure that the domains of the interpretations of these transformations are well-defined, and we do so by fixing a scope and restricting the terms under consideration. We let V be a finite set of variables and let $\mathcal{L}(\mathfrak{S})_V \subset \mathcal{L}(\mathfrak{S})$ be the subset of terms containing only variables from V, as in Definition 2.7.

Definition 2.72 (Model of basic transformations)**.** In Definition 2.56, we constructed a function (2.58) on terms with scope V denoted

$$|\cdot|_V\colon \mathcal{L}(\mathfrak{S})_V \to \coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V|,|\mathtt{t}|)$$

with $\mathrm{cod}(|t|_V) = |\tau(t)|$. We shall extend this function to basic transformations, sending each basic transformation φ between terms in $\mathcal{L}(\mathfrak{S})_V$ to a 2-morphism between 1-morphisms $|\mathrm{dom}(\varphi)|_V, |\mathrm{cod}(\varphi)|_V\colon |V| \rightrightarrows |\tau(\varphi)|$ (recall from Definition 2.32 that $\tau(\varphi) = \tau(\mathrm{dom}(\varphi)) = \tau(\mathrm{cod}(\varphi))$). We do so recursively, starting with atomic transformations.

(1) From Definition 2.69, we already have a 2-morphism $|\varepsilon|\colon |\mathrm{dom}(\varepsilon)|_{V(\varepsilon)} \Rightarrow |\mathrm{cod}(\varepsilon)|_{V(\varepsilon)}$ between 1-morphisms $|V(\varepsilon)| \rightrightarrows |\tau(\varepsilon)|$ for every edge label $\varepsilon \in \mathcal{E}$. Then given an atomic transformation $\varepsilon_{\mathbf{s}}\colon \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}] \Rightarrow \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ with $\mathbf{s} = (s_1,\dots,s_n)$ of appropriate types such that $s_i \in \mathcal{L}(\mathfrak{S})_V$ for all i, we define $|\varepsilon_{\mathbf{s}}|_V$ to be the whiskered 2-morphism

$$\begin{tikzcd}[column sep=large] |V| \arrow[r, "|\mathbf{s}|_V"] & |\tau(s_1)|\times\cdots\times|\tau(s_n)| = |V(\varepsilon)| \arrow[rr, bend left, "|\mathrm{dom}(\varepsilon)|_{V(\varepsilon)}"] \arrow[rr, bend right, "|\mathrm{cod}(\varepsilon)|_{V(\varepsilon)}"'] \arrow[rr, phantom, "\Downarrow|\varepsilon|"] & & |\tau(\varepsilon)|, \end{tikzcd}$$

or $|\varepsilon|_{|\mathbf{s}|_V}$ for short. Lemma 2.65 ensures the domain and codomain of this 2-morphism are given by $|\mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}]|_V$ and $|\mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]|_V$, respectively, and are thus compatible with $|\cdot|_V$.

(2) Then recursively, given terms $r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n \in \mathcal{L}(\mathfrak{S})_V$, a basic transformation φ between two terms in $\mathcal{L}(\mathfrak{S})_V$, and an n-ary function symbol $f \in F$ with $\mathrm{dom}(f) = (\tau(r_1), \ldots, \underbrace{\tau(\varphi)}_{k^{\text{th}} \text{ term}}, \ldots, \tau(r_n))$, let $\left|f(\mathrm{id}_{r_1}, \ldots, \mathrm{id}_{r_{k-1}}, \varphi, \mathrm{id}_{r_{k+1}}, \ldots, \mathrm{id}_{r_n})\right|_V$ be the whiskered 2-morphism[14]

$$\begin{tikzcd}
{|V|} \arrow[rr, bend left, "{\langle |r_1|_V, \ldots, |r_{k-1}|_V, |\mathrm{dom}(\varphi)|_V, |r_{k+1}|_V, \ldots, |r_n|_V \rangle}"] \arrow[rr, bend right, "{\langle |r_1|_V, \ldots, |r_{k-1}|_V, |\mathrm{cod}(\varphi)|_V, |r_{k+1}|_V, \ldots, |r_n|_V \rangle}"'] \arrow[rr, phantom, "\Downarrow"] & & {|\mathrm{dom}(f)|} \arrow[r, "|f|"] & {|\mathrm{cod}(f)|,}
\end{tikzcd}$$

where the unlabeled 2-morphism is

$$\langle \mathrm{id}_{|r_1|_V}, \ldots, \mathrm{id}_{|r_{k-1}|_V}, |\varphi|_V, \mathrm{id}_{|r_{k+1}|_V}, \ldots, \mathrm{id}_{|r_n|_V} \rangle$$

given by the universal property of 2-products in $\underline{\mathbf{K}}$ (see Definition 2.52). As shorthand, omitting subscripts, we write the whiskered 2-morphism as $|f|\langle \mathrm{id}, \ldots, |\varphi|_V, \ldots, \mathrm{id} \rangle$. Definition 2.56 ensures the domain and codomain of this whiskered 2-morphism are

$$|f(r_1, \ldots, r_{k-1}, \mathrm{dom}(\varphi), r_{k+1}, \ldots, r_n)|_V \text{ and } |f(r_1, \ldots, r_{k-1}, \mathrm{cod}(\varphi), r_{k+1}, \ldots, r_n)|_V,$$

respectively, and are thus compatible with $|\cdot|_V$.

This extends $|\cdot|_V$ to the set of all edges Bs_V in the full subgraph $\mathscr{B}_V \subset \mathscr{B}(\mathfrak{S}, \mathfrak{T})$ spanned by the vertices in $\mathcal{L}(\mathfrak{S})_V$, making $|\cdot|_V$ a graph morphism

$$|\cdot|_V \colon \mathscr{B}_V \to \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|),$$

where the codomain is the underlying graph whose vertices are 1-morphisms and whose edges are 2-morphisms. We can therefore extend $|\cdot|_V$ to a functor out of the free category on $\mathscr{B}_V$. Moreover, if $\mathrm{Bs}_{V,\mathfrak{I}} \subseteq \mathrm{Bs}_V$ is the set of all $\mathfrak{I}$-marked basic transformations between terms with scope V, then $|\cdot|_V$ maps $\mathrm{Bs}_{V,\mathfrak{I}}$ into the subset of invertible 2-morphisms in

[14]Implicitly, we interpret id_r for $r \in \mathcal{L}(\mathfrak{S})_V$ as $|\mathrm{id}_r|_V = \mathrm{id}_{|r|_V}$, although we need not formally state this.

$\coprod_{\mathtt{t}\in T}\underline{\mathbf{K}}(|V|,|\mathtt{t}|)$ by the vertical 2-functoriality of 2-products and whiskering, so we may further extend $|\cdot|_V$ to a functor

$$|\cdot|_V\colon F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}]) \to \coprod_{\mathtt{t}\in T}\underline{\mathbf{K}}(|V|,|\mathtt{t}|) \tag{2.73}$$

with $\left|\varphi^{-1}\right|_V = |\varphi|_V^{-1}$ for each $\varphi\in\mathrm{Bs}_{V,\mathfrak{I}}$.

For every morphism ν of $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$, we call the 2-morphism $|\nu|_V$ of $\underline{\mathbf{K}}$ a *coherence map* for the model $|\cdot|\colon(\mathfrak{S},\mathfrak{T},\mathfrak{I})\to\underline{\mathbf{K}}$. In particular, it is the coherence map between two *operations* for this model, as in Definition 2.51. For a basic transformation $\varphi\in\mathrm{Bs}_{\mathrm{V},\mathfrak{I}}$, we call $|\varphi|_V$ a *basic coherence map*; and for an atomic transformation $\varepsilon_{\mathbf{s}}$, we call $|\varepsilon_{\mathbf{s}}|_V$ an *atomic coherence map*.

We characterize the functor $|\cdot|_V$ defined in (2.73) in the following result.

Proposition 2.74. *Let $|\cdot|\colon\mathfrak{C}\to\underline{\mathbf{K}}$ be a model of the marked transformation schema $(\mathfrak{S},\mathfrak{T},\mathfrak{I})$ in the cartesian monoidal 2-category $\underline{\mathbf{K}}$, and let V be a finite set of variables for $\mathfrak{S}$. Then there is a unique functor*

$$|\cdot|_V\colon F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}]) \to \coprod_{\mathtt{t}\in T}\underline{\mathbf{K}}(|V|,|\mathtt{t}|) \tag{2.75}$$

satisfying conditions (1–3) from Proposition 2.62 on objects as well as the following conditions on morphisms:

(4) given atomic transformation $\varepsilon_{\mathbf{s}}\colon \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}]\Rightarrow\mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ for edge label $\varepsilon\in\mathcal{E}$ and terms $\mathbf{s}=(s_1,\ldots,s_n)$ with compatible types such that each $s_i\in\mathcal{L}(\mathfrak{S})_V$, we have

$$|\varepsilon_{\mathbf{s}}|_V = |\varepsilon|_{|\mathbf{s}|_V},$$

as in Definition 2.72 (1), with $|\varepsilon|$ given by the model;

(5) given basic transformation $f(\mathrm{id}_{r_1},\ldots,\varphi,\ldots,\mathrm{id}_{r_n})$ for terms $r_1,\ldots,r_n\in\mathcal{L}(\mathfrak{S})_V$, basic transformation $\varphi\in\mathrm{Bs}_V$ between two terms in $\mathcal{L}(\mathfrak{S})_V$, and n-ary function symbol $f\in F$ with $\mathrm{dom}(f)=(\tau(r_1),\ldots,\tau(\varphi),\ldots,\tau(r_n))$, we have

$$|f(\mathrm{id}_{r_1},\ldots,\varphi,\ldots,\mathrm{id}_{r_n})| = |f|\langle\mathrm{id},\ldots,|\varphi|_V,\ldots,\mathrm{id}\rangle$$

as in Definition 2.72 (2), with $|f|$ given by the model; and

(6) given an $\mathfrak{I}$-marked basic transformation $\varphi \in \mathrm{Bs}_{V,\mathfrak{I}}$, we have $\left|\varphi^{-1}\right|_V = |\varphi|_V^{-1}$.[15]

More generally, given a variable-finite, subterm-closed subset $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})_V$, let $\mathscr{B}_{\mathfrak{O}}$ be the full subgraph of $\mathscr{B}_V$ spanned by $\mathfrak{O}$ (with terms in $\mathfrak{O}$ as vertices and all basic transformations between terms in $\mathfrak{O}$ as edges), and let $\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}} \subseteq \mathrm{Bs}_{V,\mathfrak{I}}$ be the set of all $\mathfrak{I}$-marked basic transformations between terms in $\mathfrak{O}$. Then there is also a unique functor

$$|\cdot|_V \colon F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}^{-1}]) \to \coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|) \tag{2.76}$$

satisfying conditions (1–3) from Proposition 2.62 for terms in $\mathfrak{O}$, condition (4) for atomic transformations between terms in $\mathfrak{O}$, condition (5) for all other basic transformations between terms in $\mathfrak{O}$, and condition (6) for all $\mathfrak{I}$-marked basic transformations between terms in $\mathfrak{O}$.

Proof. The functor (2.75) satisfying the given conditions is precisely our construction (2.73) from Definition 2.72: on objects, it is the function (2.58) from Definition 2.56, which uniquely satisfies conditions (1–3) from Proposition 2.62; while on morphisms, it satisfies conditions (4–6) by construction. Its well-definedness and uniqueness are given by induction and the Parsing Lemma on the graph $\mathscr{B}_V$, condition (6) fixes $|\cdot|_V$ uniquely on $\mathrm{Bs}_{V,\mathfrak{I}}^{-1}$, and freeness uniquely extends $|\cdot|_V$ to $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$.

Then the functor (2.76) is obtained by restricting the domain of (2.75) to $F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}^{-1}])$ with subterm-closed object-set $\mathfrak{O}$. On objects, (2.76) is the function (2.64) from Proposition 2.62, which ensures its uniqueness. On morphisms, uniqueness is given by induction on the basic transformations between terms in $\mathfrak{O}$ that form the edges of $\mathscr{B}_{\mathfrak{O}}$. Condition (4) fixes the functor on atomic transformations between terms in $\mathfrak{O}$; then given a basic transformation

$$f(\mathrm{id}_{r_1}, \ldots, \varphi, \ldots, \mathrm{id}_{r_n}) \colon f(r_1, \ldots, \mathrm{dom}(\varphi), \ldots, r_n) \Rightarrow f(r_1, \ldots, \mathrm{cod}(\varphi), \ldots, r_n)$$

whose domain and codomain are in $\mathfrak{O}$, subterm-closure ensures $\mathrm{dom}(\varphi)$ and $\mathrm{cod}(\varphi)$ are in $\mathfrak{O}$ as well, so φ is also an edge in $\mathscr{B}_{\mathfrak{O}}$. Therefore condition (5) fixes the functor on basic

[15] This condition is necessary because φ^{-1} in the free category is merely a formal symbol, and *not* the categorical inverse of φ.

transformations between terms in $\mathfrak{O}$ by induction. Finally, (6) fixes the functor on $\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}^{-1}$, and uniqueness extends to the free category by functoriality. $\square$

We may now state what a model for the entire coherence schema $\mathfrak{C} = (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R})$ is in our cartesian monoidal 2-category $\underline{\mathbf{K}}$. Recall that $\mathscr{B}$ is the graph of terms in $\mathcal{L}(\mathfrak{S})$ and basic transformations between them from the transformation schema $\mathfrak{T}\colon \mathcal{E} \to \mathcal{L}(\mathfrak{S}) \times_T \mathcal{L}(\mathfrak{S})$, and $\mathrm{Bs}_{\mathfrak{I}}$ is the set of $\mathfrak{I}$-marked basic transformations, with $\mathfrak{R}$ a local binary relation on the free category $F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{I}}^{-1}])$. We employ the following notation for the variables appearing in a morphism of $F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{I}}^{-1}])$.

Definition 2.77 (Model of coherence schema)**.** A *model for* $\mathfrak{C}$ *in* $\underline{\mathbf{K}}$ is a model $|\cdot|$ for $(\mathfrak{S}, \mathfrak{T}, \mathfrak{I})$, as in Definition 2.69, such that for any pair $\nu, \nu'\colon t \Rightarrow t'$ of $\mathfrak{R}$-related transformations in $F(\mathscr{B}[\mathrm{Bs}_{\mathfrak{I}}^{-1}])$, the coherence maps $|\nu|_V$ and $|\nu'|_V$ as defined in Definition 2.72 are the same 2-morphism in $\underline{\mathbf{K}}(|V|, |\tau(t)|)$, where $V \coloneqq V(\nu) \cup V(\nu')$.

Example 2.78 (Model of coherence schemata for monoidal categories)**.** A model of the coherence schema $\mathfrak{C}_{\mathrm{LM}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing, \mathfrak{R}^{\mathrm{LM}})$ for lax monoidal categories from Example 2.44 in $\underline{\mathbf{Cat}}$ is a model $(\mathscr{C}, e^{\mathscr{C}}, \otimes^{\mathscr{C}}, \alpha^{\mathscr{C}}, \lambda^{\mathscr{C}}, \rho^{\mathscr{C}})$ from Example 2.70 of the transformation schema $(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing)$ in $\underline{\mathbf{Cat}}$ such that the interpretations of the five diagrams defining $\mathfrak{R}^{\mathrm{LM}}$ from Example 2.44 as diagrams of natural transformations between functors (one between functors $\mathscr{C}^4 \rightrightarrows \mathscr{C}$ and the other between functors $\mathscr{C}^2 \rightrightarrows \mathscr{C}$) commute. This is precisely the definition of a *lax monoidal category.* A model of the coherence schema $\mathfrak{C}_{\mathrm{M}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{M}})$ for monoidal categories from Example 2.44 in $\underline{\mathbf{Cat}}$ is then a lax monoidal category in which the associator $\alpha^{\mathscr{C}}$, the left unitor $\lambda^{\mathscr{C}}$, and the right unitor $\rho^{\mathscr{C}}$ are all natural isomorphisms and only the interpretations of the first two diagrams defining $\mathfrak{R}^{\mathrm{M}}$ need commute. This is precisely the definition of a *monoidal category.*

Finally, a model of the coherence schema $\mathfrak{C}_{\mathrm{SM}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{SM}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{SM}})$ for symmetric monoidal categories from Example 2.44 in $\underline{\mathbf{Cat}}$ is a monoidal category equipped with an additional braiding natural transformation $\beta^{\mathscr{C}}$ (as in Example 2.70) such that the three diagrams specified by the additional relations in $\mathfrak{R}^{\mathrm{SM}}$ commute (one of which makes the braiding a natural isomorphism). This is precisely the definition of a *symmetric monoidal category.*

Example 2.79 (Model of coherence schema for bicategories)**.** Given a set O, a model of the coherence schema $\mathfrak{C}_{\mathrm{LB}}(O) = (\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \varnothing, \mathfrak{R}^{\mathrm{LB}}(O))$ from Example 2.45 in $\underline{\mathbf{Cat}}$ is a model from Example 2.71 of the transformation schema $(\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \varnothing)$ in $\underline{\mathbf{Cat}}$ such that the interpretations of the five classes of diagrams defining $\mathfrak{R}^{\mathrm{LB}}(O)$ from Example 2.44 as classes of diagrams of natural transformations between functors commute.

Meanwhile, a model of the coherence schema $\mathfrak{C}_{\mathrm{B}}(O) = (\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \mathcal{E}_{\mathrm{B}}(O), \mathfrak{R}^{\mathrm{B}}(O))$ from Example 2.45 in $\underline{\mathbf{Cat}}$ is then a model from Example 2.71 of the transformation schema $(\mathfrak{S}_{\mathrm{B}}(O), \mathfrak{T}_{\mathrm{B}}(O), \mathcal{E}_{\mathrm{B}}(O))$ in $\underline{\mathbf{Cat}}$, where the associators and unitors are all natural isomorphisms, such that the interpretations of the two classes of diagrams defining $\mathfrak{R}^{\mathrm{B}}(O)$ from Example 2.44 as classes of diagrams of natural transformations between functors commute. This is precisely the definition of a *(small) bicategory with object-set O*.

We would like to extend Proposition 2.74 to (restricted) coherence schemata, but before we do so, we must prove some technical lemmas about interpreting basic transformations. First, we extend Lemma 2.65 to basic transformations to show that models remain compatible with scope expansion and substitution.

Lemma 2.80. *Fix a basic transformation $\varphi\colon t \Rightarrow t'$ for which t and t' have common scope $W = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}$ and $\mathbf{r} = (r_1, \ldots, r_n)$ consisting of terms $r_1, \ldots, r_n \in \mathcal{L}(\mathfrak{S})$ with common scope V such that $\tau(r_j) = \mathtt{t}_j$ for all $1 \leq j \leq n$. Let $\varphi[W/\mathbf{r}]$ denote the basic transformation obtained by substituting r_j for each $x_{i_j}^{\mathtt{t}_j}$ in φ for $1 \leq j \leq n$ from Definition 2.38. Then $|\varphi[W/\mathbf{r}]|_V$ is the whiskered 2-morphism $(|\varphi|_W)_{|\mathbf{r}|_V}$, or*

$$\begin{tikzcd} {|V|} \arrow[r, "{\langle |r_1|_V, \ldots, |r_n|_V \rangle}"] & {|\mathtt{t}_1| \times \cdots \times |\mathtt{t}_n| = |W|} \arrow[r, bend left, "{|t|_W}"] \arrow[r, bend right, "{|t'|_W}"'] \arrow[r, phantom, "\Downarrow |\varphi|_W"] & {|\tau(\varphi)|.} \end{tikzcd} \tag{2.81}$$

Proof. We proceed by induction on φ. Given $\varphi = \varepsilon_{s_1, \ldots, s_m}$ with each $s_i \in \mathcal{L}(\mathfrak{S})_W$, by Definition 2.38, $\varphi[W/\mathbf{r}] = \varepsilon_{s_1[W/\mathbf{r}], \ldots, s_m[W/\mathbf{r}]}$. So by the construction of the coherence maps $|\varphi|_W$ and $|\varphi[W/\mathbf{r}]|_V$ in Definition 2.72, it suffices to show that

$$\langle |s_1|_W, \ldots, |s_m|_W \rangle \circ |\mathbf{r}|_V = \langle |s_1[W/\mathbf{r}]|_V, \ldots, |s_m[W/\mathbf{r}]|_V \rangle,$$

but this follows directly from Lemma 2.65.

Now fix terms $s_1, \ldots, s_{k-1}, s_{k+1}, \ldots, s_m \in \mathcal{L}(\mathfrak{S})_W$, a basic transformation φ between terms in $\mathcal{L}(\mathfrak{S})_W$, and an m-ary function symbol $f \in F$ whose domain is given by $\mathrm{dom}(f) = (\tau(s_1), \ldots, \tau(s_{k-1}), \tau(\varphi), \tau(s_{k+1}), \ldots, \tau(s_m))$, so that we have another basic transformation

$$\varphi' = f(\mathrm{id}_{s_1}, \ldots, \underbrace{\varphi}_{k^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}_{s_m}).$$

By Definition 2.38, $\varphi'[W/\mathbf{r}] = f(\mathrm{id}_{s_1[W/\mathbf{r}]}, \ldots, \varphi[W/\mathbf{r}], \ldots, \mathrm{id}_{s_m[W/\mathbf{r}]})$, so by the construction of the coherence maps $|\varphi'|_W$ and $|\varphi'[W/\mathbf{r}]|_V$ in Definition 2.72, it suffices to show that the 2-morphism $\langle \mathrm{id}_{|s_1[W/\mathbf{r}]|_V}, \ldots, |\varphi[W/\mathbf{r}]|_V, \ldots, \mathrm{id}_{|s_m[W/\mathbf{r}]|_V} \rangle$ can be obtained by whiskering the 2-morphism $\langle \mathrm{id}_{|s_1|_W}, \ldots, |\varphi|_W, \ldots, \mathrm{id}_{|s_m|_W} \rangle$ with the 1-morphism $\langle |r_1|_V, \ldots, |r_n|_V \rangle$. This follows by induction for the k^{th} component and by Lemma 2.65 elsewhere. $\square$

Corollary 2.82. *Fix a basic transformation $\varphi\colon t \Rightarrow t'$ for which t and t' have common scope $W = \{x_{i_1}^{\mathtt{t}_1} < \cdots < x_{i_n}^{\mathtt{t}_n}\}$, and let V be a finite set of variables containing W. Then $|\varphi|_V$ is the whiskered 2-morphism $(|\varphi|_W)_\pi$ for the projection $\pi\colon |V| \to |W|$.*

Proof. Analogously to Corollary 2.67, the result follows from Lemma 2.80 with $r_j := x_{i_j}^{\mathtt{t}_j}$ for all $1 \le j \le n$, since substituting each $x_{i_j}^{\mathtt{t}_j}$ for itself yields $\varphi[W/\mathbf{r}] = \varphi$ itself, while $|\mathbf{r}|_V = \langle |x_{i_1}^{\mathtt{t}_1}|_V, \ldots, |x_{i_n}^{\mathtt{t}_n}|_V \rangle$ is precisely the projection $|V| \to |W|$. $\square$

Recall the commutative triangle of functions in (2.68) for a variable-finite, subterm-closed subset $\mathfrak{O} \subset \mathcal{L}(\mathfrak{S})$ and a finite set of variables V containing $V(\mathfrak{O})$, the set of all variables appearing in $\mathfrak{O}$. The previous corollary and the vertical 2-functoriality of whiskering shows that we may extend that triangle to a commutative triangle of functors $|\cdot|_V$ and $|\cdot|_{V(\mathfrak{O})}$ from (2.76):

$$\begin{tikzcd} F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{J}}^{-1}]) \arrow[r, "|\cdot|_{V(\mathfrak{O})}"] \arrow[dr, "|\cdot|_V"'] & \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V(\mathfrak{O})|, |\mathtt{t}|) \arrow[d] \\ & \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|). \end{tikzcd} \tag{2.83}$$

Here the vertical map is given by whiskering with the projection $\pi\colon |V| \to |V(\mathfrak{O})|$.

We also show that models are compatible with basic transformations that act on substituted terms as follows. Note that the construction given in the following hypothesis generalizes the constructions from Definition 2.46 (c).

Lemma 2.84. *Let t be a term with scope $W = \{x_{n_1}^{\mathtt{t}_1} < \cdots < x_{n_k}^{\mathtt{t}_k}\}$, and let $s_1, \ldots, s_k$ and s'_ℓ be terms with common scope V such that $\tau(s_j) = \mathtt{t}_j$ for $1 \leq j \leq k$ and $\tau(s'_\ell) = \mathtt{t}_\ell$ for some $1 \leq \ell \leq k$. Let $\mathbf{s} = (s_1, \ldots, s_k)$ and $\mathbf{s}' = (s_1, \ldots, s_{\ell-1}, s'_\ell, s_{\ell+1}, \ldots, s_k)$.*

Now let $\psi\colon s_\ell \Rightarrow s'_\ell$ be a basic transformation, and let the pairs $(a_1, b_1), \ldots, (a_m, b_m)$ denote the positions of the instances of s_ℓ in $t_0 = t[W/\mathbf{s}]$, ordered from right to left (so $a_m < b_m < \cdots < a_1 < b_1$), arising from substitutions of s_ℓ for $x_{n_\ell}^{\mathtt{t}_\ell}$ in t. For $1 \leq i \leq m$, recursively define $\psi_i\colon t_{i-1} \Rightarrow t_i$ to be the unique basic transformation with atom $A(\psi)$ and span $S(\psi) + a_i - 1$, so that $t_m = t[W/\mathbf{s}']$, still with s_j substituted for $x_{n_j}^{\mathtt{t}_j}$ for $j \neq \ell$, but with s'_ℓ substituted for $x_{n_\ell}^{\mathtt{t}_\ell}$. Then the vertical composite 2-morphism $|\psi_m|_V \bullet \cdots \bullet |\psi_1|_V$ is given by whiskering

$$\begin{tikzcd}
{|V|} \arrow[r, bend left, "{\langle |s_1|_V, \ldots, |s_\ell|_V, \ldots, |s_k|_V \rangle}"] \arrow[r, bend right, "{\langle |s_1|_V, \ldots, |s'_\ell|_V, \ldots, |s_k|_V \rangle}"'] \arrow[r, phantom, "\Downarrow \overline{\psi}"] & {|W|} \arrow[r, "|t|_W"] & {|\tau(t)|,}
\end{tikzcd}$$

where the 2-morphism $\overline{\psi}$ is given by

$$\langle \mathrm{id}_{|s_1|_V}, \ldots, \mathrm{id}_{|s_{\ell-1}|_V}, |\psi|_V, \mathrm{id}_{|s_{\ell+1}|_V}, \ldots, \mathrm{id}_{|s_k|_V} \rangle.$$

We write the whiskered 2-morphism as $|t|_W \overline{\psi} = |t|_W \langle \mathrm{id}, \ldots, |\psi|_V, \ldots, \mathrm{id} \rangle$ for short.

Proof. The domain 1-morphism $|t|_W \circ |\mathbf{s}|_V$ and the codomain 1-morphism $|t|_W \circ |\mathbf{s}'|_V$ of the whiskered 2-morphism are correct by Lemma 2.65. We then proceed by induction on t. If t is a constant c, substitution does nothing, $m = 0$, and the vertical composite 2-morphism in question is empty and thus the identity on the 1-morphism $|c|_V = |c| \circ !\colon |V| \to * \to |\tau(c)|$, which by the terminality of $*$ is equal to whiskering $|t|_W = |c| \circ !\colon |W| \to * \to |\tau(c)|$ with any 2-morphism $\overline{\psi}$ between 1-morphisms $|V| \rightrightarrows |W|$.

If t is a variable $x_{i_j}^{\mathtt{t}_j}$, then $\left|x_{i_j}^{\mathtt{t}_j}\right|_W$ is the j^{th} projection, and whiskering yields the j^{th} component of $\overline{\psi}$. If $j \neq \ell$, again $m = 0$ and our vertical composite 2-morphism is the identity, this time on the 1-morphism $|s_j|_V$, coinciding with the j^{th} component of $\overline{\psi}$. Otherwise, $j = \ell$, making $m = 1$, $t_0 = s_\ell$, $\psi_1 = \psi\colon s_\ell \Rightarrow s'_\ell$, and our vertical composite 2-morphism $|\psi|_V$, coinciding with the ℓ^{th} component of $\overline{\psi}$.

Now let $t = f(r_1, \dots, r_p)$ with $\mathrm{dom}(f) = (\tau(r_1), \dots, \tau(r_p))$, so W is a scope for each r_q, and by the Parsing Lemma $t_0 = t[W/\mathbf{s}] = f(r_1[W/\mathbf{s}], \dots, r_p[W/\mathbf{s}])$ and $t_m = t[W/\mathbf{s}'] = f(r_1[W/\mathbf{s}'], \dots, r_p[W/\mathbf{s}'])$. The result will follow via induction by the vertical 2-functoriality of whiskering. For $1 \leq q \leq p$, let m_q be the number of times $x_{i_\ell}^{\mathtt{t}_\ell}$ appears in r_q, so that we have $m_1 + \cdots + m_p = m$; let the pairs $(a_{q,1}, b_{q,1}), \dots, (a_{q,m_q}, b_{q,m_q})$ denote the positions of the instances of s_ℓ substituted for $x_{i_\ell}^{\mathtt{t}_\ell}$ in $r_q[W/\mathbf{s}]$ ordered right to left. Then for all $1 \leq i \leq m_q$ recursively define $r_{q,0} = r_q[W/\mathbf{s}]$ and $\psi_{q,i}\colon r_{q,i-1} \Rightarrow r_{q,i}$ to be the unique basic transformation with atom $A(\psi)$ and span $S(\psi) + a_{q,i} - 1$, so $r_{q,m_q} = r_q[W/\mathbf{s}']$. Then by induction, $\left|\psi_{q,m_q}\right|_V \bullet \cdots \bullet |\psi_{q,1}|_V$ is given by whiskering $|r_q|_W\overline{\psi}$.

By the Parsing Lemma, when $r_{q,0}$ is interpreted as the q^{th} entry of t_0, each position $(a_{q,i}, b_{q,i})$ of s_ℓ in $r_{q,0}$ corresponds to the position (a_{M_q+i}, b_{M_q+i}) of s_ℓ in t_0, where $M_q \coloneqq m_{q+1} + \cdots + m_p$. It follows from Lemma 2.37 and the construction of the t_i's, ψ_i's, $r_{q,i}$'s, and $\psi_{q,i}$'s that

$$\begin{aligned} t_{M_q+i} &= f(r_{1,0}, \dots, r_{q-1,0}, r_{q,i}, r_{q+1,m_{q+1}}, \dots, r_{p,m_p}) \\ \text{and} \quad \psi_{M_q+i} &= f(\mathrm{id}_{r_{1,0}}, \dots, \mathrm{id}_{r_{q-1,0}}, \psi_{q,i}, \mathrm{id}_{r_{q+1,m_{q+1}}}, \dots, \mathrm{id}_{r_{p,m_p}}), \end{aligned}$$

so by Definition 2.72, $\left|\psi_{M_q+i}\right|_V$ is the whiskered 2-morphism

$$\left|\psi_{M_q+i}\right|_V = |f|\langle \mathrm{id}_{|r_{1,0}|_V}, \dots, \mathrm{id}_{|r_{q-1,0}|_V}, |\psi_{q,i}|_V, \mathrm{id}_{\left|r_{q+1,m_{q+1}}\right|_V}, \dots, \mathrm{id}_{|r_{p,m_p}|_V}\rangle.$$

So by the inductive hypothesis and the vertical 2-functoriality of 2-products and of whiskering, $\left|\psi_{M_q+m_q}\right|_V \bullet \cdots \bullet \left|\psi_{M_q+1}\right|_V$ is the whiskered 2-morphism

$$|f|\langle \mathrm{id}_{|r_{1,0}|_V}, \dots, \mathrm{id}_{|r_{q-1,0}|_V}, |r_q|_W\overline{\psi}, \mathrm{id}_{\left|r_{q+1,m_{q+1}}\right|_V}, \dots, \mathrm{id}_{|r_{p,m_p}|_V}\rangle.$$

Then by Definition 2.56 (3), the vertical 2-functoriality of 2-products, and both the horizontal 1-functoriality and the vertical 2-functoriality of whiskering, $|\psi_m|_V \bullet \cdots \bullet |\psi_1|_V$ is

$$\begin{aligned} |f|\langle |r_1|_W \circ \overline{\psi}, \dots, |r_p|_W \circ \overline{\psi}\rangle &= |f|\langle |r_1|_W, \dots, |r_p|_W\rangle \overline{\psi} \\ &= |t|_W \overline{\psi} \end{aligned}$$

as desired. $\square$

Putting this all together with Propositions 2.62 and 2.74, the next theorem makes precise the sense in which the parameter categories $\mathscr{P}_V(\mathfrak{C})$ and $\mathscr{P}^{\mathrm{L}}_V(\mathfrak{C})$ of Definition 2.51 for the restricted coherence schema $(\mathfrak{C}, V, \mathcal{L}(\mathfrak{S})_V)$ are universal among models $|\cdot|\colon \mathfrak{C} \to \underline{\mathbf{K}}$.

Theorem 2.85. *Let $\mathfrak{C}$ be a coherence schema and $|\cdot|\colon \mathfrak{C} \to \underline{\mathbf{K}}$ a model for $\mathfrak{C}$ in the cartesian monoidal 2-category $\underline{\mathbf{K}}$. Fix a finite set of variables V for $\mathfrak{C}$, yielding a restricted coherence schema $(\mathfrak{C}, \mathcal{L}(\mathfrak{S})_V)$, and define the category $\mathscr{P}_V(\mathfrak{C})$ presented by $(\mathfrak{C}, \mathcal{L}(\mathfrak{S})_V)$ as in Definition 2.51. Then there is a unique functor*

$$|\cdot|_V \colon \mathscr{P}_V(\mathfrak{C}) \to \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|) \tag{2.86}$$

such that

(1) $\left|x^{\mathtt{t}}_k\right|_V$ is the projection $|V| \to |\mathtt{t}|$ to the $x^{\mathtt{t}}_k$-factor, with $|\mathtt{t}|$ given by the model, for each variable $x^{\mathtt{t}}_k \in V$;

(2) $|c|_V$ is the constant morphism $|V| \to * \xrightarrow{|c|} |\tau(c)|$, with $*$ the terminal object of $\underline{\mathbf{K}}$ and $|c|\colon * \to |\tau(c)|$ given by the model, for each constant $c \in C$;

(3) $|f(s_1, \dots, s_n)|_V = |f| \circ |\mathbf{s}|_V$ for an n-ary function symbol $f \in F$ and n-tuple of terms $\mathbf{s} = (s_1, \dots, s_n)$ from $\mathcal{L}(\mathfrak{S})_V$ for which $\mathrm{dom}(f) = (\tau(s_1), \dots, \tau(s_n))$, with $|f|$ as given by the model and $|\mathbf{s}|_V = \langle |s_1|_V, \dots, |s_n|_V\rangle$;

(4) given atomic transformation $\varepsilon_{\mathbf{s}}\colon \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}] \Rightarrow \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ for edge label $\varepsilon \in \mathcal{E}$ and terms $\mathbf{s} = (s_1, \dots, s_n)$ from $\mathcal{L}(\mathfrak{S})_V$ with compatible types, if $[\varepsilon_{\mathbf{s}}]$ is the residue of $\varepsilon_{\mathbf{s}}$ in the quotient category $\mathscr{P}_V(\mathfrak{C})$, then

$$|[\varepsilon_{\mathbf{s}}]|_V = |\varepsilon|_{|\mathbf{s}|_V}$$

as in Definition 2.72 (1), with $|\varepsilon|$ given by the model; and

(5) *given basic transformation* $f(\mathrm{id}_{r_1}, \dots, \varphi, \dots, \mathrm{id}_{r_n})$ *for terms* $r_1, \dots, r_n \in \mathcal{L}(\mathfrak{S})_V$*, basic transformation* φ *between terms in* $\mathcal{L}(\mathfrak{S})_V$*, and* n*-ary function symbol* $f \in F$ *with* $\mathrm{dom}(f) = (\tau(r_1), \dots, \tau(\varphi), \dots, \tau(r_n))$*, the functor* $|\cdot|_V$ *applied to the basic transformation's residue in* $\mathscr{P}_V(\mathfrak{C})$ *is given by*

$$|[f(\mathrm{id}_{r_1}, \dots, \varphi, \dots, \mathrm{id}_{r_n})]|_V = |f|\langle \mathrm{id}, \dots, |\varphi|_V, \dots, \mathrm{id}\rangle$$

as in Definition 2.72 (2), with $|f|$ *given by the model.*[16]

Moreover, if the functor $|\cdot|_V$ *in* (2.86) *sends the residue of every basic transformation in* Bs_V *to a monomorphism, then* $|\cdot|_V$ *factors through the quotient functor* $\mathscr{P}_V(\mathfrak{C}) \to \mathscr{P}^{\mathrm{L}}_V(\mathfrak{C})$ *from Definition 2.51.*

Proof. In Proposition 2.74, we constructed a functor (2.75) denoted

$$|\cdot|_V : F(\mathscr{B}_V[\mathrm{Bs}^{-1}_{V,\mathfrak{I}}]) \to \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|)$$

that satisfies conditions (1–3) on objects as well as (4–5) for the edges in $\mathscr{B}_V[\mathrm{Bs}^{-1}_{V,\mathfrak{I}}]$ (i.e. basic transformations φ and, if $\mathfrak{I}$-marked, their inverses) rather than their residues in $\mathscr{P}_V(\mathfrak{C})$ (i.e. $[\varphi]$ and, if φ is $\mathfrak{I}$-marked, $[\varphi^{-1}]$). The functor (2.75) further satisfies the condition that

(6) given an $\mathfrak{I}$-marked basic transformation $\varphi \in \mathrm{Bs}_{V,\mathfrak{I}}$, we have $\left|\varphi^{-1}\right|_V = |\varphi|_V^{-1}$,

and once this condition is imposed, (2.75) is necessarily unique.

We now invoke Proposition A.9 to show that the functor (2.75) descends uniquely to the quotient $\mathscr{P}_V(\mathfrak{C}) = F(\mathscr{B}_V[\mathrm{Bs}^{-1}_{V,\mathfrak{I}}])/\langle R(\mathfrak{C})_V\rangle$, which then must continue to satisfy (1–3) on objects while also satisfying (4–5) as written. To do so, it suffices to show that $|\cdot|_V$ agrees on every $R(\mathfrak{C})_V$-related pair of morphisms in $F(\mathscr{B}_V[\mathrm{Bs}^{-1}_{V,\mathfrak{I}}])$, with $R(\mathfrak{C})_V$ defined as in Definition 2.46. We consider each of the four named subrelations of $R(\mathfrak{C})_V$ in turn, using the symbols $\bullet$ to denote vertical composition and $\circ$ to denote horizontal composition of 2-morphisms in $\underline{\mathbf{K}}$.

[16] It turns out we no longer need an analogue for Proposition 2.74 (6), for when we quotient out by $\langle R(\mathfrak{C})_V\rangle$ to obtain $\mathscr{P}_V(\mathfrak{C})$, inverse laws are enforced, making $[\varphi^{-1}]$ the actual categorical inverse of $[\varphi]$ in $\mathscr{P}_V(\mathfrak{C})$ for $\varphi \in \mathrm{Bs}_{V,\mathfrak{I}}$.

(a) For all $\varphi \in \mathrm{Bs}_{V,\mathfrak{I}}$, we have $\left|\varphi^{-1}\right|_V = |\varphi|_V^{-1}$ by (6), so by functoriality of $|\cdot|_V$ we have

$$\left|\varphi^{-1} \bullet \varphi\right|_V = |\varphi|_V^{-1} \bullet |\varphi|_V = \mathrm{id}_{\mathrm{dom}(|\varphi|_V)} = \mathrm{id}_{|\mathrm{dom}\,\varphi|_V} = |\mathrm{id}_{\mathrm{dom}\,\varphi}|_V$$

and

$$\left|\varphi \bullet \varphi^{-1}\right|_V = |\varphi|_V \bullet |\varphi|_V^{-1} = \mathrm{id}_{\mathrm{cod}(|\varphi|_V)} = \mathrm{id}_{|\mathrm{cod}\,\varphi|_V} = |\mathrm{id}_{\mathrm{cod}\,\varphi}|_V.$$

Hence $|\cdot|_V$ agrees on the R^{inv}-related pairs from Definition 2.46 (a).

(b) Given basic transformations $\varphi\colon s \Rightarrow t_1$ and $\psi\colon s \Rightarrow t_2$ in $\mathscr{B}_V$ with $S_2(\varphi) < S_1(\psi)$, define $\varphi'\colon t_2 \Rightarrow u$ and $\psi'\colon t_1 \Rightarrow u$ in $\mathscr{B}_V$ as in Definition 2.46 (b) for R^{fun}. By the functoriality of $|\cdot|_V$, it suffices to show that $|\psi'|_V \bullet |\varphi|_V = |\varphi'|_V \bullet |\psi|_V$.

We proceed via induction on s. The hypothesis is impossible if s is atomic, so let $s = f(s_1, \ldots, s_n)$ for n-ary $f \in F$ and terms $s_1, \ldots, s_n \in \mathcal{L}(\mathfrak{S})_V$. By Corollary 2.16, the $S(\varphi)$-subterm $\mathrm{dom}(A(\varphi))$ of s is in its i^{th} entry s_i and the $S(\psi)$-subterm $\mathrm{dom}(A(\psi))$ of s is in its j^{th} entry s_j for some $1 \le i \le j \le n$. It follows that the $S(\varphi)$-subterm $\mathrm{dom}(A(\varphi))$ of t_2 is also in its i^{th} entry s_i, while the $(S(\psi)+\Delta(\varphi))$-subterm $\mathrm{dom}(A(\psi))$ of t_1 is also in its j^{th} entry s_j. We can therefore write

$$\begin{aligned} \varphi &= f(\mathrm{id}_{s_1}, \ldots, \mathrm{id}_{s_{i-1}}, \varphi_0, \mathrm{id}_{s_{i+1}}, \ldots, \mathrm{id}_{s_n}) \\ \text{and} \quad \psi &= f(\mathrm{id}_{s_1}, \ldots, \mathrm{id}_{s_{j-1}}, \psi_0, \mathrm{id}_{s_{j+1}}, \ldots, \mathrm{id}_{s_n}) \end{aligned}$$

for basic transformations $\varphi_0 \in \mathrm{Bs}_V$ with domain s_i and atom $A(\varphi)$ and $\psi_0 \in \mathrm{Bs}_V$ with domain s_j and atom $A(\psi)$, so that

$$\begin{aligned} t_1 &= f(s_1, \ldots, s_{i-1}, \mathrm{cod}(\varphi_0), s_{i+1}, \ldots, s_n) \\ \text{and} \quad t_2 &= f(s_1, \ldots, s_{j-1}, \mathrm{cod}(\psi_0), s_{j+1}, \ldots, s_n). \end{aligned}$$

We have two cases: either $i = j$ or $i < j$.

If $i = j$, because $S_2(\varphi) < S_1(\psi)$, we must have $S_2(\varphi_0) < S_1(\psi_0)$, so

$$\begin{aligned} \varphi' &= f(\mathrm{id}_{s_1}, \ldots, \mathrm{id}_{s_{i-1}}, \varphi_0', \mathrm{id}_{s_{i+1}}, \ldots, \mathrm{id}_{s_n}) \\ \text{and} \quad \psi' &= f(\mathrm{id}_{s_1}, \ldots, \mathrm{id}_{s_{i-1}}, \psi_0', \mathrm{id}_{s_{i+1}}, \ldots, \mathrm{id}_{s_n}), \end{aligned}$$

for some basic transformations $\varphi_0' \in \mathrm{Bs}_V$ with domain $\operatorname{cod}(\psi_0)$, atom $A(\varphi)$, and span $S(\varphi_0)$ and $\psi_0' \in \mathrm{Bs}_V$ with domain $\operatorname{cod}(\varphi_0)$, atom $A(\psi)$, span $S(\psi_0) + \Delta(\varphi_0)$, and the same codomain as φ_0'. Then by induction $|\psi_0'|_V \bullet |\varphi_0|_V = |\varphi_0'|_V \bullet |\psi|_V$, and by Proposition 2.74 (5) we have the whiskered 2-morphism

$$|\xi|_V = |f|\langle \mathrm{id}, \ldots, |\xi_0|_V, \ldots, \mathrm{id}\rangle$$

for $\xi \in \{\varphi, \psi, \varphi', \psi'\}$, so the result follows from the vertical 2-functoriality of 2-products and whiskering.

Otherwise, $i < j$, and we have

$$\varphi' = f(\mathrm{id}_{s_1}, \ldots, \underbrace{\varphi_0'}_{i^{\text{th}} \text{ entry}}, \ldots, \underbrace{\mathrm{id}_{\operatorname{cod}\psi_0}}_{j^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}_{s_n})$$

$$\text{and} \quad \psi' = f(\mathrm{id}_{s_1}, \ldots, \underbrace{\mathrm{id}_{\operatorname{cod}\varphi_0}}_{i^{\text{th}} \text{ entry}}, \ldots, \underbrace{\psi_0'}_{j^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}_{s_n})$$

for some basic transformations $\varphi_0' \in \mathrm{Bs}_V$ with domain s_i, atom $A(\varphi)$, and span $S(\varphi_0)$ and $\psi_0' \in \mathrm{Bs}_V$ with domain s_j, atom $A(\psi)$, and span $S(\psi_0)$. By uniqueness from Lemma 2.37, it follows that $\varphi_0 = \varphi_0'$ and $\psi_0 = \psi_0'$. So by Proposition 2.74 (5), we have the whiskered 2-morphisms

$$|\varphi|_V = |f|\langle \mathrm{id}, \ldots, \underbrace{|\varphi_0|_V}_{i^{\text{th}} \text{ entry}}, \ldots, \underbrace{\mathrm{id}_{|s_j|_V}}_{j^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}\rangle,$$

$$|\psi|_V = |f|\langle \mathrm{id}, \ldots, \underbrace{\mathrm{id}_{|s_i|_V}}_{i^{\text{th}} \text{ entry}}, \ldots, \underbrace{|\psi_0|_V}_{j^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}\rangle,$$

$$|\varphi'|_V = |f|\langle \mathrm{id}, \ldots, \underbrace{|\varphi_0|_V}_{i^{\text{th}} \text{ entry}}, \ldots, \underbrace{\mathrm{id}_{|\operatorname{cod}(\psi_0)|_V}}_{j^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}\rangle,$$

$$\text{and} \quad |\psi'|_V = |f|\langle \mathrm{id}, \ldots, \underbrace{\mathrm{id}_{|\operatorname{cod}(\varphi_0)|_V}}_{i^{\text{th}} \text{ entry}}, \ldots, \underbrace{|\psi_0|_V}_{j^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}\rangle.$$

It follows that $|\psi'|_V \bullet |\varphi|_V = |\varphi'|_V \bullet |\psi|_V$, for by the 2-functoriality of 2-products, both sides equal

$$|f|\langle \mathrm{id}, \ldots, \underbrace{|\varphi_0|_V}_{i^{\text{th}} \text{ entry}}, \ldots, \underbrace{|\psi_0|}_{j^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}\rangle,$$

implying that $|\cdot|_V$ agrees on R^{fun}-related pairs.

(c) Given an atomic transformation $\varepsilon_{\mathbf{s}} \colon t_0 \Rightarrow t'_0$ in $\mathscr{B}_V$ with $\mathbf{s} = (s_1, \ldots, s_k) \in \mathcal{L}(\mathfrak{S})^k_V$, fix some $1 \leq \ell \leq k$ and a basic transformation $\psi \colon s_\ell \Rightarrow s'_\ell$ in $\mathscr{B}_V$. Define integers m and n, basic transformations $\psi_i \colon t_{i-1} \Rightarrow t_i$ in $\mathscr{B}_V$ for $1 \leq i \leq m$, and basic transformations $\psi'_j \colon t'_{j-1} \Rightarrow t'_j$ in $\mathscr{B}_V$ for $1 \leq j \leq n$, as in Definition 2.46 (c) for R^{nat}; then we have an atomic transformation $\varepsilon_{\mathbf{s}'} \colon t_m \Rightarrow t'_n$ in $\mathscr{B}_V$ with $\mathbf{s}' = (s_1, \ldots, s_{\ell-1}, s'_\ell, s_{\ell+1}, \ldots, s_k)$. By the functoriality of $|\cdot|_V$, it suffices to show that

$$\left|\psi'_n\right|_V \bullet \cdots \bullet \left|\psi'_1\right|_V \bullet |\varepsilon_{\mathbf{s}}|_V = |\varepsilon_{\mathbf{s}'}|_V \bullet |\psi_m|_V \bullet \cdots \bullet |\psi_1|_V.$$

Note that $t_0 = \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ and $t'_0 = \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$, while $t_m = \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}']$ and $t'_n = \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}']$. By Proposition 2.74 (4), $|\varepsilon_{\mathbf{s}}|_V = |\varepsilon|_{|\mathbf{s}|_V}$ and $|\varepsilon_{\mathbf{s}'}|_V = |\varepsilon|_{|\mathbf{s}'|_V}$, with $|\varepsilon| \colon |\mathrm{dom}(\varepsilon)|_{V(\varepsilon)} \Rightarrow |\mathrm{cod}(\varepsilon)|_{V(\varepsilon)}$ given by the model as in Definition 2.69. Meanwhile, by Lemma 2.84, we have

$$\begin{aligned} |\psi_m|_V \bullet \cdots \bullet |\psi_1|_V &= |\mathrm{dom}(\varepsilon)|_{V(\varepsilon)}\overline{\psi} \\ \text{and} \quad \left|\psi'_n\right|_V \bullet \cdots \bullet \left|\psi'_1\right|_V &= |\mathrm{cod}(\varepsilon)|_{V(\varepsilon)}\overline{\psi} \end{aligned}$$

with $\overline{\psi} \colon |\mathbf{s}|_V \Rightarrow |\mathbf{s}'|_V$ given by $\overline{\psi} = \langle \mathrm{id}, \ldots, \underbrace{|\psi|_V}_{\ell^{\text{th}}\text{ entry}}, \ldots, \mathrm{id} \rangle$. So by interchange, both sides of our desired equation equal the horizontal composite $|\varepsilon| \circ \overline{\psi}$ depicted by

$$\begin{tikzcd} {|V|} \arrow[r, bend left, "|\mathbf{s}|_V"] \arrow[r, bend right, "|\mathbf{s}'|_V"'] \arrow[r, phantom, "\Downarrow \overline{\psi}"] & {|V(\varepsilon)|} \arrow[r, bend left, "|\mathrm{dom}(\varepsilon)|_{V(\varepsilon)}"] \arrow[r, bend right, "|\mathrm{cod}(\varepsilon)|_{V(\varepsilon)}"'] \arrow[r, phantom, "\Downarrow |\varepsilon|"] & {|\tau(\varepsilon)|,} \end{tikzcd}$$

which may be computed as the vertical composite of whiskered 2-cells in two different ways. Hence $|\cdot|_V$ agrees on R^{nat}-related pairs.

(d) For each $\mathfrak{R}$-relation

$$\nu = (\varphi_m^{\pm 1} \bullet \cdots \bullet \varphi_1^{\pm 1}) \ \mathfrak{R} \ (\psi_n^{\pm 1} \bullet \cdots \bullet \psi_1^{\pm 1}) = \nu',$$

where the φ_i's and ψ_j's are basic transformations, define the finite set of variables

$$W \coloneqq V(\nu) \cup V(\nu') = V(\varphi_1) \cup \cdots \cup V(\varphi_m) \cup V(\psi_1) \cup \cdots \cup V(\psi_n)$$

appearing in either ν or ν' as in Notation 2.42. Order and label the variables in W so that $W = \{x_{\ell_1}^{\mathtt{t}_1} < \cdots < x_{\ell_k}^{\mathtt{t}_k}\}$.

Now consider the functor $|\cdot|_W$ from (2.75) in Proposition 2.74, with W in place of V, arising from our given model $|\cdot|$ for $\mathfrak{C}$. By Definition 2.77, we have $|\nu|_W = |\nu'|_W$, so by functoriality and by (6) we have

$$|\varphi_m|_W^{\pm 1} \bullet \cdots \bullet |\varphi_1|_W^{\pm 1} = |\psi_n|_W^{\pm 1} \bullet \cdots \bullet |\psi_1|_W^{\pm 1}.$$

Then given $\mathbf{r} = (r_1, \ldots, r_k) \in \mathcal{L}_V(\mathfrak{S})^k$ with $\tau(r_\ell) = \mathtt{t}_\ell$ for all $1 \leq \ell \leq k$, we have by Lemma 2.80 that

$$|\varphi_i[W/\mathbf{r}]|_V^{\pm 1} = (|\varphi_i|_W^{\pm 1})_{|\mathbf{r}|_V} \quad \text{and} \quad |\psi_j[W/\mathbf{r}]|_V^{\pm 1} = (|\psi_j|_W^{\pm 1})_{|\mathbf{r}|_V},$$

so by vertical 2-functoriality of whiskering,

$$\begin{aligned} |\varphi_m[W/\mathbf{r}]|_V^{\pm 1} \bullet \cdots \bullet |\varphi_1[W/\mathbf{r}]|_V^{\pm 1} &= (|\varphi_m|_W^{\pm 1} \bullet \cdots \bullet |\varphi_1|_W^{\pm 1})_{|\mathbf{r}|_V} \\ &= (|\psi_n|_W^{\pm 1} \bullet \cdots \bullet |\psi_1|_W^{\pm 1})_{|\mathbf{r}|_V} \\ &= |\psi_n[W/\mathbf{r}]|_V^{\pm 1} \bullet \cdots \bullet |\psi_1[W/\mathbf{r}]|_V^{\pm 1}. \end{aligned}$$

Hence $|\cdot|_V$ agrees on the R^{com}-related pairs from Definition 2.46 (d) by functoriality of $|\cdot|_V$ and by (6).

Now suppose for induction we have basic transformations φ_i for $1 \leq i \leq m$ and ψ_j for $1 \leq j \leq n$ all with type $\mathtt{t}$ and scope V for which

$$|\varphi_m|_V^{\pm 1} \bullet \cdots \bullet |\varphi_1|_V^{\pm 1} = |\psi_n|_V^{\pm 1} \bullet \cdots \bullet |\psi_1|_V^{\pm 1}.$$

Fix a p-ary function $f \in F$ and $t_1, \ldots, t_{k-1}, t_{k+1}, \ldots, t_p \in \mathcal{L}_V(\mathfrak{S})$ so that $\mathrm{dom}(f) = (\tau(t_1), \ldots, \tau(t_{k-1}), \mathtt{t}, \tau(t_{k+1}), \ldots, \tau(t_p))$. For every basic transformation φ of type $\mathtt{t}$ and scope V, define a basic transformation $\mathcal{F}(\varphi) = f(\mathrm{id}_{t_1}, \ldots, \mathrm{id}_{t_{k-1}}, \varphi, \mathrm{id}_{t_{k+1}}, \ldots, \mathrm{id}_{t_p})$, also of scope V. Then

$$|\mathcal{F}(\varphi)|_V = |f|\langle \mathrm{id}_{|t_1|_V}, \ldots, \mathrm{id}_{|t_{k-1}|_V}, |\varphi|_V, \mathrm{id}_{|t_{k+1}|_V}, \ldots, \mathrm{id}_{|t_p|_V} \rangle,$$

so by vertical 2-functoriality of 2-products and whiskering,

$$\begin{aligned}|\mathcal{F}(\varphi_m)|_V^{\pm 1} \bullet \cdots \bullet |\mathcal{F}(\varphi_1)|_V^{\pm 1} &= |f|\langle \mathrm{id}_{|t_1|_V}, \ldots, \underbrace{|\varphi_m|_V^{\pm 1} \bullet \cdots \bullet |\varphi_1|_V^{\pm 1}}_{k^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}_{|t_p|_V}\rangle \\ &= |f|\langle \mathrm{id}_{|t_1|_V}, \ldots, \underbrace{|\psi_n|_V^{\pm 1} \bullet \cdots \bullet |\psi_1|_V^{\pm 1}}_{k^{\text{th}} \text{ entry}}, \ldots, \mathrm{id}_{|t_p|_V}\rangle \\ &= |\mathcal{F}(\psi_n)|_V^{\pm 1} \bullet \cdots \bullet |\mathcal{F}(\psi_1)|_V^{\pm 1}.\end{aligned}$$

It follows by induction that $|\cdot|_V$ agrees on $R(\mathfrak{C})_V$-related pairs. Hence the functor $|\cdot|_V$ from (2.75) with domain $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ descends uniquely to a functor (2.86) with domain $\mathscr{P}_V(\mathfrak{C})$, and this new functor satisfies (1–5) as written. The uniqueness of the functor (2.86) satisfying (1–5) follows from the the universal property of quotient categories, the uniqueness result of Proposition 2.74, and the fact that the quotient functor $[\cdot]\colon F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}]) \to \mathscr{P}_V(\mathfrak{C})$ necessarily satisfies $[\varphi^{-1}] = [\varphi]^{-1}$ for all $\varphi \in \mathrm{Bs}_{V,\mathfrak{I}}$ by the construction of $R^{\mathrm{inv}} \subseteq R(\mathfrak{C})_V$.

Finally, every morphism in $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}^{-1}}]$ is the composite of edges in Bs_V and isomorphisms. Since monomorphisms are closed under composition and every isomorphism is a monomorphism, the last statement follows from Proposition A.13. □

Our theorem implies a more general result for any restricted coherence schema. We repeat the conditions for general restricted coherence schemata for convenience.

Corollary 2.87. *Following Definition 2.51, let $\mathscr{P}(\mathfrak{C}, V, \mathfrak{O})$ be the category presented by the restricted coherence schema $(\mathfrak{C}, V, \mathfrak{O})$, and let $|\cdot|\colon \mathfrak{C} \to \underline{\mathbf{K}}$ be a model for $\mathfrak{C}$ in the cartesian monoidal 2-category $\underline{\mathbf{K}}$. Then there is a unique functor*

$$|\cdot|_V\colon \mathscr{P}(\mathfrak{C}, V, \mathfrak{O}) \to \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|) \tag{2.88}$$

such that

(1) $\left|x_k^{\mathtt{t}}\right|_V$ is the projection $|V| \to |\mathtt{t}|$ to the $x_k^{\mathtt{t}}$-factor, with $|\mathtt{t}|$ given by the model, for each variable $x_k^{\mathtt{t}} \in V \cap \mathfrak{O}$;

(2) $|c|_V$ is the constant morphism $|V| \to * \xrightarrow{|c|} |\tau(c)|$, with $*$ the terminal object of $\underline{\mathbf{K}}$ and $|c|\colon * \to |\tau(c)|$ given by the model, for each constant $c \in C \cap \mathfrak{O}$;

(3) $|f(s_1,\dots,s_n)|_V = |f| \circ |\mathbf{s}|_V$ *for an n-ary function symbol $f \in F$ and n-tuple of terms $\mathbf{s} = (s_1,\dots,s_n)$ from $\mathcal{L}(\mathfrak{S})_V$ for which $\mathrm{dom}(f) = (\tau(s_1),\dots,\tau(s_n))$ such that $f(s_1,\dots,s_n) \in \mathfrak{O}$, with $|f|$ as given by the model and $|\mathbf{s}|_V = \langle |s_1|_V,\dots,|s_n|_V\rangle$;*

(4) given atomic transformation $\varepsilon_{\mathbf{s}}\colon \mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}] \Rightarrow \mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ for edge label $\varepsilon \in \mathcal{E}$ and terms $\mathbf{s} = (s_1,\dots,s_n)$ from $\mathcal{L}(\mathfrak{S})_V$ with compatible types, if both $\mathrm{dom}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ and $\mathrm{cod}(\varepsilon)[V(\varepsilon)/\mathbf{s}]$ are in $\mathfrak{O}$ and $[\varepsilon_{\mathbf{s}}]$ is the residue of $\varepsilon_{\mathbf{s}}$ in the quotient category $\mathscr{P}_V(\mathfrak{C})$, then

$$|[\varepsilon_{\mathbf{s}}]|_V = |\varepsilon|_{|\mathbf{s}|_V}$$

as in Definition 2.72 (1), with $|\varepsilon|$ given by the model; and

(5) given basic transformation $f(\mathrm{id}_{r_1},\dots,\varphi,\dots,\mathrm{id}_{r_n})$ for terms $r_1,\dots,r_n \in \mathcal{L}(\mathfrak{S})_V$, basic transformation φ between terms in $\mathcal{L}(\mathfrak{S})_V$, and n-ary function symbol $f \in F$ with $\mathrm{dom}(f) = (\tau(r_1),\dots,\tau(\varphi),\dots,\tau(r_n))$ such that

$$f(r_1,\dots,\mathrm{dom}(\varphi),\dots,r_n) \quad \text{and} \quad f(r_1,\dots,\mathrm{cod}(\varphi),\dots,r_n)$$

are both in $\mathfrak{O}$, the functor $|\cdot|_V$ applied to the basic transformation's residue in $\mathscr{P}_V(\mathfrak{C})$ is given by

$$|[f(\mathrm{id}_{r_1},\dots,\varphi,\dots,\mathrm{id}_{r_n})]|_V = |f|\langle \mathrm{id},\dots,|\varphi|_V,\dots,\mathrm{id}\rangle$$

as in Definition 2.72 (2), with $|f|$ given by the model.

Furthermore, the functor $|\cdot|_V$ factors through the canonical functors

$$\mathscr{P}(\mathfrak{C},V,\mathfrak{O}) \to \mathscr{P}(\mathfrak{C},V(\mathfrak{O}),\mathfrak{O})$$

Moreover, if the functor $|\cdot|_V$ in (2.88) sends the residue of every basic transformation in $\mathrm{Bs}_{\mathfrak{O}}$ to a monomorphism, then $|\cdot|_V$ factors through the quotient functor $\mathscr{P}(\mathfrak{C},V,\mathfrak{O}) \to \mathscr{P}^{\mathrm{L}}(\mathfrak{C},V,\mathfrak{O})$ from Definition 2.51.

Proof. Restricting the domain $\mathscr{P}_V(\mathfrak{C})$ of the functor $|\cdot|_V$ from (2.86) in Theorem 2.85 to its subcategory $\mathscr{P} := \mathscr{P}(\mathfrak{C},V,\mathfrak{O})$ yields the desired functor (2.88) satisfying the given conditions. Conditions (1–3) are satisfied as every term in $\mathfrak{O}$ is an object of $\mathscr{P}$, and conditions (4–5) are satisfied as every basic transformation φ between terms in $\mathfrak{O}$ gives rise to a morphism $[\varphi]$ in $\mathscr{P}$.

For uniqueness, compose a functor (2.88) with quotient functor $[\cdot]\colon F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}^{-1}]) \to \mathscr{P}$, which satisfies $[\varphi^{-1}] = [\varphi]^{-1}$ for all $\varphi \in \mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}$ because of the construction of $R^{\mathrm{inv}} \subseteq R(\mathfrak{C})_V$ in Definition 2.46 (a). This combined with our five given conditions on (2.88) imply that the composite functor $|[\cdot]|_V$ satisfies the six conditions on (2.76) given in Proposition 2.74. So uniqueness follows from the uniqueness result in Proposition 2.74. Finally, the last statement once again follows from Proposition A.13. $\square$

Typically we will work with the minimal scope $V(\mathfrak{O})$ in place of an arbitrary set of variables V containing $V(\mathfrak{O})$. It turns out the functor from (2.88) is related to the same functor but with $V(\mathfrak{O})$ in place of V as follows (recall from Definition 2.51 that we denote $\mathscr{P}(\mathfrak{C}, V(\mathfrak{O}), \mathfrak{O})$ simply by $\mathscr{P}(\mathfrak{C}, \mathfrak{O})$).

Proposition 2.89. *Given a restricted coherence schema $(\mathfrak{C}, V, \mathfrak{O})$ and a model $|\cdot|\colon \mathfrak{C} \to \underline{\mathbf{K}}$ for $\mathfrak{C}$ in the cartesian monoidal 2-category $\underline{\mathbf{K}}$, there is a commutative square*

$$\begin{tikzcd}
\mathscr{P}(\mathfrak{C}, \mathfrak{O}) = \mathscr{P}(\mathfrak{C}, V(\mathfrak{O}), \mathfrak{O}) \arrow[r, "|\cdot|_{V(\mathfrak{O})}"] \arrow[d] & \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V(\mathfrak{O})|, |\mathtt{t}|) \arrow[d] \\
\mathscr{P}(\mathfrak{C}, V, \mathfrak{O}) \arrow[r, "|\cdot|_V"] & \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V|, |\mathtt{t}|),
\end{tikzcd}$$

where the top arrow is the unique functor from (2.88); the bottom arrow is the unique functor from (2.88) with the minimal $V(\mathfrak{O})$ in place of V; the left arrow is the quotient functor induced by the inclusion $R(\mathfrak{C})_{V(\mathfrak{O})} \subseteq R(\mathfrak{C})_V$; and the right arrow is given by whiskering with the projection $\pi\colon |V| \to |V(\mathfrak{O})|$.

Proof. This follows directly from the commutativity of (2.83), from Corollary 2.87, and from the fact that quotient functors are epic. $\square$

Chapter 3

COHERENT NORMALIZATION PROBLEMS

In Chapter 2, we introduced formalisms leading to coherence schemata in Definition 2.43, and we explained how Mac Lane-style presentations of categorical structures could be translated as models of such coherence schemata, culminating in Definition 2.77. Furthermore, in Definition 2.51, we defined the universal parameter categories $\mathscr{P} := \mathscr{P}(\mathfrak{C}, \mathfrak{O})$ and $\mathscr{P}^{\mathrm{L}} := \mathscr{P}^{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$ associated to any restricted coherence schema $(\mathfrak{C}, \mathfrak{O})$, and in Theorem 2.85, we explained how to construct interpretation functors from $\mathscr{P}$ and $\mathscr{P}^{\mathrm{L}}$ into certain mapping categories associated with a given model of $\mathfrak{C}$. The objects of $\mathscr{P}$ and $\mathscr{P}^{\mathrm{L}}$, given by $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})_V$, parametrize a choice of operations on the model; while the morphisms of $\mathscr{P}$ and $\mathscr{P}^{\mathrm{L}}$ parametrize natural transformations between the chosen operations.

In general, a categorical coherence theorem asserts that some class of diagrams commute. In our framework, such theorems may sometimes be recast as the assertion that some or all of the hom-sets in $\mathscr{P}$ or $\mathscr{P}^{\mathrm{L}}$ are singletons: if $|\mathscr{P}(s,t)| = 1$, then any transformations from s to t composed of basic transformations and their inverses have the same interpretation, regardless of the model of $\mathfrak{C}$ we are working in. A more refined result, like the coherence theorem for braided monoidal categories, would identify the congruence classes in $\mathscr{P}(s,t)$ using properties of their representatives, but we shall not pursue this line of inquiry here.

Thus, the problem is to understand the structure of the categories $\mathscr{P}$ and $\mathscr{P}^{\mathrm{L}}$ associated to a given restricted coherence schema $(\mathfrak{C}, \mathfrak{O})$. In fact, we will work in greater generality: both $\mathscr{P}$ and $\mathscr{P}^{\mathrm{L}}$ are categories of the form $F(\mathscr{G}[I^{-1}])/\sim$, where $\mathscr{G}$ is a directed graph, I is a marked set of edges in $\mathscr{G}$, and $\sim$ is a congruence relation on the free category $F(\mathscr{G}[I^{-1}])$. It turns out that characterizing the quotient $F(\mathscr{G}[I^{-1}])/\sim$ is often closely related to the problem of finding suitable normal forms for its objects. We call these problems *coherent normalization problems (CNPs)*, which we introduce in the next section and study throughout this chapter.

3.1 Introducing CNPs and their solutions

Definition 3.1 (Coherence normalization problem & solution)**.** A *coherent normalization problem* (or *CNP* for short) is a triple $\mathfrak{G} = (\mathscr{G}, I, \sim)$ such that:

(a) $\mathscr{G} = [\mathrm{dom}, \mathrm{cod}\colon E \rightrightarrows V]$ is a directed graph, called the *underlying graph of* $\mathfrak{G}$;

(b) $I \subseteq E$ is a set of edges in $\mathscr{G}$ (i.e. $(\mathscr{G}, I)$ is a marked directed graph) called the *marked edges of* $\mathfrak{G}$; and

(c) $\sim$ is a congruence relation on the free category $F(\mathscr{G}[I^{-1}])$, called the *congruence relation of* $\mathfrak{G}$, such that for any edge $i \in I$, denoted $i\colon \mathrm{dom}(i) \Rightarrow \mathrm{cod}(i)$, we have $(i^{-1} \bullet i) \sim \mathrm{id}_{\mathrm{dom}(i)}$ and $(i \bullet i^{-1}) \sim \mathrm{id}_{\mathrm{cod}(i)}$.[1]

We call $\mathscr{P} \coloneqq F(\mathscr{G}[I^{-1}])/\sim$ the *category presented by* $\mathfrak{G}$.

A *solution to the CNP* $\mathfrak{G}$ is a set of vertices $N \subseteq V = \mathrm{Ob}(F(\mathscr{G}[I^{-1}]))$ such that for any vertex $t \in V$:

(1) there is a unique $n(t) \in N$ such that there exists a morphism $\nu\colon t \Rightarrow n(t)$ in $F(\mathscr{G}[I^{-1}])$ (or equivalently in $\mathscr{P}$); and moreover

(2) if $\nu, \nu'\colon t \Rightarrow n(t)$ are two morphisms in $F(\mathscr{G}[I^{-1}])$, then $\nu \sim \nu'$ (or equivalently $\nu = \nu'$ for two such morphisms in $\mathscr{P}$).

If N is a solution to $\mathfrak{G}$, we call $n(t)$ the *N-normal form of t*. Thus (1) asserts the existence and uniqueness of normal forms, while (2) asserts the uniqueness of normalization maps up to congruence.

We typically think of the edges in E as basic transformations and those in I as invertible basic transformations. Then the morphisms of $F(\mathscr{G}[I^{-1}])$ are more general transformations consisting of sequences of composable transformations from E and inverted transformations

[1] We use $\Rightarrow$ arrows for morphisms and $\bullet$ for composition in $F(\mathscr{G}[I^{-1}])/\sim$ because in our applications, they will correspond to natural transformations and vertical composition.

from I, while the congruence relation $\sim$ identifies certain pairs of these transformations as equivalent. Each morphism $t \Rightarrow t'$ records a way to transform t to t'.

We are interested in ways to explicitly construct and verify solutions to CNPs. First, we may immediately deduce the following lemma about normal forms.

Lemma 3.2. *Let $\mathscr{G} = [E \rightrightarrows V]$ be a directed graph and $\mathfrak{G} = (\mathscr{G}, I, \sim)$ be a CNP with solution $N \subseteq V$. If there is an edge $e\colon t \Rightarrow u$ in $\mathscr{G}$, then t and u have the same N-normal form: $n(t) = n(u)$. Moreover, morphisms $\nu\colon t \Rightarrow n(t)$ and $\nu'\colon u \Rightarrow n(u)$ satisfy $\nu \sim (\nu' \bullet e)$.*

Proof. Let ν' be any morphism $u \Rightarrow n(u)$ in $F(\mathscr{G}[I^{-1}])$ exhibiting $n(u)$ as the N-normal form of u. Then we have the composite $(\nu' \bullet e)\colon t \Rightarrow u \Rightarrow n(u) \in N$, so by the uniqueness of N-normal forms, $n(t) = n(u)$. Furthermore, given any other morphism $\nu\colon t \Rightarrow n(t) = n(u)$, we have $\nu \sim (\nu' \bullet e)$. □

The importance of finding solutions to CNPs is summarized below.

Theorem 3.3. *Suppose that $\mathfrak{G} = (\mathscr{G}, I, \sim)$ is a CNP and that N is a solution to $\mathfrak{G}$. Let $\mathscr{P} = F(\mathscr{G}[I^{-1}])/\sim$ be the category presented by $\mathfrak{G}$, and for every vertex $n \in N$, let $\mathscr{P}[n]$ denote the connected component category of $n \in \mathrm{Ob}(\mathscr{P})$ in $\mathscr{P}$. Then*

$$\mathscr{P} \cong \coprod_{n \in N} \mathscr{P}[n]$$

and for each $n \in N$, the object n is terminal in $\mathscr{P}[n]$.

Proof. The inclusions $\mathscr{P}[n] \hookrightarrow \mathscr{P}$ induce a functor $F\colon \coprod_{n \in N} \mathscr{P}[n] \to \mathscr{P}$, which we must prove is an isomorphism.

To start, we prove that for each $m \in N$, the object m is terminal in $\mathscr{P}[m]$. Given $t \in \mathscr{P}[m]$, there is a zigzag of morphisms of $\mathscr{P}$ connecting t to m, which lifts to a zigzag in $F(\mathscr{G}[I^{-1}])$. That is, there are vertices $t = t_0, \ldots, t_k = m$ such that for each $0 \leq j < k$, there is a morphism in $F(\mathscr{G}[I^{-1}])$ that is either $t_j \Rightarrow t_{j+1}$ or $t_{j+1} \Rightarrow t_j$. Either way, by Lemma 3.2, $n(t_j) = n(t_{j+1})$. So by induction

$$n(t) = n(t_0) = \cdots = n(t_k) = n(m) = m,$$

with the final equality exhibited by the identity morphism $\mathrm{id}_m\colon m \Rightarrow m \in N$. In particular, there is a morphism $\nu\colon t \Rightarrow n(t) = m$ in $F(\mathscr{G}[I^{-1}])$ and hence a morphism $[\nu]\colon t \Rightarrow m$ in $\mathscr{P}$. If $[\nu']\colon t \Rightarrow m$, then $\nu, \nu'\colon t \Rightarrow m \in N$ in $F(\mathscr{G}[I^{-1}])$, and hence $\nu \sim \nu'$ by Definition 3.1 (2). Thus $[\nu] = [\nu']$, which proves that m is terminal in $\mathscr{P}[m]$.

Next, we prove that $F\colon \coprod_{n\in N} \mathscr{P}[n] \to \mathscr{P}$ is bijective on objects. For surjectivity, suppose $t \in \mathrm{Ob}(\mathscr{P})$ and let $m = n(t) \in N$. Then there exists a morphism $t \Rightarrow m$ in $F(\mathscr{G}[I^{-1}])$ and thus in $\mathscr{P}$, so $t \in \mathscr{P}[m]$ is in the image of F. For injectivity, it is immediate that F is injective on each component $\mathscr{P}[n]$, so it suffices to show that $\mathscr{P}[n]$ and $\mathscr{P}[n']$ are disjoint if $n \neq n'$ for $n, n' \in N$. However, if $t \in \mathscr{P}[n] \cap \mathscr{P}[n']$, then by our previous argument $n = n(t) = n'$. Thus F is bijective on objects.

Finally, we show that F is fully faithful. As each inclusion functor is fully faithful, it suffices to show that, as in $\coprod_{n\in N} \mathscr{P}[n]$, there are no morphisms $s \Rightarrow t$ in $\mathscr{P}$ whenever $s \in \mathscr{P}[n]$ and $t \in \mathscr{P}[n']$ for $n, n' \in N$ with $n \neq n'$. Indeed, if there were, n and n' would be in the same connected component, yet we saw previously that $\mathscr{P}[n]$ and $\mathscr{P}[n']$ are necessarily disjoint. Hence F is fully faithful. $\square$

Passing to topology, we obtain the following result. Let $B\mathscr{C}$ denote the classifying space (i.e. the geometric realization of the nerve) of a category $\mathscr{C}$.

Corollary 3.4. *Given the notation and hypotheses of Theorem 3.3, the classifying space*

$$B\mathscr{P} \cong \coprod_{n\in N} B\mathscr{P}[n],$$

and each of the component spaces $B\mathscr{P}[n]$ is contractible.

Proof. The classifying space functor preserves coproducts and sends categories with terminal objects to contractible spaces. $\square$

If $\mathfrak{G} = (\mathscr{G}, I, \sim)$ is a CNP and the congruence relation $\sim$ is left-cancelable, e.g. if $\mathfrak{G}$ is the CNP presenting $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}, \mathfrak{D})$, then the existence of a solution to $\mathfrak{G}$ has even stronger implications for the category presented by $\mathfrak{G}$.

Theorem 3.5. *Suppose that $\mathfrak{G} = (\mathscr{G}, I, \sim)$ is a CNP and that N is a solution to $\mathfrak{G}$. Let $F(\mathscr{G}[I^{-1}])/\!\sim\; = \mathscr{P} \cong \coprod_{n\in N} \mathscr{P}[n]$ as in Theorem 3.3.*

(1) If $\sim$ is left-cancelable, then for each $n \in N$, $\mathscr{P}[n]$ is a preorder category: every diagram in it commutes.

(2) If I is the entire edge-set of $\mathscr{G}$, then $\sim$ is left-cancelable, and the category $\mathscr{P}[n]$ is equivalent to the terminal category $*$ for each $n \in N$. In other words, the inclusion functor from N viewed as a discrete category into $\mathscr{P}$ is an equivalence.

Proof. For (1), by Proposition A.13, every morphism in $\mathscr{P}$ is monic. Then given two parallel morphisms $f, f' \colon t \rightrightarrows u$ in some $\mathscr{P}[n]$, by Theorem 3.3, n is terminal in $\mathscr{P}[n]$, so there exists a morphism $! \colon u \to n$. Terminality of n further implies that $! \bullet f = ! \bullet f'$, so since $!$ is monic, $f = f'$. Hence $\mathscr{P}[n]$ is a preorder category.

For (2), every edge in $\mathscr{G}[I^{-1}]$ is invertible as a morphism in $F(\mathscr{G}[I^{-1}])$, and every morphism in $F(\mathscr{G}[I^{-1}])$ is composed of edges, so every morphism in $F(\mathscr{G}[I^{-1}])$ is invertible. Moreover, given two parallel morphisms $f, f' \colon t \rightrightarrows u$ and a morphism $g \colon u \to v$ in $F(\mathscr{G}[I^{-1}])$ with $g \bullet f \sim g \bullet f'$, since g is invertible, we have $f \sim f'$. Hence $\sim$ is left-cancelable, so by (1), each $\mathscr{P}[n]$ is a preorder category. As each $\mathscr{P}[n]$ is also a groupoid with terminal object n, every object $w \in \mathscr{P}[n]$ has an invertible map $w \to n$, making $\mathscr{P}[n]$ equivalent to a terminal category. $\square$

Part (1) then allows for coherence theorems in the style of Laplaza's theorem from [13] and [9], where the interpretations of certain basic transformations are assumed to be merely monomorphisms instead of isomorphisms, while part (2) encapsulates the most familiar coherence theorems where the interpretation of every basic transformation is invertible.

We may now phrase the problem of proving a coherence theorem for a restricted coherence schema in terms of solving a corresponding CNP. To that end, we define the following CNPs for a given restricted coherence schema $(\mathfrak{C}, \mathfrak{O})$.

Notation 3.6 (CNPs for restricted coherence schemata)**.** Let $\mathfrak{C} = (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R})$ be a coherence schema and $\mathfrak{O} \subseteq \mathcal{L}(\mathfrak{S})$ be a variable-finite, subterm-closed subset. Then the CNPs given by the restricted coherence schema $(\mathfrak{C}, \mathfrak{D})$ are

$$\begin{aligned} \mathrm{CN}(\mathfrak{C}, \mathfrak{D}) &:= \Big(\mathscr{B}_{\mathfrak{D}}, \mathrm{Bs}_{\mathfrak{D},\mathfrak{J}}, \langle R(\mathfrak{C})_{V(\mathfrak{D})}\rangle\Big) \\ \text{and} \quad \mathrm{CN}_{\mathrm{L}}(\mathfrak{C}, \mathfrak{D}) &:= \Big(\mathscr{B}_{\mathfrak{D}}, \mathrm{Bs}_{\mathfrak{D},\mathfrak{J}}, \langle R(\mathfrak{C})_{V(\mathfrak{D})}\rangle^{\mathrm{L}}\Big), \end{aligned}$$

where

- the graph $\mathscr{B}_{\mathfrak{D}}$ has the terms in $\mathfrak{D}$ as vertices and the basic transformations $\mathrm{Bs}_{\mathfrak{D}}$ generated by $\mathfrak{T}$ as edges, as defined in Definition 2.32;

- the subset $\mathrm{Bs}_{\mathfrak{D},\mathfrak{J}} \subseteq \mathrm{Bs}_{\mathfrak{D}}$ consists of all $\mathfrak{J}$-marked basic transformations, as defined in Definition 2.40; and

- the local binary relation $R(\mathfrak{C})_{V(\mathfrak{D})}$ is defined on the free category $F(\mathscr{B}_{V(\mathfrak{D})}[\mathrm{Bs}^{-1}_{V(\mathfrak{D}),\mathfrak{J}}])$ (where $\mathscr{B}_{V(\mathfrak{D})}$ is the graph of basic transformations between terms with scope $V(\mathfrak{D})$ and $\mathrm{Bs}_{V(\mathfrak{D}),\mathfrak{J}}$ is its set of $\mathfrak{J}$-marked edges) as in Definition 2.46, yielding congruences $\langle R(\mathfrak{C})_{V(\mathfrak{D})}\rangle$ and $\langle R(\mathfrak{C})_{V(\mathfrak{D})}\rangle^{\mathrm{L}}$ that are subsequently restricted to the subcategory $F(\mathscr{B}_{\mathfrak{D}}[\mathrm{Bs}^{-1}_{\mathfrak{D},\mathfrak{J}}]) \subseteq F(\mathscr{B}_{V(\mathfrak{D})}[\mathrm{Bs}^{-1}_{V(\mathfrak{D}),\mathfrak{J}}])$.

Note that the inclusion of R^{inv} in $R(\mathfrak{C})_{V(\mathfrak{D})}$ in Definition 2.46 (a) guarantees that Definition 3.1 (c) is satisfied. These CNPs present $\mathscr{P}(\mathfrak{C}, \mathfrak{D})$ and $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}, \mathfrak{D})$, respectively, from Definition 2.51.

We specialize the previous results (Theorems 3.3 and 3.5 and Corollary 3.4) to these CNPs presenting our universal parameter categories, then leverage Corollary 2.87 to prove our desired coherence theorems.

Theorem 3.7. *Using Notation 3.6, let $\mathfrak{N} \subseteq \mathfrak{D}$ be a solution to $\mathrm{CN}(\mathfrak{C}, \mathfrak{D})$ presenting $\mathscr{P} := \mathscr{P}(\mathfrak{C}, \mathfrak{D})$, and let $|\cdot| \colon \mathfrak{C} \to \underline{\mathbf{K}}$ be a model for $\mathfrak{C}$ in a cartesian monoidal 2-category $\underline{\mathbf{K}}$, giving rise to the functor*

$$|\cdot| := |\cdot|_{V(\mathfrak{D})} \colon \mathscr{P} \to \coprod_{\mathtt{t} \in T} \underline{\mathbf{K}}(|V(\mathfrak{D})|, |\mathtt{t}|)$$

from (2.88) (with the minimal $V(\mathfrak{D})$ in place of V). Then we have the following.

(1) For every term $t \in \mathfrak{O}$, there exists a unique term $n(t) \in \mathfrak{N}$ such that there is a morphism $t \Rightarrow n(t)$ in $\mathscr{P}$. Moreover, there is a unique morphism $\nu\colon t \Rightarrow n(t)$ in $\mathscr{P}$. In particular, $\tau(t) = \tau(n(t))$. So the classifying space $B\mathscr{P}$ is a coproduct over $\mathfrak{N}$ of contractible spaces.

Consequently, any pair of parallel morphisms η_1, η_2 in $\coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V(\mathfrak{O})|, |\mathtt{t}|)$ that lifts via $|\cdot|$ to a pair of parallel morphisms $t \rightrightarrows n(t)$ in $\mathscr{P}$ satisfies $\eta_1 = \eta_2$.

(2) Assume every edge label of our coherence schema is marked, so $\mathfrak{I} = \mathcal{E}$. Then every diagram in $\mathscr{P}$ commutes, and the inclusion from $\mathfrak{N}$ viewed as a discrete category to $\mathscr{P}$ is an equivalence. That is, given terms $t, t' \in \mathfrak{O}$ with $n(t) = n(t')$, there exists a unique morphism $t \Rightarrow t'$ in $\mathscr{P}$.

Consequently, every diagram in $\coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V(\mathfrak{O})|, |\mathtt{t}|)$ that factors through the functor $|\cdot|$ is a commutative diagram of isomorphisms.

Now assume instead $\mathfrak{N}_{\mathrm{L}} \subseteq \mathfrak{O}$ is a solution[2] *to $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$ presenting $\mathscr{P}^{\mathrm{L}} := \mathscr{P}^{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$.*

(3) The category $\mathscr{P}^{\mathrm{L}}$ is a preorder category: every diagram in it commutes.

If, furthermore, $|\cdot|$ sends the residue of every basic transformation in $\mathrm{Bs}_{\mathfrak{O}}$ to a monomorphism, then $|\cdot|$ factors through $\mathscr{P}^{\mathrm{L}}$, so any diagram in $\coprod_{\mathtt{t}\in T} \underline{\mathbf{K}}(|V(\mathfrak{O})|, |\mathtt{t}|)$ that factors through the functor $|\cdot|$ commutes.

In short, by Corollary 2.87, every model $|\cdot|\colon \mathfrak{C} \to \underline{\mathbf{K}}$ (for a cartesian monoidal 2-category $\underline{\mathbf{K}}$ such as $\underline{\mathbf{Cat}}$) extends uniquely to a functor out of the category $\mathscr{P}(\mathfrak{C}, \mathfrak{O})$ (and, under certain conditions, to $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$), so the theorem above actually yields a coherence theorem for every model of $\mathfrak{C}$ in $\underline{\mathbf{K}}$. Finding a solution to the CNP presenting the universal parameter category $\mathscr{P}(\mathfrak{C}, \mathfrak{O})$ (or the monomorphic parameter category $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$) then directly implies a coherence theorem for models of $\mathfrak{C}$ restricted to $\mathfrak{O}$. In the remainder of this chapter, we shall develop general techniques for constructing solutions to CNPs and apply them to prove coherence theorems for familiar categorical structures.

---

[2] In fact, we will see in all of our examples that the solution to $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$ is just the solution to $\mathrm{CN}(\mathfrak{C}, \mathfrak{O})$ by Corollary 3.12, although there may be cases where $\mathrm{CN}(\mathfrak{C}, \mathfrak{O})$ has no solution but $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$ does.

3.2 Constructing solutions to CNPs

We begin with a simple example of a CNP that is guaranteed to have a solution by the axiom of choice.

Example 3.8. Let $\mathscr{G} = [E \rightrightarrows V]$ be a graph and $\mathfrak{G} = (\mathscr{G}, E, \sim)$ be a CNP with every edge marked and $\sim$ maximal, so that the category $\mathscr{P}$ presented by $\mathfrak{G}$ is the preorder reflection of $F(\mathscr{G}[E^{-1}])$, with a unique morphism $t \Rightarrow t'$ for vertices $t, t' \in V$ whenever t and t' are connected by a zigzag of paths in $\mathscr{G}$. Then there exists a solution N to $\mathfrak{G}$ comprised of one representative from each connected component of $\mathscr{G}$.

In general, however, there is no guarantee that a solution set N exists. The basic problems we shall consider going forward are:

(1) whether a given CNP has a solution, and

(2) how to convert a solution to one CNP into a solution to another CNP.

As solutions to a CNP are defined purely in terms of the category $\mathscr{P}$ it presents and not in terms of the edges in E and I that generate the free category, we shall be most interested in the following notion of a morphism between normalization problems.

Definition 3.9 (CNP morphism)**.** Suppose that $\mathfrak{G}_i = (\mathscr{G}_i, I_i, \sim_i)$ are CNPs for $i = 1, 2$. A *morphism* $f\colon \mathfrak{G}_1 \to \mathfrak{G}_2$ is a functor $f\colon F(\mathscr{G}_1[I_1^{-1}]) \to F(\mathscr{G}_2[I_2^{-1}])$ that preserves congruence relations, i.e. $\nu \sim_1 \nu'$ implies $f(\nu) \sim_2 f(\nu')$ for all parallel morphisms ν, ν' in $F(\mathscr{G}_1[I_1^{-1}])$. We write **CN** for the category of CNPs and these morphisms.

The next result follows directly and justifies our definition of morphisms of CNPs.

Proposition 3.10. *Let $\mathscr{G}_i = [E_i \rightrightarrows V_i]$ be a graph and $\mathfrak{G}_i = (\mathscr{G}_i, I_i, \sim_i)$ be a CNP for $i = 1, 2$, and let $f\colon \mathfrak{G}_1 \to \mathfrak{G}_2$ be a morphism that is surjective on objects and full. Then f preserves solutions: if N is a solution to $\mathfrak{G}_1$, then $f(N)$ is a solution to $\mathfrak{G}_2$. In particular, isomorphisms of CNPs preserve solutions.*

Proof. Since f is surjective on objects, each vertex in V_2 is $f(t)$ for some $t \in V_1$. As N is a solution to $\mathfrak{G}_1$, there is a morphism $\nu\colon t \Rightarrow n(t) \in N$ in $F(\mathscr{G}[I_1^{-1}])$ that f sends to a morphism $f(\nu)\colon f(t) \Rightarrow f(n(t)) \in f(N)$ in $F(\mathscr{G}[I_2^{-1}])$.

Now given $n' \in N$ for which there is a morphism $\xi\colon f(t) \Rightarrow f(n')$, fullness means there is a morphism $\nu'\colon t \Rightarrow n'$ with $f(\nu') = \xi$. Then $n(t) = n'$ and $\nu \sim_1 \nu'$, so $f(n(t)) = f(n')$ and $f(\nu) \sim_2 f(\nu') = \xi$. Hence $f(N)$ is a solution to $\mathfrak{G}_2$.

If $f\colon \mathfrak{G}_1 \to \mathfrak{G}_2$ is an isomorphism, then $f\colon F(\mathscr{G}[I_1^{-1}]) \to F(\mathscr{G}[I_2^{-1}])$ is an isomorphism of categories, so it is necessarily surjective on objects and full. $\square$

Corollary 3.11. *Let $\mathscr{G} = [E \rightrightarrows V]$ be a graph and $\mathfrak{G}_i = (\mathscr{G}, I, \sim_i)$ be a CNP for $i = 1, 2$ with $\sim_1 \subseteq \sim_2$: that is, every $\sim_1$-related pair of morphisms in $F(\mathscr{G}[I^{-1}])$ is $\sim_2$-related. Then if N is a solution to $\mathfrak{G}_1$, it is also a solution to $\mathfrak{G}_2$.*

Proof. As $\sim_1 \subseteq \sim_2$, the identity functor on $F(\mathscr{G}[I^{-1}])$ is a morphism $\mathfrak{G}_1 \to \mathfrak{G}_2$. Since the identity functor is both surjective on objects and full, it preserves solutions. $\square$

We can apply this result to the CNPs for our restricted coherence schema as defined in Notation 3.6.

Corollary 3.12. *If $\mathfrak{N} \subseteq \mathfrak{O}$ is a solution to $\mathrm{CN}(\mathfrak{C}, \mathfrak{O})$, then $\mathfrak{N}$ is also a solution to $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}, \mathfrak{O})$.*

Proof. This follows from the previous corollary, as $\langle R(\mathfrak{C})_{V(\mathfrak{O})}\rangle \subseteq \langle R(\mathfrak{C})_{V(\mathfrak{O})}\rangle^{\mathrm{L}}$. $\square$

In the remainder of this section and the next, we shall examine when solutions $N \subseteq V$ to a CNP $\mathfrak{G}$ exist and how they may be constructed. We will introduce several relevant conditions on the graph $\mathscr{G}$. These generalize familiar properties for relations, which we will review as well, thinking of them as graphs with no parallel edges.

Definition 3.13 (Minimality, well-foundedness, confluence)**.** Let $\mathscr{G} = [E \rightrightarrows V]$ be a directed graph. For consistency with the graphs constructed in Chapter 2, we shall denote vertices of $\mathscr{G}$ by letters such as s, t, u and edges of $\mathscr{G}$ by double-arrows such as $e\colon s \Rightarrow t$. We say that $t \in V$ is:

(1) *$\mathscr{G}$-minimal* (or just *minimal*) if there is no $u \in V$ for which an edge $t \Rightarrow u$ exists in $\mathscr{G}$, i.e. there are no edges out of t. (We say *minimal* instead of *maximal* because we

typically think of an arrow $e\colon s \Rightarrow t$ as a "reduction" of the vertex s into a "simpler" vertex t.)

(2) *$\mathscr{G}$-minimal* (or just *minimal*) *relative to* A if $t \in A \subseteq V$ and there is no $u \in A$ for which an edge $t \Rightarrow u$ exists in $\mathscr{G}$, i.e. there are no edges from t to vertices in A.

(3) *$\mathscr{G}$-well-founded* (or just *well-founded*) if there is no infinite sequence $t = t_0, t_1, t_2, \ldots$ for which an edge $t_i \Rightarrow t_{i+1}$ exists in $\mathscr{G}$ for all $i \geq 0$, i.e. there are no infinite pathsout of t.

We say that $\mathscr{G}$ is *well-founded* if every $t \in V$ is well-founded, i.e. there are no infinite paths in $\mathscr{G}$ or, equivalently, every nonempty subset has a minimal element.

If further $\sim$ is a congruence relation on the free category $F(\mathscr{G})$, we say that t is:

(4) $(\mathscr{G}, \sim)$*-confluent* (or just *confluent*) if, for any morphisms $\gamma\colon t \Rightarrow a$ and $\delta\colon t \Rightarrow b$ in $F(\mathscr{G})$, there exist morphisms $\alpha\colon a \Rightarrow c$ and $\beta\colon b \Rightarrow c$ in $F(\mathscr{G})$ for which $(\alpha \bullet \gamma) \sim (\beta \bullet \delta)$:

$$\begin{tikzcd} & t \arrow[dl, Rightarrow, "\gamma"'] \arrow[dr, Rightarrow, "\delta"] & \\ a \arrow[dr, Rightarrow, "\exists\,\alpha"'] & \sim & b \arrow[dl, Rightarrow, "\exists\,\beta"] \\ & c. & \end{tikzcd}$$

We say that $(\mathscr{G}, \sim)$ is *confluent* if every $t \in V$ is confluent.

Now let V be a set and R be a binary relation on V. Then we define minimality and well-foundedness for (V, R) to coincide with minimality and well-foundedness for the graph $\mathscr{G} = [\pi_1, \pi_2\colon E \rightrightarrows V]$ with $E = \{(t, u) \in V \times V \mid t \, R \, u\}$. Explicitly, given $t \in V$, we say that t is:

(1) *R-minimal* if there is no $u \in V$ such that $t \, R \, u$.

(2) *R-minimal relative to* A if $t \in A \subseteq V$ and there is no $u \in A$ such that $t \, R \, u$.

(3) *R-well-founded* if there is no infinite sequence $t = t_0, t_1, t_2, \ldots$ such that $t_i \, R \, t_{i+1}$ for all $i \geq 0$.

We say that (V, R) is *well-founded* (or that R is a *well-founded* relation on V) if every $t \in V$ is R-well-founded.

Remark 3.14. Binary relations $\Rightarrow$ such that no infinite chains $x_0 \Rightarrow x_1 \Rightarrow x_2 \Rightarrow \cdots$ exist are sometimes called *noetherian*, in analogy to the ascending chain condition in algebra. However, we will often think of our edges $\Rightarrow$ as reductions, so we refer to graphs with such edges as well-founded.

The next two propositions are standard. They record the well-founded analogues of the well-ordering principle and induction for both relations and graphs.

Proposition 3.15. *Suppose that (V, R) is well-founded and let $>_R$ denote the transitive (but not reflexive) closure of R, so that $t >_R u$ if and only if there is a* positive integer *$k > 0$ and elements $t = t_0 \, R \, \cdots \, R \, t_k = u$ in V. Then the following hold.*

(1) *Every nonempty subset $A \subseteq V$ contains an element $t \in A$ that is $>_R$-minimal relative to A.*

(2) (Ordinary Induction) *Suppose $P \subseteq V$ such that*[3]

  (i) *for every R-minimal $t \in V$, we have $t \in P$, and*

  (ii) *for every $t \in V$ that is not R-minimal, if $u \in P$ for all $u \in V$ with $t \, R \, u$, then $t \in P$.*

  Then $P = V$.

(3) (Strong Induction) *Suppose $P \subseteq V$ such that*

  (i) *for every R-minimal $t \in V$, we have $t \in P$, and*

  (ii) *for every $t \in V$ that is not R-minimal, if $u \in P$ for all $u \in V$ with $t >_R u$, then $t \in P$.*

---

[3] Typically P is defined by a property $\varphi \colon V \to \{\bot, \top\}$ that may or may not be true for each element $v \in V$, so that $P \coloneqq \varphi^{-1}(\top)$. Then (2)(i) gives a base case and (2)(ii) gives an inductive step to verify that the property is true for every element of V, making $\varphi^{-1}(\top) = V$.

Then $P = V$.

Proof of (1). Suppose not. Fix $t_0 \in A$. Then for $k \geq 0$, given elements

$$t_0 >_R \cdots >_R t_k$$

in A, we know t_k is not $>_R$-minimal relative to A, so there exists $t_{k+1} \in A$ with $t_k >_R t_{k+1}$. Continuing recursively, we obtain an infinite sequence

$$t_0 >_R t_1 >_R t_2 >_R \cdots$$

in A. Each pair $t_i >_R t_{i+1}$ is exhibited by a sequence $t_i = t_i^0 \, R \, \cdots \, R \, t_i^{k_i} = t_{i+1}$ in V with $k_i > 0$, yielding an infinite sequence

$$t_0 = t_0^0 \, R \, \cdots \, R \, t_0^{k_0} = t_1 = t_1^0 \, R \, \cdots \, R \, t_1^{k_1} = t_2 = t_2^0 \, R \, \cdots$$

in V that contradicts the well-foundedness of (V, R). □

Proof of (2). Suppose for contradiction that $V \setminus P$ is nonempty. Then by (1), there exists $t \in V \setminus P$ that is $>_R$-minimal relative to $V \setminus P$. As $t \notin P$, (2)(i) implies that t is not R-minimal. Moreover, for all $u \in V$ with $t \, R \, u$, we have $t >_R u$, so $u \notin V \setminus P$ by the $>_R$-minimality of t relative to $V \setminus P$. Hence $u \in P$. Then by (2)(ii), we have $t \in P$, a contradiction. □

Proof of (3). Again suppose for contradiction $V \setminus P$ is nonempty and take $t \in V \setminus P$ that is $>_R$-minimal relative to $V \setminus P$, so $t \notin P$ and thus not R-minimal by (3)(i). For $u \in V$ with $t >_R u$, we have $u \notin V \setminus P$ by the $>_R$-minimality of t relative to $V \setminus P$, so $u \in P$. Then by (3)(ii), we have $t \in P$, a contradiction. □

Often when applying the above proposition, our set V will be the vertex-set of a graph $\mathscr{G}$ and our relation R will be defined by $t \, R \, u$ whenever there is an edge from t to u in that graph. In that case, the transitive closure $>_R$ of R is given by $t >_R u$ whenever there is a nonempty path from t to u in $\mathscr{G}$, i.e. whenever there is a nonidentity morphism from t to u in the free category $F(\mathscr{G})$.

To apply the properties of well-founded relations to directed graphs, we introduce the concept of a strict weight.

Definition 3.16 (Strict weight)**.** Let $\mathscr{G} = [E \rightrightarrows V]$ be a directed graph. A *strict weight* on $\mathscr{G}$, denoted $w\colon \mathscr{G} \to (W, R)$, consists of

- a set W,
- a well-founded relation R on W, and
- a function $w\colon V \to W$,

such that for every edge $e\colon s \Rightarrow t$ of $\mathscr{G}$, we have $w(s)\, R\, w(t)$.

The following lemma is a triviality, but it reframes the definitions in a form that has useful generalizations.

Lemma 3.17. *Suppose $\mathscr{G} = [E \rightrightarrows V]$ is a directed graph. Then $\mathscr{G}$ is well-founded if and only if it has a strict weight.*

Proof. Let $\to$ be the relation on V given by $t \to u$ if and only if there is an edge $t \Rightarrow u$ in $\mathscr{G}$. Then if $\mathscr{G}$ is well-founded, so is $\to$, and the identity function on V yields a strict weight $\mathscr{G} \to (V, \to)$. Conversely, any strict weight $w\colon \mathscr{G} \to (W, R)$ maps paths $t_0 \Rightarrow t_1 \Rightarrow \cdots$ in $\mathscr{G}$ to sequences $w(t_0)\, R\, w(t_1)\, R\, \cdots$ in W. Hence if R is well-founded, so is $\mathscr{G}$. $\square$

In practice, we often take $(W, R) = (\mathbb{N}, >)$, as the natural numbers are well-ordered. Then a strict weight $w\colon \mathscr{G} \to (\mathbb{N}, >)$ measures some feature of the vertices of $\mathscr{G}$ that strictly decreases along the edges of $\mathscr{G}$. But in fact Lemma 3.17 is completely general.

Our basic result for solving CNPs is the following.

Proposition 3.18. *Suppose $\mathscr{G} = [E \rightrightarrows V]$ is a graph and $\sim$ is a congruence relation on $F(\mathscr{G})$. Let $M \subseteq V$ be the set of all $\mathscr{G}$-minimal vertices.*

(1) If $\mathscr{G}$ is well-founded (or, equivalently by Lemma 3.17, has a strict weight), then for every $t \in V$, there is at least one vertex $m \in M$ for which there exists a morphism $t \Rightarrow m$ in $F(\mathscr{G})$.

(2) If $(\mathscr{G}, \sim)$ is confluent, then for every $t \in V$, there is at most one vertex $m \in M$ for which there exists a morphism $t \Rightarrow m$ in $F(\mathscr{G})$. Moreover, given $\nu, \nu' \colon t \Rightarrow m$, we have $\nu \sim \nu'$.

(3) If $\mathscr{G}$ is well-founded (i.e. has a strict weight) and $(\mathscr{G}, \sim)$ is confluent, then M is a solution to the CNP $(\mathscr{G}, \varnothing, \sim)$.

Proof of (1). Let $A = \{u \in V \mid F(\mathscr{G})(t, u) \neq \varnothing\}$, let $\to$ be the relation on V with $t \to u$ if and only if there is an edge $t \Rightarrow u$ in $\mathscr{G}$, and let $\twoheadrightarrow$ be the transitive closure of $\to$. As $t \in A$, by Proposition 3.15 (1), there is a vertex $m \in A$ that is $\twoheadrightarrow$-minimal relative to A. Thus there is a morphism $\alpha \colon t \Rightarrow m$ in $F(\mathscr{G})$: we claim m is $\mathscr{G}$-minimal. Suppose not. Then there is an edge $e \colon m \Rightarrow u$ in $\mathscr{G}$, so we have a morphism $(e \bullet \alpha) \colon t \Rightarrow m \Rightarrow u$ in $F(\mathscr{G})$ and hence $u \in A$. But e exhibits the relation $m \twoheadrightarrow u$, contradicting the $\twoheadrightarrow$-minimality of m relative to A. Therefore $m \in M$. $\square$

Proof of (2). Given morphisms $\nu \colon t \Rightarrow m$ and $\nu' \colon t \Rightarrow m'$ in $F(\mathscr{G})$ with $m, m' \in M$, since $(\mathscr{G}, \sim)$ is confluent, there exist morphisms $\alpha \colon m \Rightarrow c$ and $\alpha' \colon m' \Rightarrow c$ in $F(\mathscr{G})$ for which $(\alpha \bullet \nu) \sim (\alpha' \bullet \nu')$. By the $\mathscr{G}$-minimality of m and m', the paths in $\mathscr{G}$ corresponding to α and α' must be empty, so α and α' are identities. It follows that $m = c = m'$ and $\nu = (\alpha \bullet \nu) \sim (\alpha' \bullet \nu') = \nu'$. $\square$

Proof of (3). As $\mathscr{G}[\varnothing^{-1}] = \mathscr{G}$, by (1) and (2), M is a solution to $(\mathscr{G}, \varnothing, \sim)$. $\square$

The problem now becomes establishing confluence conditions on $(\mathscr{G}, \sim)$. A key insight, originally due to Newman [23] in the combinatorial case and rephrased by Huet [8] in terms of well-founded induction, is that confluence can often be reduced to local confluence in the sense of the next definition. Such results are often called Diamond Lemmas.

Definition 3.19 (Local confluence)**.** Suppose $\mathscr{G} = [E \rightrightarrows V]$ is a directed graph and $\sim$ is a congruence relation on $F(\mathscr{G})$. We say that a vertex $t \in V$ is $(\mathscr{G}, \sim)$-*locally confluent* if, for any edges $e \colon t \Rightarrow a$ and $f \colon t \Rightarrow b$ in $\mathscr{G}$, there exist morphisms $\alpha \colon a \Rightarrow c$ and $\beta \colon b \Rightarrow c$ in $F(\mathscr{G})$ for which $(\alpha \bullet e) \sim (\beta \bullet f)$. We say that $(\mathscr{G}, \sim)$ is *locally confluent* if every $t \in V$ is $(\mathscr{G}, \sim)$-locally confluent.

Note that local confluence is a priori weaker than confluence, as the former is a condition on all *edges* with a common domain in $\mathscr{G}$, whereas the latter is a condition on all *morphisms* with a common domain in $F(\mathscr{G})$. But the Diamond Lemma will show that the weaker condition implies the stronger one given well-foundedness. We now present our categorical analogue of the Diamond Lemma, whose proof we adapt from Huet [8].

Lemma 3.20 (Diamond Lemma)**.** *Let $\mathscr{G} = [E \rightrightarrows V]$ be a well-founded directed graph and $\sim$ be a congruence relation on $F(\mathscr{G})$. Then $(\mathscr{G}, \sim)$ is confluent if and only if it is locally confluent.*

Proof. The "only if" direction is immediate, as every edge in $\mathscr{G}$ is a morphism in $F(\mathscr{G})$. For the "if" direction, suppose $(\mathscr{G}, \sim)$ is locally confluent. We shall prove that every vertex in V is $(\mathscr{G}, \sim)$-confluent by well-founded induction.

Suppose $t \in V$ is $\mathscr{G}$-minimal, so the only path out of t in $\mathscr{G}$ is empty. Thus morphisms $\gamma\colon t \Rightarrow a$ and $\delta\colon t \Rightarrow b$ in $F(\mathscr{G})$ are identities, so $a = t = b$ and

$$(\mathrm{id}_t \bullet \gamma) = \gamma = \mathrm{id}_t \sim \mathrm{id}_t = \delta = (\mathrm{id}_t \bullet \delta).$$

Therefore t is $(\mathscr{G}, \sim)$-confluent.

Now suppose that $t \in V$ is not $\mathscr{G}$-minimal and that u is $\mathscr{G}$-confluent whenever there is an edge $t \Rightarrow u$ in $\mathscr{G}$. Fix morphisms $\gamma\colon t \Rightarrow a$ and $\delta\colon t \Rightarrow b$ in $F(\mathscr{G})$. If one of these is the identity, assume without loss of generality $\gamma = \mathrm{id}_t$. Then

$$(\delta \bullet \gamma) = (\delta \bullet \mathrm{id}_t) = \delta \sim \delta = (\mathrm{id}_b \bullet \delta),$$

so t is $(\mathscr{G}, \sim)$-confluent. Otherwise, γ and δ are paths of length at least 1 in $\mathscr{G}$, so we may factor them as

$$\gamma\colon t \overset{e}{\Rightarrow} a' \overset{\gamma'}{\Longrightarrow} a \quad \text{and} \quad \delta\colon t \overset{f}{\Rightarrow} b' \overset{\delta'}{\Longrightarrow} b$$

for edges $e\colon t \Rightarrow a'$ and $f\colon t \Rightarrow b'$ in $\mathscr{G}$ and morphisms $\gamma'\colon a' \Rightarrow a$ and $\delta'\colon b' \Rightarrow b$ in $F(\mathscr{G})$.

Now consider the diagram

$$\begin{tikzcd}
& & & t \arrow[dl, Rightarrow, "e"'] \arrow[dr, Rightarrow, "f"] & & & \\
& & a' \arrow[ddll, Rightarrow, "\gamma'"'] \arrow[dr, Rightarrow, "\alpha_1"'] & \sim & b' \arrow[dl, Rightarrow, "\beta_1"] \arrow[ddrr, Rightarrow, "\delta'"] & & \\
& \sim & & c_1 \arrow[ddll, Rightarrow, "\beta_2"] & & & \\
a \arrow[dr, Rightarrow, "\alpha_2"'] & & & & \sim & & b \arrow[dddlll, Rightarrow, "\beta"'] \\
& c_2 \arrow[ddrr, Rightarrow, "\alpha_3"'] & & & & & \\
& & & & & & \\
& & & c. & & &
\end{tikzcd}$$

Morphisms $\alpha_1\colon a' \Rightarrow c_1$ and $\beta_1\colon b' \Rightarrow c_1$ in $F(\mathscr{G})$ satisfying $(\alpha_1 \bullet e) \sim (\beta_1 \bullet f)$ exist because $(\mathscr{G}, \sim)$ is locally confluent. Then morphisms $\alpha_2\colon a \Rightarrow c_2$ and $\beta_2\colon c_1 \Rightarrow c_2$ satisfying $(\alpha_2 \bullet \gamma') \sim (\beta_2 \bullet \alpha_1)$ exist because a' is $(\mathscr{G}, \sim)$-confluent by induction. Finally, morphisms $\alpha_3\colon c_2 \Rightarrow c$ and $\beta\colon b \Rightarrow c$ satisfying $(\alpha_3 \bullet \beta_2 \bullet \beta_1) \sim (\beta \bullet \delta')$ exist because b' is $(\mathscr{G}, \sim)$-confluent by induction. As $\sim$ is a congruence relation,

$$(\alpha_3 \bullet \alpha_2 \bullet \gamma) = (\alpha_3 \bullet \alpha_2 \bullet \gamma' \bullet e) \sim (\beta \bullet \delta' \bullet f) = (\beta \bullet \delta),$$

so t is $(\mathscr{G}, \sim)$-confluent. Thus $(\mathscr{G}, \sim)$ is confluent by well-founded induction. $\square$

We therefore have yet another result for solving CNPs, which is the first that we shall apply directly.

Proposition 3.21. *Suppose $\mathscr{G} = [E \rightrightarrows V]$ is a graph and $\sim$ is a congruence relation on $F(\mathscr{G})$. Let $M \subseteq V$ be the set of all $\mathscr{G}$-minimal vertices. If $\mathscr{G}$ is well-founded (i.e. has a strict weight) and $(\mathscr{G}, \sim)$ is locally confluent, then M is a solution to the CNP $(\mathscr{G}, \varnothing, \sim)$.*

Proof. By Lemma 3.20, $(\mathscr{G}, \sim)$ is confluent, so this follows from Proposition 3.18 (3). $\square$

3.3 Coherence for associators

Here we begin to recast Mac Lane's coherence theorem for monoidal categories from [16] into our language of CNPs, focusing first on (lax, i.e. not necessarily invertible) associators. We define a coherence schema $\mathfrak{C}_{\mathrm{A}} := (\mathfrak{S}_{\otimes}, \mathfrak{T}_{\mathrm{A}}, \varnothing, \mathfrak{R}^{\mathrm{A}})$ as follows.

- The signature $\mathfrak{S}_{\otimes}$ consists of:

  - one type, $T \coloneqq \{\mathtt{m}\}$;
  - no constant symbols, $C \coloneqq \varnothing$; and
  - one function symbol, $F \coloneqq \{\otimes\}$, with $\mathrm{ar}(\otimes) = 2$.

  We denote its language by $\mathcal{M} \coloneqq \mathcal{L}(\mathfrak{S}_{\otimes})$. Explicitly, $\mathcal{M}$ consists of:

  (1) the variable $x_i \coloneqq x_i^{\mathtt{m}}$ for $i > 0$; and

  (2) the term $\otimes(s, t)$ for $s, t \in \mathcal{L}^{\mathrm{M}}$.

  We may think of $\mathcal{M}$ as the free magma[4] on $\{x_1 < x_2 < \cdots\}$, where the product of $s, t \in \mathcal{M}$ is given by $\otimes(s, t)$.

- The transformation schema is $\mathfrak{T}_{\mathrm{A}} \colon \{\alpha\} \to \mathcal{M} \times \mathcal{M}$, with the domain and codomain of the sole edge label in $\mathcal{E} = \{\alpha\}$ given by

  $$\alpha \colon \otimes(\otimes(x_1, x_2), x_3) \Rightarrow \otimes(x_1, \otimes(x_2, x_3)),$$

  yielding a directed graph $\mathscr{B}^{\mathrm{A}}$ whose vertices are terms in $\mathcal{M}$ and whose edges are basic transformations in $\mathrm{Bs}^{\mathrm{A}} \coloneqq \mathrm{Bs}(\mathfrak{S}_{\otimes}, \mathfrak{T}_{\mathrm{A}})$. Explicitly, Bs^{A} is given as follows.

  (1) For $s, t, u \in \mathcal{M}$, there is an atomic transformation in Bs^{A} given by

  $$\alpha_{s,t,u} \colon \otimes(\otimes(s, t), u) \Rightarrow \otimes(s, \otimes(t, u)).$$

  (2) For $s, t, u \in \mathcal{M}$ and $\varphi \colon s \Rightarrow t$ in Bs^{A}, there are basic transformations in Bs^{A} given by

  $$\otimes(\varphi, \mathrm{id}_u) \colon \otimes(s, u) \Rightarrow \otimes(t, u) \quad \text{and} \quad \otimes(\mathrm{id}_u, \varphi) \colon \otimes(u, s) \Rightarrow \otimes(u, t).$$

- There are no marked edges.

---

[4] A magma is a set equipped with a binary operation, with no other conditions.

- The commutativity schema $\mathfrak{R}^{\mathrm{A}}$ is the local binary relation on the free category $\mathscr{F}^{\mathrm{A}} := F(\mathscr{B}^{\mathrm{A}})$ with only one related pair, consisting of the two paths around the following diagram (with subscripts in morphisms suppressed for readability):

$$
\begin{tikzcd}
\otimes(\otimes(\otimes(x_1,x_2),x_3),x_4) \arrow[r, Rightarrow, "\otimes(\alpha,\mathrm{id})"] \arrow[dd, Rightarrow, "\alpha"'] & \otimes(\otimes(x_1,\otimes(x_2,x_3)),x_4) \arrow[d, Rightarrow, "\alpha"] \\
\mathfrak{R}^{\mathrm{A}} & \otimes(x_1,\otimes(\otimes(x_2,x_3),x_4)) \arrow[d, Rightarrow, "\otimes(\mathrm{id},\alpha)"] \\
\otimes(\otimes(x_1,x_2),\otimes(x_3,x_4)) \arrow[r, Rightarrow, "\alpha"] & \otimes(x_1,\otimes(x_2,\otimes(x_3,x_4))).
\end{tikzcd}
\tag{3.22}
$$

So a model of this coherence schema $\mathfrak{C}_{\mathrm{A}} = (\mathfrak{S}_{\otimes}, \mathfrak{T}_{\mathrm{A}}, \varnothing, \mathfrak{R}^{\mathrm{A}})$ in **<u>Cat</u>** is a category with a binary operation and a lax associator for that binary relation that obeys the pentagon axiom from (3.22).

Now set $V_k := \{x_1 < \cdots < x_k\}$ for $k \in \mathbb{N}$. We will restrict our attention to the set of all terms $\mathcal{M}_k \subset \mathcal{M}$ with scope V_k. Then $\mathcal{M}_k$ induces the full subgraph $\mathscr{B}_k^{\mathrm{A}} \subset \mathscr{B}^{\mathrm{A}}$, which in turn induces the free category $\mathscr{F}_k^{\mathrm{A}} := F(\mathscr{B}_k^{\mathrm{A}})$.

We will work with the restricted coherence schema $(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k)$, yielding a local binary relation $R_k^{\mathrm{A}} := R(\mathfrak{C}_{\mathrm{A}})_{V_k}$ on $\mathscr{F}_k^{\mathrm{A}}$ as in Definition 2.46. Then $\mathscr{P}(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k) = \mathscr{F}_k^{\mathrm{A}}/\langle R_k^{\mathrm{A}}\rangle$ is the category presented by $(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k)$. We will also consider the monomorphic category $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k) = \mathscr{F}_k^{\mathrm{A}}/\langle R_k^{\mathrm{A}}\rangle^{\mathrm{L}}$ presented by $(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k)$.

We can now use Corollary 3.44 to solve the corresponding CNPs, denoted $\mathrm{CN}(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k) = (\mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{A}}\rangle)$ and $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k) = (\mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{A}}\rangle^{\mathrm{L}})$, by showing that $\mathscr{B}_k^{\mathrm{A}}$ has a strict weight and that the pair $(\mathscr{B}_k^{\mathrm{A}}, \langle R_k^{\mathrm{A}}\rangle)$ is locally confluent. We first define a function $\ell\colon \mathcal{M} \to \mathbb{N}$ so that $\ell(s)$ is the (necessarily positive) number of variables, counting repeats, appearing in $s \in \mathcal{M}$. Then for our proposed strict weight, we define a function $\|\cdot\|\colon \mathcal{M} \to \mathbb{N}$ recursively as follows (the Parsing Lemma ensures well-definedness in the recursive step).

(i) For each variable $x_i \in V \subseteq \mathcal{M}$, set $\|x_i\| := 0$.

(ii) For terms $s, t \in \mathcal{M}$, set $\|\otimes(s,t)\| := \|s\| + \|t\| + \ell(s) - 1$.

The idea is that the added $\ell(s)$ term in the formula for $\|\otimes(s,t)\|$ penalizes longer left entries, so moving parentheses rightward reduces weight. We formalize this as follows.

Lemma 3.23. *The function* $\|\cdot\|\colon \mathcal{M} \to \mathbb{N}$ *defined above, when restricted to* $\mathcal{M}_k \subset \mathcal{M}$, *induces a strict weight* $\mathscr{B}_k^{\mathrm{A}} \to (\mathbb{N}, >)$.

Proof. Given a basic transformation $\varphi\colon t \Rightarrow t'$ between $t, t' \in \mathcal{M}_k$, we want to show that $\|t\| > \|t'\|$. We proceed by induction on φ.

If φ is atomic, then $\varphi = \alpha_{u,v,w}\colon \otimes(\otimes(u,v),w) \Rightarrow \otimes(u,\otimes(v,w))$ for $u, v, w \in \mathcal{M}_k$, and

$$\begin{aligned} \|\otimes(\otimes(u,v),w)\| &= \|u\| + \|v\| + \|w\| + 2\ell(u) + \ell(v) - 2, \quad \text{whereas} \\ \|\otimes(u,\otimes(v,w))\| &= \|u\| + \|v\| + \|w\| + \ell(u) + \ell(v) - 2. \end{aligned}$$

Therefore $\|\otimes(\otimes(u,v),w)\| > \|\otimes(u,\otimes(v,w))\|$.

Now assume $\varphi\colon t \Rightarrow t'$ is a basic transformation satisfying $\|t\| > \|t'\|$. Then for all $s \in \mathcal{M}_k$, the domain and codomain of $\otimes(\mathrm{id}_s, \varphi)$ satisfy

$$\|\otimes(s,t)\| = \|s\| + \|t\| + \ell(s) - 1 > \|s\| + \left\|t'\right\| + \ell(s) - 1 = \left\|\otimes(s,t')\right\|.$$

Meanwhile, the domain and codomain of each atomic transformation $\alpha_{u,v,w}$ contain the same number of variables, so by induction $\ell(t) = \ell(t')$. Hence the domain and codomain of $\otimes(\varphi, \mathrm{id}_s)$ satisfy

$$\|\otimes(t,s)\| = \|t\| + \|s\| + \ell(t) - 1 > \left\|t'\right\| + \|s\| + \ell(t') - 1 = \left\|\otimes(t',s)\right\|.$$

So $\|\cdot\|\colon \mathscr{B}_k^{\mathrm{A}} \to (\mathbb{N}, >)$ is a strict weight by induction. $\square$

Next, we prove local confluence, leveraging the subrelations $R^{\mathrm{fun}}, R^{\mathrm{nat}}$, and R^{com} of R_k^{A} as defined in Definition 2.46.

Lemma 3.24. *The pair* $(\mathscr{B}_k^{\mathrm{A}}, \langle R_k^{\mathrm{A}} \rangle)$ *is locally confluent.*

Proof. We show that every $p \in \mathcal{M}_k$ is $(\mathscr{B}_k^{\mathrm{A}}, \langle R_k^{\mathrm{A}} \rangle)$-locally confluent by induction on the length of p. There are no basic transformations with atomic domains, so assume $p = \otimes(q,r)$ and let $\varphi_n\colon p \Rightarrow p_n$ be a basic transformation for $n = 1, 2$. We view their spans $S(\varphi_1)$ and $S(\varphi_2)$ as intervals and divide into cases depending on where they appear. If the spans are either both in the left entry q or both in the right entry r, the recursive definition of R_k^{A} implies that we are done by induction. On the other hand, if one span is in the left entry and the other is in the right, local confluence follows from the construction of R^{fun}, exhibiting bifunctoriality of the interpretation of $\otimes$.

By Corollaries 2.16 and 2.17, the remaining case to consider is if the domain of the atom of one of the basic transformations—say, without loss of generality, φ_1—is the entire domain p. This would make φ_1 atomic with domain $p = \otimes(q, r) = \otimes(\otimes(s, t), r)$; that is, we have $\varphi_1 = \alpha_{s,t,r} \colon \otimes(\otimes(s, t), r) \Rightarrow \otimes(s, \otimes(t, r))$. Meanwhile, the domain of the atom of φ_2 is either also p; in one of s, t, or r; or exactly $\otimes(s, t)$.

In the first case, if the spans of φ_1 and φ_2 coincide, then for this transformation schema, so must their atoms $A(\varphi_1) = A(\varphi_2)$. Then by Lemma 2.37, $\varphi_1 = \varphi_2$ and we are done.

If the span of φ_2 is in s, t, or r, local confluence follows from a R^{nat}-relation exhibiting the naturality of the interpretation of α. For instance, if the span of φ_2 is in t, we can write $\varphi_2 = \otimes(\otimes(\mathrm{id}_s, \varphi'), \mathrm{id}_r)$ for some basic transformation $\varphi' \colon t \Rightarrow t'$. Then as in Definition 2.46 (c), there exists a basic transformation $\psi \colon \otimes(s, \otimes(t, r)) \Rightarrow \otimes(s, \otimes(t', r))$ satisfying

$$(\psi \bullet \alpha_{s,t,r})\ R^{\text{nat}}\ (\alpha_{s,t',r} \bullet \varphi_2),$$

proving local confluence; in fact, we would take $\psi = \otimes(\mathrm{id}_s, \otimes(\varphi', \mathrm{id}_r))$.

Finally, if the domain of the atom of φ_2 is $\otimes(s, t)$, we can write $s = \otimes(u, v)$ in

$$p = \otimes(\otimes(s, t), r) = \otimes(\otimes(\otimes(u, v), t), r)$$

so that $\varphi_1 = \alpha_{\otimes(u,v),t,r}$ and $\varphi_2 = \otimes(\alpha_{u,v,t}, \mathrm{id}_r)$. But then local confluence follows from substituting (u, v, t, r) for (x_1, x_2, x_3, x_4) in (3.22) to obtain the desired R^{com}-relation. $\square$

Remark 3.25. The final step of the proof demonstrates why we need (3.22) in our commutativity schema: it covers exactly the case we need to prove local confluence not already implied by functoriality and naturality.

Now that we have shown $(\mathscr{B}_k^{\mathrm{A}}, \langle R_k^{\mathrm{A}} \rangle)$ to be locally confluent, we can apply Proposition 3.21 to deduce that the set of minimal vertices in $\mathscr{B}_k^{\mathrm{A}}$ is a solution to the coherent normalization problem $\mathrm{CN}(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k)$. These minimal elements are exactly the terms in $\mathcal{M}_k$ that have no subterms of the form $\otimes(\otimes(u, v), w)$ for any $u, v, w \in \mathcal{M}_k$. By Corollary 3.12, the same set of minimal vertices is also a solution to $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k)$. We summarize this as follows.

Definition 3.26 (Right-normalized term)**.** We say that a term $t \in \mathcal{M}_k$ is *right-normalized* if t contains no subterms of the form $\otimes(\otimes(u,v),w)$ for any $u,v,w \in \mathcal{M}_k$. We write $\mathcal{N}_k \subset \mathcal{M}_k$ for the set of all right-normalized terms in $\mathcal{M}_k$.

We have the following alternative characterizations of right-normalized terms.

Proposition 3.27. *For $t \in \mathcal{M}_k$, the following are equivalent:*

(1) $\|t\| = 0$*;*

(2) t *is right-normalized: it contains no subterms of the form* $\otimes(\otimes(u,v),w)$ *for any* $u,v,w \in \mathcal{M}_k$*;*

(3) either t *is a variable or* t *may be written as* $t = \otimes(x_i, s)$ *for a variable* x_i *and another right-normalized term* s*;*

(4) $t = \otimes(x_{i_1}, \otimes(x_{i_2}, \cdots, \otimes(x_{i_{n-1}}, x_{i_n})\cdots))$ *for the* n *variables* $x_{i_1}, \ldots, x_{i_n} \in V_k$ *appearing in that order in* t*.*

Proof. First, we show that (1) $\implies$ (2). As $|\cdot|\colon \mathscr{B}_k^{\mathrm{A}} \to (\mathbb{N}, >)$ is a strict weight and 0 is minimal in $\mathbb{N}$, if $\|t\| = 0$, then t must also be a minimal vertex of $\mathscr{B}_k^{\mathrm{A}}$, making t right-normalized.

Next, we show that (2) $\implies$ (3). Say t is right-normalized. If t is a variable, we are done; otherwise, $t = \otimes(u,s)$ for some $u, s \in \mathcal{M}_k$. Then by definition, both u and s must also be right-normalized. Moreover, if u is not a variable, then we could write $u = \otimes(v,w)$ for some $v, w \in \mathcal{M}_k$, making $t = \otimes(\otimes(v,w),s)$ *not* right-normalized. So u must be a variable, satisfying (3).

We may now show that (3) $\implies$ (4) by induction on the number of variables n (possibly with repeats) appearing in t. The base case $n = 1$ is automatic; otherwise, $t = \otimes(x_i, s)$, and by induction we may write $s = \otimes(x_{i_2}, \ldots, \otimes(x_{i_{n-1}}, x_{i_n})\cdots)$, so (4) follows.

Finally, we also show that (4) $\implies$ (1) by induction on the number of variables n appearing in t. The base case $n = 1$ is automatic; otherwise, $t = \otimes(x_{i_1}, s)$, and by induction we may write

$$\|t\| = |x_{i_1}| + |s| + \ell(x_{i-1}) - 1 = 0 + 0 + 1 - 1 = 0,$$

so (1) follows. □

Corollary 3.28. *The subset $\mathcal{N}_k \subset \mathcal{M}_k$ is a solution to the coherent normalization problems* $\mathrm{CN}(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k)$ *and* $\mathrm{CN}_\mathrm{L}(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k)$.

Then applying Theorem 3.7 yields the following coherence theorem for our restricted coherence schema $(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k)$. Recall that $\mathfrak{C}_\mathrm{A}$ has only 1 type, $\mathtt{m}$, while the minimal common scope for $\mathcal{M}_k$ is the set of k variables $V_k = \{x_1 < \cdots < x_k\}$.

Theorem 3.29. *For each $t \in \mathcal{M}_k$, there is a unique $n(t) \in \mathcal{N}_k$ such that there is a morphism $t \Rightarrow n(t)$ in the category $\mathscr{P}(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k)$, and there is a unique morphism $t \Rightarrow n(t)$ in $\mathscr{P}(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k)$. So the classifying space $B[\mathscr{P}(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k)]$ is a coproduct over $\mathcal{N}_k$ of contractible spaces. Moreover, $\mathscr{P}^\mathrm{L}(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k)$ is a preorder category: every diagram in it commutes.*

More generally, let $|\cdot| \colon \mathfrak{C}_\mathrm{A} \to \underline{\mathbf{K}}$ be a model for $\mathfrak{C}$ in a cartesian monoidal 2-category $\underline{\mathbf{K}}$, giving rise to the functor

$$|\cdot| \coloneqq |\cdot|_{V_k} \colon \mathscr{P}(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k) \to \underline{\mathbf{K}}(|\mathtt{m}|^k, |\mathtt{m}|)$$

from Corollary 2.87. Then for any diagram in $\underline{\mathbf{K}}(|\mathtt{m}|^k, |\mathtt{m}|)$ that factors through $|\cdot|$, the morphisms parametrized by any pair of parallel morphisms in $\mathscr{P}$ from some $t \in \mathcal{M}_k$ to $n(t) \in \mathcal{N}_k$ are equal. If further $|\cdot|$ sends the residue of every edge of $\mathscr{B}_k^\mathrm{A}$ to a monomorphism, then $|\cdot|$ factors through $\mathscr{P}^\mathrm{L}(\mathfrak{C}_\mathrm{A}, \mathcal{M}_k)$, so any diagram in $\underline{\mathbf{K}}(|\mathtt{m}|^k, |\mathtt{m}|)$ that factors through $\mathscr{P}$ commutes.

3.4 CNPs presenting translation categories

The coherent Diamond Lemma (Lemma 3.20) proven and applied above is a useful method for solving CNPs presented by suitably "directed" reductions (e.g. for monoidal categories). We now present another method for solving CNPs arising from monoid and group actions, which may be used to prove a coherence theorem for permutative categories (i.e. strictly associative and unital symmetric monoidal categories).

We will make extensive use of the definition of *translation categories* for monoid actions and *translation groupoids* for group actions, as given in Definition B.1, where we write $\mathbb{T}_M X$

for the translation category of the action of a monoid M on a set X. We are particularly interested in translation groupoids of group actions that are free, as such translation groupoids can be written as the coproduct of categories with terminal objects, just as categories presented by CNPs can per Theorem 3.3. In fact, *every* object of a translation groupoid of a group action that is free is a terminal object of its connected component, as we prove in the following proposition. Here, given a group Γ, we denote the orbit of an element x of a (left) Γ-set by Γx, itself viewed as a (left) Γ-set.

Proposition 3.30. *Let Γ be a group and X be a free left Γ-set, and let $O(X)$ be a set of orbit representatives of X. Then $\mathbb{T}_\Gamma X$ decomposes into its connected components as*

$$\mathbb{T}_\Gamma X \cong \coprod_{x \in O(X)} \mathbb{T}_\Gamma(\Gamma x),$$

and each $x \in X$ is terminal in $\mathbb{T}_\Gamma(\Gamma x)$.

Proof. The decomposition of $\mathbb{T}_\Gamma X$ into its connected components, each of the form $\mathbb{T}_\Gamma(\Gamma x)$, follows directly from the standard orbit decomposition of Γ-sets, so it remains to show that x is terminal in $\mathbb{T}_\Gamma(\Gamma x)$. Given $y \in \mathrm{Ob}(\mathbb{T}_\Gamma(\Gamma x)) = \Gamma x$, there exists $g \in \Gamma$ for which $y = gx$, so there exists a morphism $g^{-1} \colon y \to x$ in $\mathbb{T}_\Gamma(\Gamma x)$. Then given another morphism $h \colon y \to x$ in $\mathbb{T}_\Gamma(\Gamma x)$, we have $x = hy$ and thus $y = ghy$, so by freeness $gh = e$ and thus $h = g^{-1}$, proving uniqueness. $\square$

If a CNP presents a coproduct of categories with terminal objects as above, then choosing a terminal object of each component yields a solution to the original CNP. So as a special case, if a CNP presents the translation groupoid $\mathbb{T}_\Gamma(\Gamma/e)$, where Γ/e is the free orbit of Γ (i.e. the group Γ itself viewed as a left Γ-set), then any object (group element) $\gamma \in \Gamma/e$ forms a singleton solution $\{\gamma\}$ to the CNP. We therefore describe a method of presenting translation categories and groupoids for left actions from monoid and group presentations.

Here we use the standard monoid constructions from Definition B.2, where we fix our notation: the monoid M presented (as a monoid) by generators $G \subseteq M$ and relations R is the quotient of the free monoid $F_{\mathrm{mon}}G$ on the set G by the congruence $\langle R \rangle$. Given a presentation of a monoid and a set that it acts on, we seek a CNP that presents the corresponding translation category. We do so by constructing a multi-object categorical analogue of the monoid presentation: its objects are elements of the set acted on, and its morphisms are monoid elements, generated by generators, sending their domains to their codomains. We take care to ensure the congruence of the CNP imposes precisely the relations given by R.

Proposition 3.31. *Suppose M is a monoid and X is a left M-set. Suppose further that M is presented (as a monoid) by generators G and relations R. We define a CNP $\mathfrak{G} = (\mathscr{G}, \varnothing, \sim)$ as follows.*

(1) The directed graph $\mathscr{G}$ has vertex-set X and edge-set

$$E = \{(x, g, y) \in X \times G \times X \mid gx = y\},$$

with $\operatorname{dom}(x, g, y) = x$ and $\operatorname{cod}(x, g, y) = y$.

(2) The relation $\sim$ is defined as follows: let $\mathbb{T}_{F_{\mathrm{mon}}G}X$ be the translation category of X as an $F_{\mathrm{mon}}G$-set, let $\mathscr{G} \to \mathbb{T}_{F_{\mathrm{mon}}G}X$ be the graph morphism that is the identity on vertices and sends edge (x, g, gx) to the morphism $g\colon x \to gx$, and let $\varepsilon\colon F(\mathscr{G}) \to \mathbb{T}_{F_{\mathrm{mon}}G}X$ be its free functorial extension. We define a local binary relation $\widehat{R}$ on $F(\mathscr{G})$ as follows: given parallel morphisms f, f' in $F(\mathscr{G})$, write

$$f \mathrel{\widehat{R}} f' \quad \text{if and only if} \quad \varepsilon(f) \mathrel{R} \varepsilon(f')$$

(recall that hom-sets in $\mathbb{T}_{F_{\mathrm{mon}}G}X$ are subsets of $F_{\mathrm{mon}}G$, so it makes sense to ask if $\varepsilon(f)$ and $\varepsilon(f')$ are R-related elements of $F_{\mathrm{mon}}G$). We let $\sim$ be the congruence relation on $F(\mathscr{G})$ generated by $\widehat{R}$.

Then the category $F(\mathscr{G})/{\sim}$ presented by $\mathfrak{G}$ is isomorphic to the translation category $\mathbb{T}_M X$ of X as an M-set.

Proof. By the functoriality of Definition B.1, the canonical map $\pi\colon F_{\mathrm{mon}}G \to M$ induces a functor $\mathbb{T}_\pi X\colon \mathbb{T}_{F_{\mathrm{mon}}G}X \to \mathbb{T}_M X$ that is the identity on objects and π on morphisms. Then composing with ε yields a functor $(\mathbb{T}_\pi X \circ \varepsilon)\colon F(\mathscr{G}) \to \mathbb{T}_M X$. As π agrees on R-related products of generators in G, given parallel morphisms f, f' in $F(\mathscr{G})$ for which $f \mathrel{\widehat{R}} f'$

and thus $\varepsilon(f)\; R\; \varepsilon(f')$ in $F_{\mathrm{mon}}G$, we have $\pi(\varepsilon(f)) = \pi(\varepsilon(f'))$. Hence $\mathbb{T}_\pi X \circ \varepsilon$ agrees on $\widehat{R}$-related morphisms, so by Proposition A.9, $\mathbb{T}_\pi X \circ \varepsilon$ descends uniquely to the quotient category $F(\mathscr{G})/{\sim}$, yielding a functor $\iota\colon F(\mathscr{G})/{\sim} \to \mathbb{T}_M X$. It remains to show that ι is an isomorphism.

As ε and $\mathbb{T}_\pi X$ are both identities on objects and ι agrees with $\mathbb{T}_\pi X \circ \varepsilon$ on objects, ι is the identity on objects as well. Then on hom-sets, ι is the function

$$\iota\colon F(\mathscr{G})(x,y)/{\sim} \to \{\mu \in M \mid \mu x = y\}$$

for $x, y \in X$ that, given $f\colon x \Rightarrow y$ in $F(\mathscr{G})$, sends the residue $[f]$ to $\pi(\varepsilon(f))$. We seek to show that this function is bijective.

First note that for $m = g_k \cdots g_1 \in F_{\mathrm{mon}}G$ with $g_1, \ldots, g_k \in G$ and $x \in X$, there is a unique morphism $\alpha_x(m)\colon x \Rightarrow \pi(m)x$ in $F(\mathscr{G})$ that ε sends to $m\colon x \to \pi(m)x$ in $\mathbb{T}_{F_{\mathrm{mon}}G}X$. Freeness of $F_{\mathrm{mon}}G$ and the fact that ε sends edges to generators implies that $\alpha_x(m)$ must be k edges long, of the form

$$\alpha_x(m) = (x_{k-1}, g_k, x_k) \bullet \cdots \bullet (x_0, g_1, x_1)$$

for some $x = x_0, x_1, \ldots, x_{k-1}, x_k = \pi(m)x \in X$. Then each $x_i = g_i x_{i-1} = \pi(g_i \cdots g_1)x$ for $1 \le i \le x$, including $x_k = \pi(g_k \cdots g_1)x = \pi(m)x$, completing our unique construction of $\alpha_x(m)$.

Now for surjectivity, take $\mu \in M$ such that $\mu x = y$. We can write $\mu = \pi(g_k \cdots g_1)$ for generators $g_1, \ldots, g_k \in G$. Then $\alpha \coloneqq \alpha_x(g_k \cdots g_1)$ is a morphism $x \Rightarrow y$ in $F(\mathscr{G})$ that ε sends to the morphism $g_k \cdots g_1\colon x \to y$ in $\mathbb{T}_{F_{\mathrm{mon}}G}X$. Hence $\iota([\alpha]) = \pi(\varepsilon(\alpha)) = \pi(g_k \cdots g_1) = \mu$, proving fullness.

For injectivity, take $f, f'\colon x \Rightarrow y$ in $F(\mathscr{G})$ such that $\pi(\varepsilon(f)) = \pi(\varepsilon(f'))$, so $\varepsilon(f)\,\langle R\rangle\,\varepsilon(f')$. Explicitly, there are $\varepsilon(f) = m_0, \ldots, m_n = \varepsilon(f') \in F_{\mathrm{mon}}G$ for which $m_{j-1} = s_j t_j u_j$ and $m_j = s_j t'_j u_j$ for some $s_j, t_j, t'_j, u_j \in F_{\mathrm{mon}}G$ with $t_j\; R\; t'_j$ or $t'_j\; R\; t_j$ for $1 \le j \le n$, such that $\pi(m_0)x = y = \pi(m_n)x$, while $f = \alpha_x(m_0)$ and $f' = \alpha_x(m_n)$. It suffices to show that $\alpha_x(m_0) \sim \alpha_x(m_n)$, and we will do so by induction on n.

The $n = 0$ case follows directly from reflexivity, so let $n \ge 1$ and assume inductively that $\alpha_x(m_0) \sim \alpha_x(m_{n-1})$. Either $t_n\; R\; t'_n$ or $t'_n\; R\; t_n$, so $\pi(t_n) = \pi(t'_n)$ and thus

$\pi(t_n u_n) = \pi(t'_n u_n)$. Then we have parallel morphisms $\alpha_{\pi(u_n)x}(t_n)\colon \pi(u_n)x \Rightarrow \pi(t_n u_n)x$ and $\alpha_{\pi(u_n)x}(t'_n)\colon \pi(u_n)x \Rightarrow \pi(t'_n u_n)x$ for which $\varepsilon(\alpha_{\pi(u_n)x}(t_n)) = t_n$ and $\varepsilon(\alpha_{\pi(u_n)x}(t'_n)) = t'_n$, so $\alpha_{\pi(u_n)x}(t_n) \sim \alpha_{\pi(u_n)x}(t'_n)$ and thus

$$\alpha_{\pi(u_n t_n)x}(s_n) \bullet \alpha_{\pi(u_n)x}(t_n) \bullet \alpha_x(u_n) \sim \alpha_{\pi(u_n t_n)x}(s_n) \bullet \alpha_{\pi(u_n)x}(t'_n) \bullet \alpha_x(u_n).$$

Since ε sends the left hand side to $s_n t_n u_n = m_{n-1}$ and the right hand side to $s_n t'_n u_n = m_n$, we can rewrite both sides to get $\alpha_x(m_{n-1}) \sim \alpha_x(m_n)$, so the result follows inductively from transitivity. $\square$

This proposition applies to group presentations and translation groupoids as well, as long as the group is presented as a monoid, with relations that ensure every generator has an inverse but that do not contain any formal inverses of generators. By Lemma B.3, such group presentations coincide with monoid presentations. We may leverage this to construct CNPs for groups and group actions. Note that when the group action is the group's free orbit, the CNP's graph is precisely the group's Cayley graph.

We will use Proposition 3.31 to translate the presentation of the r^{th} symmetric group Σ_r from Proposition B.15 by block transpositions $B_r = \{\tau[i,j,k] \in \Sigma_r \mid 0 \le i < j < k \le r\}$, as defined in Definition B.8, to a CNP. For reasons that will become apparent, we will also include an additional generator id and an additional relation that relates id to the identity permutation, which does not change the group generated from Σ_r. (See Section B.3 for a comprehensive review of the presentation of Σ_r by block transpositions.)

Proposition 3.32. *Let Σ_r denote the r^{th} symmetric group. There is a CNP $\mathfrak{G}_{\Sigma_r} = (\mathscr{S}_r, \varnothing, \langle R_{\Sigma_r}\rangle)$ that presents the translation groupoid $\mathbb{T}_{\Sigma_r}(\Sigma_r/\mathrm{id})$, defined as follows.*

- *We define the graph $\mathscr{S}_r$ as a Cayley graph for the presentation of Σ_r by block transformations as follows. The vertex-set of $\mathscr{S}_r$ is the left Σ_r-set*

  $$\mathcal{S}_r = \{x_{\sigma^{-1}(1)} x_{\sigma^{-1}(2)} \cdots x_{\sigma^{-1}(r)} \mid \sigma \in \Sigma_r\}.$$

  We denote $\sigma^* \coloneqq x_{\sigma^{-1}(1)} x_{\sigma^{-1}(2)} \cdots x_{\sigma^{-1}(r)}$, where $\rho \cdot \sigma^* = (\rho\sigma)^*$, which we denote simply by $\rho\sigma^*$, for $\rho \in \Sigma_r$.

The edge-set of $\mathscr{S}_r$ consists of edges

$$\tau[i,j,k]\colon \sigma^* \Rightarrow \tau[i,j,k]\sigma^*$$

for $\sigma \in \Sigma_r$ and $0 \le i < j < k \le r$ as well as edges

$$\mathrm{id}\colon \sigma^* \Rightarrow \sigma^*$$

for $\sigma \in \Sigma_r$.

- *There are no marked edges.*

- *Adapt the relation R' from Proposition B.15 to define the local binary relation R_{Σ_r} to relate either path around the following diagrams for all $\sigma \in \Sigma_r$:*

  (0) we have

$$\begin{tikzcd} \sigma^* \arrow[r, Rightarrow, bend left, "\mathrm{id}"] \arrow[r, equal, bend right] & \sigma^*, \end{tikzcd} \tag{3.33}$$

  where the top arrow is the edge $\mathrm{id}\colon \sigma^* \Rightarrow \sigma^*$ of $\mathscr{S}_r$, while the bottom arrow is the identity morphism in $F(\mathscr{S}_r)$ corresponding to the empty path of $\mathscr{S}_r$ at σ^*;

  (1) for $0 \le i < j < k \le r$, we have

$$\begin{tikzcd} & \tau[i,j,k]\sigma^* \arrow[dr, Rightarrow, "{\tau[i,i-j+k,k]}"] & \\ \sigma^* \arrow[ur, Rightarrow, "{\tau[i,j,k]}"] \arrow[rr, equal] & & \sigma^*, \end{tikzcd} \tag{3.34}$$

  where the horizontal arrow is again the identity morphism in $F(\mathscr{S}_r)$ corresponding to the empty path of $\mathscr{S}_r$ at σ^*;

  (2) for $0 \le i < j < k \le i' < j' < k' \le r$, we have

$$\begin{tikzcd} \sigma^* \arrow[r, Rightarrow, "{\tau[i,j,k]}"] \arrow[d, Rightarrow, "{\tau[i',j',k']}"'] & \tau[i,j,k]\sigma^* \arrow[d, Rightarrow, "{\tau[i',j',k']}"] \\ \tau[i',j',k']\sigma^* \arrow[r, Rightarrow, "{\tau[i,j,k]}"] & \tau[i',j',k']\tau[i,j,k]\sigma^*; \end{tikzcd} \tag{3.35}$$

(3) for $0 \le i' < j' \le i < j < k \le k' \le r$

$$\begin{tikzcd} \sigma^* \arrow[r, Rightarrow, "{\tau[i,j,k]}"] \arrow[d, Rightarrow, "{\tau[i',j',k']}"'] & \tau[i,j,k]\sigma^* \arrow[d, Rightarrow, "{\tau[i',j',k']}"] \\ \tau[i',j',k']\sigma^* \arrow[r, Rightarrow, "{\tau([i,j,k]-(j'-i'))}"] & \tau[i',j',k']\tau[i,j,k]\sigma^*, \end{tikzcd} \tag{3.36}$$

where $\tau([i,j,k]-(j'-i')) \coloneqq \tau[i-(j'-i'), j-(j'-i'), k-(j'-i')]$*, and*

(4) for $0 \le i < j < k < \ell \le r$*;*

$$\begin{tikzcd} \sigma^* \arrow[rr, Rightarrow, "{\tau[i,j,\ell]}"] \arrow[dr, Rightarrow, "{\tau[i,j,k]}"'] & & \tau[i,j,\ell]\sigma^* \\ & \tau[i,j,k]\sigma^* \arrow[ur, Rightarrow, "{\tau[i-j+k,k,\ell]}"'] & \end{tikzcd} \tag{3.37}$$

Then $\langle R_{\Sigma_r} \rangle$ *is the congruence relation on* $F(\mathscr{S}_r)$ *generated by* R_{Σ_r}*.*

Any $\sigma \in \Sigma_r$ *forms a singleton solution* $\{\sigma\}$ *to this CNP.*

Proof. We apply Proposition 3.31 to the presentation of Σ_r from Proposition B.15 (with an additional generator id and an additional relation relating id to the identity element, so that the presentation still presents Σ_r), yielding a CNP presenting $\mathbb{T}_{\Sigma_r}(\Sigma_r/\mathrm{id})$, the translation groupoid for the free orbit of Σ_r. So any singleton subset of Σ_r is a solution to $\mathfrak{G}_{\Sigma_r}$. □

We will use this CNP to prove the coherence theorem for symmetric monoidal categories in Section 3.8.

3.5 New CNP solutions from old

We now consider when known solutions to CNPs may be used to produce a solution to another, presumably more difficult, CNP. We first consider the simplest way to combine many CNPs into one.

Definition 3.38 (Coproduct CNP)**.** Let $\mathscr{G}_i = [E_i \rightrightarrows V_i]$ be a graph and $\mathfrak{G}_i = (\mathscr{G}_i, I_i, \sim_i)$ be a CNP for $i = 1, 2$. Then define the *coproduct CNP* $\mathfrak{G}_1 \sqcup \mathfrak{G}_2$ to be the CNP $(\mathscr{G}, I, \sim)$, with $\mathscr{G} = \mathscr{G}_1 \sqcup \mathscr{G}_2 = [E_1 \sqcup E_2 \rightrightarrows V_1 \sqcup V_2]$ given by the coproduct of graphs, $I = I_1 \sqcup I_2$, and $f \sim g$ for parallel morphisms f, g in $F(\mathscr{G}[I^{-1}]) \cong F(\mathscr{G}_1[I_1^{-1}]) \sqcup F(\mathscr{G}_2[I_2^{-1}])$ if and only if either $f \sim_1 g$ or $f \sim_2 g$.

More generally, for a finite indexing set J, let $\mathscr{G}_j = [E_j \rightrightarrows V_j]$ be a graph and $\mathfrak{G}_j = (\mathscr{G}_j, I_j, \sim_j)$ be a CNP for $j \in J$. Then define the *coproduct CNP* $\coprod_{j\in J} \mathfrak{G}_j$ to be the CNP $(\mathscr{G}, I, \sim)$, with

$$\mathscr{G} = \coprod_{j\in J} \mathscr{G}_j = \left[\coprod_{j\in J} E_j \rightrightarrows \coprod_{j\in J} V_j\right]$$

given by the coproduct of graphs, $I = \coprod_{j\in J} I_j$, and $f \sim g$ for parallel morphisms f, g in

$$F(\mathscr{G}[I^{-1}]) \cong \coprod_{j\in J} F(\mathscr{G}_j[I_j^{-1}])$$

if and only if $f \sim_j g$ for some $j \in J$.

Remark 3.39. The coproduct CNP defined above is indeed the categorical coproduct in the category **CN**.

Example 3.40. Every CNP $\mathfrak{G} = (\mathscr{G}, I, \sim)$ whose corresponding graph $\mathscr{G} = \coprod_{j\in J} \mathscr{G}_j$ decomposes into a coproduct of graphs $\mathscr{G}_j = [E_j \rightrightarrows V_j]$ for $j \in J$ with J finite may be decomposed into CNPs $\mathfrak{G} = \coprod_{j\in J} \mathfrak{G}_j$, as follows. Define $I_j = I \cap E_j$ so that

$$\coprod_{j\in J} I_j = \coprod_{j\in J} (I \cap E_j) = I \cap \coprod_{j\in J} E_j = I \cap E = I,$$

and define $\sim_j$ to be the restriction of $\sim$ to the subcategory $F(\mathscr{G}_j[I_j^{-1}]) \subseteq F(\mathscr{G}[I^{-1}])$, as defined in Example A.7 (5)(i). Then every pair of $\sim_j$-related morphisms is $\sim$-related, and conversely, any two $\sim$-related morphisms must both be in the same component $F(\mathscr{G}_j[I_j^{-1}])$ for some $j \in J$ and would thus $\sim_j$-related. So $\mathfrak{G} = \coprod_{j\in J} (\mathfrak{G}_j, I \cap E_j, \sim_j)$.

Proposition 3.41. *Let J be a finite indexing set. Then let $\mathscr{G}_j = [E_j \rightrightarrows V_j]$ be a graph and $\mathfrak{G}_j = (\mathscr{G}_j, I_j, \sim_j)$ a CNP for $j \in J$. If $N_j \subseteq V_j$ is a solution to $\mathfrak{G}_j$ for $j \in J$, then $\coprod_{j\in J} N_j$ is a solution to $\coprod_{j\in J} \mathfrak{G}_j$. Conversely, if N is a solution to $\coprod_{j\in J} \mathfrak{G}_j$, then $N \cap V_j$ is a solution to $\mathfrak{G}_j$.*

Proof. Each object and morphism in $F(\mathscr{G}[I^{-1}])$ is in some subcategory $F(\mathscr{G}_j[I_j^{-1}])$, and any two parallel morphisms in $F(\mathscr{G}[I^{-1}])$ are in the same $F(\mathscr{G}_j[I_j^{-1}])$, so the result follows immediately. $\square$

Next, we consider reduction to subproblems, showing how the solution to a subproblem may serve as a solution to the original problem.

Definition 3.42 ((Marked) subgraph, subproblem)**.** Let $\mathscr{G}_j = [\mathrm{dom}_j, \mathrm{cod}_j \colon E_j \rightrightarrows V_j]$ be a graph for $j = 1, 2$. As in Definition A.2, we say that $\mathscr{G}_1$ is a *subgraph* of $\mathscr{G}_2$ and write $\mathscr{G}_1 \subseteq \mathscr{G}_2$ if we have subset inclusions $V_1 \subseteq V_2$ and $E_1 \subseteq E_2$ as well as function restrictions $\mathrm{dom}_1 = \mathrm{dom}_2|_{E_1}$ and $\mathrm{cod}_1 = \mathrm{cod}_2|_{E_1}$.

Now fix $I_j \subseteq E_j$ so that $(\mathscr{G}_j, I_j)$ is a marked directed graph for $j = 1, 2$. We say that $(\mathscr{G}_1, I_1)$ is a *marked subgraph* of $(\mathscr{G}_2, I_2)$ and write $(\mathscr{G}_1, I_1) \subseteq (\mathscr{G}_2, I_2)$ if $\mathscr{G}_1$ is a subgraph of $\mathscr{G}_2$ and $I_1 \subseteq I_2$.

Finally, fix a congruence $\sim_j$ on $F(\mathscr{G}_j[I_j^{-1}])$ so that $\mathfrak{G}_j = (\mathscr{G}_j, I_j, \sim_j)$ is a CNP for $j = 1, 2$ with $(\mathscr{G}_1, I_1)$ a marked subgraph of $(\mathscr{G}_2, I_2)$. By Example A.7 (5)(i), we may restrict $\sim_2$ to the subcategory $F(\mathscr{G}_1[I_1^{-1}]) \subseteq F(\mathscr{G}_2[I_2^{-1}])$; abusing notation, we also denote this restriction by $\sim_2$. Then if $\sim_1$ is contained in $\sim_2$ as congruences on $F(\mathscr{G}_1[I_1^{-1}])$, we say that $\mathfrak{G}_1$ a *subproblem* of $\mathfrak{G}_2$ and write $\mathfrak{G}_1 \subseteq \mathfrak{G}_2$.

Given a CNP $\mathfrak{G} = (\mathscr{G}, J, \sim)$, we may obtain a subproblem $\mathfrak{G}' = (\mathscr{G}, I, \sim')$ with the same underlying graph by contracting the set of marked edges J to a subset $I \subset J$, then restricting $\sim$ to the subcategory $F(\mathscr{G}[I^{-1}]) \subset F(\mathscr{G}[J^{-1}])$ before possibly contracting it as well to a congruence $\sim'$ contained in $\sim$. The following proposition shows that a solution to the CNP $\mathfrak{G}'$ is also a solution to the ambient CNP $\mathfrak{G}$. In some cases, the subproblem may be easier to solve because passing from J to I makes the edges in $J \setminus I$ irreversible, making Diamond Lemma-style arguments possible. Our proof adapts Mac Lane's analysis of the top diagram on [16, p. 167].

Proposition 3.43. *Suppose $\mathscr{G} = [s, t \colon E \rightrightarrows V]$ is a graph and $(\mathscr{G}, I, \sim')$ is a subproblem of $(\mathscr{G}, J, \sim)$ (so $I \subseteq J \subseteq E$ and $\sim'$ is contained in the restriction of $\sim$). If $N \subseteq V$ is a solution to the CNP $(\mathscr{G}, I, \sim')$, then N is also a solution to $(\mathscr{G}, J, \sim)$.*

Proof. Given $t \in V$, there is a unique $n(t) \in N$ for which there is a morphism $\nu \colon t \Rightarrow n(t)$ in $F(\mathscr{G}[I^{-1}])$, so ν is in $F(\mathscr{G}[J^{-1}])$ as well.

Now suppose there is another morphism $\nu' \colon t \Rightarrow n' \in N$ in $F(\mathscr{G}[J^{-1}])$ arising from a path in $\mathscr{G}[J^{-1}]$ as depicted in the top row of the diagram

$$\begin{tikzcd} t = t_0 \arrow[r, Rightarrow, "e_1"] \arrow[d, Rightarrow, "\nu \sim_I \nu_0"'] \arrow[dr, phantom, "\sim"] & t_1 \arrow[r, Rightarrow, "e_2"] \arrow[d, Rightarrow, "\nu_1"] \arrow[dr, phantom, "\sim"] & \cdots \arrow[r, Rightarrow, "e_{k-1}"] \arrow[d, phantom, "\cdots"] \arrow[dr, phantom, "\sim"] & t_{k-1} \arrow[r, Rightarrow, "e_k"] \arrow[d, Rightarrow, "\nu_{k-1}"] \arrow[dr, phantom, "\sim"] & t_k = n' \arrow[d, Rightarrow, "\nu_k \sim_I \mathrm{id}_{n'}"] \\ n(t) = n(t_0) \arrow[r, equal] & n(t_1) \arrow[r, equal] & \cdots \arrow[r, equal] & n(t_{k-1}) \arrow[r, equal] & n(t_k) = n(n') = n', \end{tikzcd}$$

where each $e_i \colon t_{i-1} \Rightarrow t_i$ is either an edge in E or the inverse $e_i = j_i^{-1}$ of an edge $j_i \colon t_i \Rightarrow t_{i-1}$ in $J \subseteq E$, so $\nu' = e_k \bullet \cdots \bullet e_1$. We also have morphisms in $F(\mathscr{G}[I^{-1}])$ that we denote by $\nu_i \colon t_i \Rightarrow n(t_i)$ that are unique up to $\sim'$. Applying Lemma 3.2 to whichever of the edges $e_i \colon t_{i-1} \Rightarrow t_i$ or $j_i \colon t_i \Rightarrow t_{i-1}$ is in E, we conclude that $n(t_{i-1}) = n(t_i)$. Moreover, in the former case, $\nu_i \bullet e_i \colon t_{i-1} \Rightarrow n(t_i)$ is parallel to $\nu_{i-1} \colon t_{i-1} \Rightarrow n(t_{i-1})$ in $F(\mathscr{G}[I^{-1}])$, so $\nu_i \bullet e_i \sim' \nu_{i-1}$; while in the latter case, similar reasoning shows $\nu_{i-1} \bullet j_i \sim' \nu_i$, but viewing either side as morphisms in the ambient category $F(\mathscr{G}[J^{-1}])$ and composing each with $e_i = j_i^{-1}$ yields $\nu_{i-1} \sim \nu_i \bullet e_i$. By induction, we conclude that

$$n(t) = n(t_0) = \cdots = n(t_k) = n(n') = n',$$

with the final equality exhibited by the identity morphism $\mathrm{id}_{n'} \colon n' \Rightarrow n' \in N$, and

$$\nu \sim' \nu_0 \sim (\nu_1 \bullet e_1) \sim \cdots \sim (\nu_k \bullet e_k \bullet \cdots \bullet e_1) = (\nu_k \bullet \nu') \sim' (\mathrm{id}_{n'} \bullet \nu') = \nu',$$

so $\nu \sim \nu'$. $\square$

Combining our results thus far (taking $\varnothing$ in place of I and I in place of J above) yields the following.

Corollary 3.44. *Suppose $\mathscr{G} = [E \rightrightarrows V]$ is a graph and $\mathfrak{G} = (\mathscr{G}, I, \sim)$ is a CNP. Let $\sim'$ be a congruence on $F(\mathscr{G})$ contained in the restriction of $\sim$ to $F(\mathscr{G}) \subseteq F(\mathscr{G}[I^{-1}])$. If $\mathscr{G}$ has a strict weight and $(\mathscr{G}, \sim')$ is locally confluent, then the set $M = \{t \in V \mid t \text{ is } \mathscr{G}\text{-minimal}\}$ is a solution to $\mathfrak{G}$.*

Moreover, if $I = E$ (so every edge is marked), the inclusion functor from M viewed as a discrete category to the category $F(\mathscr{G}[E^{-1}])/{\sim}$ presented by $\mathfrak{G}$ is an equivalence.

Proof. By Lemma 3.17, $\mathscr{G}$ is well-founded, so by Lemma 3.20, $(\mathscr{G}, \sim')$ is confluent. It follows from Proposition 3.18 that M is a solution to $(\mathscr{G}, \varnothing, \sim')$. As $\varnothing \subseteq I$ and $\sim'$ is contained in the restriction of $\sim$ to $F(\mathscr{G})$, the CNP $(\mathscr{G}, \varnothing, \sim')$ is a subproblem of $\mathfrak{G}$, so Proposition 3.43 implies that M is also a solution to $\mathfrak{G}$. The case of $E = I$ follows from Theorem 3.5 (2). $\square$

Applying Proposition 3.43 to our CNPs for restricted coherence schemata yields the following.

Corollary 3.45. *Suppose $(\mathfrak{C}, \mathfrak{O})$ is a restricted coherence schema with $\mathfrak{C} = (\mathfrak{S}, \mathfrak{T}, \mathfrak{I}, \mathfrak{R})$ as in Example 2.44 and $V \coloneqq V(\mathfrak{O})$, and let $\mathfrak{N} \subseteq \mathfrak{O}$ be a solution to its coherent normalization problem $\mathrm{CN}(\mathfrak{C}, \mathfrak{O}) = (\mathscr{B}_{\mathfrak{O}}, \mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}, \langle R(\mathfrak{C})_V\rangle)$, as defined in Notation 3.6. Construct a new coherence schema $\overline{\mathfrak{C}} = (\mathfrak{S}, \mathfrak{T}, \mathcal{E}, \mathfrak{R})$, all of whose edges are marked, yielding a new coherent normalization problem $\mathrm{CN}(\overline{\mathfrak{C}}, \mathfrak{O}) = (\mathscr{B}_{\mathfrak{O}}, \mathrm{Bs}_{\mathfrak{O}}, \langle R(\overline{\mathfrak{C}})_V\rangle)$, where the entire edge-set $\mathrm{Bs}_{\mathfrak{O}}$ of $\mathscr{B}_{\mathfrak{O}}$ is marked. Then $\mathfrak{N}$ is also a solution to $\mathrm{CN}(\overline{\mathfrak{C}}, \mathfrak{O})$, and the inclusion functor from $\mathfrak{N}$ viewed as a discrete category to the universal parameter category $\mathscr{P}(\overline{\mathfrak{C}}, \mathfrak{O})$ is an equivalence.*

Proof. Set $V \coloneqq V(\mathfrak{O})$, and recall that $\langle R(\mathfrak{C})_V\rangle$ (resp. $\langle R(\overline{\mathfrak{C}})_V\rangle$) are initially constructed in Definition 2.46 on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ (resp. $F(\mathscr{B}_V[\mathrm{Bs}_V^{-1}])$) before being restricted to the subcategory $F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}^{-1}])$ (resp. $F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O}}^{-1}])$). As $\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}} \subseteq \mathrm{Bs}_{\mathfrak{O}}$, to apply Proposition 3.43, it suffices to show that the congruence $\langle R(\mathfrak{C})_V\rangle$ restricted to $F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}^{-1}]) \subseteq F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ is contained in the congruence $\langle R(\overline{\mathfrak{C}})_V\rangle$ restricted to $F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O},\mathfrak{I}}^{-1}]) \subseteq F(\mathscr{B}_{\mathfrak{O}}[\mathrm{Bs}_{\mathfrak{O}}^{-1}]) \subseteq F(\mathscr{B}_V[\mathrm{Bs}_V^{-1}])$. With $\mathrm{Bs}_{V,\mathfrak{I}} \subseteq \mathrm{Bs}_V$ and the functoriality of congruence restriction, it is enough to show that $\langle R(\mathfrak{C})_V\rangle$ on $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}])$ as originally defined is contained in the restriction of $\langle R(\overline{\mathfrak{C}})_V\rangle$ to $F(\mathscr{B}_V[\mathrm{Bs}_{V,\mathfrak{I}}^{-1}]) \subseteq F(\mathscr{B}_V[\mathrm{Bs}_V^{-1}])$. But this follows from the fact that $R(\mathfrak{C})_V \subseteq R(\overline{\mathfrak{C}})_V$, as changing only the marked edges from $\mathfrak{I}$ to $\mathcal{E}$ does not change the constituent relations $R^{\mathrm{fun}}, R^{\mathrm{nat}}$, and R^{com} and only expands the constituent relation R^{inv}.

Then the result follows from Proposition 3.43 and Theorem 3.5 (2). $\square$

Example 3.46. In Section 3.3, we defined a restricted coherence schema $(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k)$ with $\mathfrak{C}_{\mathrm{A}} = (\mathfrak{S}_{\otimes}, \mathfrak{T}_{\mathrm{A}}, \varnothing, \mathfrak{R}^{\mathrm{A}})$ parametrizing a binary operation with a *lax* associator satisfying a pentagon axiom, yielding a coherent normalization problem $\mathrm{CN}(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k)$ whose solution is the set $\mathcal{N}_k \subset \mathcal{M}_k$ of right-normalized terms defined in Definition 3.26. The associator was lax because the sole edge label α for $\mathfrak{T}_{\mathrm{A}}$ was unmarked. But we may mark it to obtain a new coherence schema $\overline{\mathfrak{C}_{\mathrm{A}}} = (\mathfrak{S}_{\otimes}, \mathfrak{T}_{\mathrm{A}}, \{\alpha\}, \mathfrak{R}^{\mathrm{A}})$ parametrizing a binary operation with an *invertible* associator satisfying a pentagon axiom.

Then by Corollary 3.45, $\mathcal{N}_k$ is also a solution to the coherent normalization problem $\mathrm{CN}(\overline{\mathfrak{C}_{\mathrm{A}}}, \mathcal{M}_k)$. Furthermore, the inclusion functor from $\mathcal{N}_k$ viewed as a discrete category to the universal parameter category $\mathscr{P}(\overline{\mathfrak{C}_{\mathrm{A}}}, \mathcal{M}_k)$ is an equivalence. Following Theorem 3.7 (2), we may extend this to a coherence theorem for every model of $\overline{\mathfrak{C}_{\mathrm{A}}}$.

We now consider when the set of noninvertible relations may be reduced, using a relative variant of the Diamond Lemma. First we weaken our notion of weights.

Definition 3.47 (Weak weight, relative local confluence)**.** Suppose $\mathscr{G} = [\mathrm{dom}, \mathrm{cod}\colon E \rightrightarrows V]$ is a directed graph, $(\mathscr{G}, I, \sim)$ is a CNP, W is a set with a well-founded relation R, and $w\colon V \to W$ is a function such that, for every edge $e\colon t \Rightarrow u$ in $\mathscr{G}$, either $w(t) = w(u)$ or $w(t)\, R\, w(u)$ (note that well-foundedness implies these are necessarily mutually exclusive). Define

$$\begin{aligned} E_0 &= \left\{ e \in E \,\middle|\, w(\mathrm{dom}(e)) = w(\mathrm{cod}(e)) \right\}, \\ I_0 &= \left\{ i \in I \,\middle|\, w(\mathrm{dom}(i)) = w(\mathrm{cod}(i)) \right\} = I \cap E_0, \\ \text{and} \quad E_> &= \left\{ e \in E \,\middle|\, w(\mathrm{dom}(e))\, R\, w(\mathrm{cod}(e)) \right\} = E \setminus E_0. \end{aligned}$$

Let $\mathscr{G}_0$ (resp. $\mathscr{G}_>$) be the subgraph $[\mathrm{dom}, \mathrm{cod}\colon E_0 \rightrightarrows V]$ (resp. $[\mathrm{dom}, \mathrm{cod}\colon E_> \rightrightarrows V]$) of $\mathscr{G}$ with all vertices in V but only the edges in E_0 (resp. $E_>$). We say that w is a *weak weight* on $(\mathscr{G}, I, \sim)$, denoted $w\colon (\mathscr{G}, I, \sim) \to (W, R)$, if the following conditions are satisfied.[5]

(1) For any edge $e\colon t \Rightarrow u$ in $\mathscr{G}_0$, the vertex t is $\mathscr{G}_>$-minimal if and only if the vertex u is $\mathscr{G}_>$-minimal.

(2) For any edge $e\colon t \Rightarrow t'$ in $\mathscr{G}_0[I_0^{-1}]$ and edge $f'\colon t' \Rightarrow u'$ in $\mathscr{G}_>$, there exists an edge $f\colon t \Rightarrow u$ in $\mathscr{G}_>$ and a morphism $\varphi\colon u \Rightarrow u'$ in $F(\mathscr{G}[I_0^{-1}])$ such that $(f' \bullet e) \sim (\varphi \bullet f)$

[5] The intent behind these conditions is to guarantee that the existence of edges in E_0 in between the edges in $E_>$ will not disrupt arguments relying on the well-foundedness of $\mathscr{G}_>$ when, say, inducting on the vertices of $\mathscr{G}[I_0^{-1}]$ as a whole. By (2), every path in $\mathscr{G}[I_0^{-1}]$ is $\sim$-related to a path in $\mathscr{G}[I_0^{-1}]$ with all the edges in $E_>$ preceding all the edges in $E_0 \cup I_0^{-1}$. Then (1) guarantees that if the vertex at the end of the $E_>$ portion of this new path is $\mathscr{G}_>$-minimal, the vertex at the end of the entire path is $\mathscr{G}_>$-minimal, too.

as morphisms in $F(\mathscr{G}[I_0^{-1}])$, as depicted below, where we use dashed arrows for edges in $\mathscr{G}_0[I_0^{-1}]$ and tailed arrows for those in $\mathscr{G}_>$.

$$\begin{tikzcd} & t \arrow[dl, Rightarrow, dashed, "e"'] \arrow[dr, Rightarrow, tail, "\exists\, f"] & \\ t' \arrow[dr, Rightarrow, tail, "f'"'] & \sim & u \arrow[dl, Rightarrow, "\exists\, \varphi"] \\ & u' & \end{tikzcd}$$

Furthermore, we say that $(\mathscr{G}, I, \sim)$ *locally confluent relative to* w or simply w*-locally confluent* if, for any edges $e\colon t \Rightarrow a$ and $f\colon t \Rightarrow b$ in $\mathscr{G}_>$, there exist morphisms $\alpha\colon a \Rightarrow c$ and $\beta\colon b \Rightarrow c$ in $F(\mathscr{G}[I_0^{-1}])$ such that $(\alpha \bullet e) \sim (\beta \bullet f)$, as depicted below, where we continue to use tailed arrows for edges in $\mathscr{G}_>$.

$$\begin{tikzcd} & t \arrow[dl, Rightarrow, tail, "e"'] \arrow[dr, Rightarrow, tail, "f"] & \\ a \arrow[dr, Rightarrow, "\exists\, \alpha"'] & \sim & b \arrow[dl, Rightarrow, "\exists\, \beta"] \\ & c. & \end{tikzcd}$$

Remark 3.48. Every strict weight is also a weak weight: if $(\mathscr{G}, I, \sim)$ is a CNP and $w\colon \mathscr{G} \to (W, R)$ is a strict weight, the sets E_0 and I_0 defined above are empty, satisfying conditions (1–2) vacuously. Moreover $\mathscr{G}[I_0^{-1}] = \mathscr{G} = \mathscr{G}_>$, so $(\mathscr{G}, I, \sim)$ is w-locally confluent if and only if $(\mathscr{G}, \sim)$ is locally confluent.

The following result produces solutions to CNPs $(\mathscr{G}, I, \sim)$ equipped with weak weights, letting us restrict attention to $\mathscr{G}_>$-minimal elements and the edges between them.

Proposition 3.49. *Let* $\mathscr{G} = [E \rightrightarrows V]$ *be a graph and* $(\mathscr{G}, I, \sim)$ *be a CNP with a* weak *weight* $w\colon (\mathscr{G}, I, \sim) \to (W, R)$ *(with* $\mathscr{G}_0$ *and* $\mathscr{G}_>$ *defined as in Definition 3.47) such that* $(\mathscr{G}, I, \sim)$ *is* w*-locally confluent. Let*

$$M = \{t \in V \mid t \text{ is } \mathscr{G}_>\text{-minimal}\},$$

and let $\mathscr{G}_M$ *be the full subgraph of* $\mathscr{G}$ *spanned by* M*, let* I_M *be the intersection of* I *with the edges of* $\mathscr{G}_M$*, and denote the restriction of* $\sim$ *to* $F(\mathscr{G}_M[I_M^{-1}])$ *also by* $\sim$*, so that the CNP* $(\mathscr{G}_M, I_M, \sim)$ *is a subproblem of* $(\mathscr{G}, I, \sim)$*. If* N *is a solution to* $(\mathscr{G}_M, I_M, \sim)$*, then* N *is also a solution to* $(\mathscr{G}, I, \sim)$*.*

In preparation for the proof, we have the following lemma.

Lemma 3.50. *Under the hypotheses of Proposition 3.49, for $t \in M$, there is a unique element $n \in N$ for which there exists a morphism $t \Rightarrow n$ in $F(\mathscr{G}[I_0^{-1}])$. Furthermore, for every pair of morphisms $\nu, \nu' \colon t \Rightarrow n$ in $F(\mathscr{G}[I_0^{-1}])$, we have $\nu \sim \nu'$ for $\sim$ restricted to $F(\mathscr{G}[I_0^{-1}])$.*

As N is a solution to $(\mathscr{G}_M, I_M, \sim)$, we already know for $t \in M$ that there is a unique morphism $t \Rightarrow n(t) \in N$ in $F(\mathscr{G}_M[I_M^{-1}])$ up to $\sim$. This lemma expands the category $F(\mathscr{G}_M[I_M^{-1}])$ to $F(\mathscr{G}[I_0^{-1}])$, while Proposition 3.49 will expand M to V.

Proof of Lemma 3.50. Since N is a solution to $(\mathscr{G}_M, I_M, \sim)$, there is a unique element $n = n(t) \in N$ for which there exists a morphism $\nu \colon t \Rightarrow n$ in $F(\mathscr{G}_M[I_M^{-1}])$, and morphisms $t \Rightarrow n$ in $F(\mathscr{G}_M[I_M^{-1}])$ are unique up to $\sim$.

Given an edge $i \colon u \Rightarrow v$ in I_M, we have that $i \in I \subseteq E$ and that u is $\mathscr{G}_>$-minimal, so $i \notin E_>$ and thus $i \in E_0 \cap I = I_0$. Hence $I_M \subseteq I_0$, so $F(\mathscr{G}_M[I_M^{-1}]) \subseteq F(\mathscr{G}[I_0^{-1}])$. Therefore $\nu \colon t \Rightarrow n$ is also a morphism in $F(\mathscr{G}[I_0^{-1}])$.

Now suppose there is another morphism $\nu' \colon t \Rightarrow n' \in N$ in $F(\mathscr{G}[I_0^{-1}])$. We want to show $n = n'$ and $\nu \sim \nu'$. We can write $\nu' = e_k \bullet \cdots \bullet e_1$ for edges $e_j \colon t_{j-1} \Rightarrow t_j$ in $\mathscr{G}[I_0^{-1}]$ with $t = t_0, \ldots, t_k = n'$. So each e_j is either the inverse of an edge in $I_0 \subseteq E_0$ or an edge of $\mathscr{G}$. We claim that if e_j is an edge of $\mathscr{G}$, then e_j is an edge of $\mathscr{G}_0$: for if not, we can let ℓ be the least index with e_ℓ an edge of $\mathscr{G}_>$, so that e_j is an edge of $\mathscr{G}_0[I_0^{-1}]$ for $0 \le j < \ell$. Then applying Definition 3.47 (2) inductively yields in particular an edge with domain t in $\mathscr{G}_>$, contradicting the $\mathscr{G}_>$-minimality of t. Therefore every e_j is either an edge of $\mathscr{G}_0$ or the inverse of an edge of $\mathscr{G}_0$. As $t_0 = t$ is $\mathscr{G}_>$-minimal, applying Definition 3.47 (1) inductively then implies that every t_j is $\mathscr{G}_>$-minimal, so $t = t_0, \ldots, t_k = n' \in M$, and every e_j is either an edge of $\mathscr{G}_M$ or the inverse of an edge in I_M. Hence $\nu' \colon t \Rightarrow n' \in N$ is a morphism in $F(\mathscr{G}_M[I_M^{-1}])$. As N is a solution to $(\mathscr{G}_M, I_M, \sim)$, it follows that $n = n'$ and $\nu \sim \nu'$. $\square$

We are now ready to prove Proposition 3.49. The argument below adapts Mac Lane's analysis of the second diagram on [16, p. 167].

Proof of Proposition 3.49. Recall that $w\colon \mathscr{G} \to (W, R)$ is a weak weight making the CNP $(\mathscr{G}, I, \sim)$ w-locally confluent; that M is the set of all $\mathscr{G}_{>}$-minimal vertices of $\mathscr{G}$; that $\mathscr{G}_M$ is the full subgraph of $\mathscr{G}$ spanned by M; that I_M is the set of edges in both $\mathscr{G}_M$ and I; that $(\mathscr{G}_M, I_M, \sim)$ is a subproblem of $(\mathscr{G}, I, \sim)$; and that N is a solution to $(\mathscr{G}_M, I_M, \sim)$.

By Proposition 3.43, it suffices to show that N is a solution to the subproblem $(\mathscr{G}, I_0, \sim)$ with $\sim$ restricted to $F(\mathscr{G}[I_0^{-1}])$. We will show that every $t \in V$ has a unique N-normal form $n(t) \in N$ satisfying Definition 3.1 (1–2) for the CNP $(\mathscr{G}, I_0, \sim)$ by well-founded strong induction on $w(t) \in W$.

For our base case, suppose first that $w(t)$ is R-minimal. Then t is $\mathscr{G}_{>}$-minimal: for if not, there is an edge $t \Rightarrow u$ in $\mathscr{G}_{>}$ that w sends to $w(t)\, R\, w(u)$ in W, contradicting the R-minimality of $w(t)$. So $t \in M$, and by Lemma 3.50, t has a unique N-normal form $n(t)$ for the CNP $(\mathscr{G}, I_0, \sim)$.

Now let $>_R$ be the transitive closure of R, fix $t \in V$, and suppose inductively that for all $t' \in V$ with $w(t) >_R w(t')$, there is a unique N-normal form $n(t')$ of t'. We want to show that there is a unique N-normal form $n(t)$ of t. If t is $\mathscr{G}_{>}$-minimal, we are again done by Lemma 3.50. Otherwise, we have an edge $e\colon t \Rightarrow t'$ in $\mathscr{G}_{>}$, so $w(t)\, R\, w(t')$ in W, so by our inductive hypothesis there is a unique N-normal form $n(t')$ of t' equipped with a morphism $\nu\colon t' \Rightarrow n(t')$ unique up to $\sim$ in $F(\mathscr{G}[I_0^{-1}])$. Composing with e yields a morphism $(\nu \bullet e)\colon t \Rightarrow n(t')$ in $F(\mathscr{G}[I_0^{-1}])$, so $n(t')$ is our candidate for the N-normal form of t.

It remains to show uniqueness. Suppose we have a morphism $\nu'\colon t \Rightarrow n' \in N$ in $F(\mathscr{G}[I_0^{-1}])$. We want to show $n(t') = n'$ and $(\nu \bullet e) \sim \nu'$. We can write $\nu' = e_k \bullet \cdots \bullet e_1$ for edges $e_j\colon t_{j-1} \Rightarrow t_j$ in $\mathscr{G}[I_0^{-1}]$ with $t = t_0, \ldots, t_k = n'$. If every e_j were an edge of $\mathscr{G}_0[I_0^{-1}]$ and thus either an edge or the inverse of an edge in $\mathscr{G}_0$, since $t_k = n' \in N \subseteq M$ is $\mathscr{G}_{>}$-minimal, Definition 3.47 (1) implies inductively that $t_0 = t$ is $\mathscr{G}_{>}$-minimal, a contradiction. So some e_j is in $\mathscr{G}_{>}$: let ℓ be the least index such that e_ℓ is an edge of $\mathscr{G}_{>}$, so that e_j is an edge of $\mathscr{G}_0[I_0^{-1}]$ for $0 \leq j < \ell$. Then applying Definition 3.47 (2) inductively yields an edge $e'\colon t \Rightarrow t_1'$ in $\mathscr{G}_{>}$ and a morphism $\varphi\colon t_1' \Rightarrow n'$ in $F(\mathscr{G}[I_0^{-1}])$ such that $\nu' \sim (\varphi \bullet e')$. Now consider the diagram in $F(\mathscr{G}[I_0^{-1}])$ below:

$$\begin{tikzcd}
 & & t \arrow[dll, Rightarrow, "e"'] \arrow[drr, Rightarrow, "e'"] & & \\
t' \arrow[dd, Rightarrow, "\nu"'] \arrow[drr, Rightarrow, "\alpha"] & & \sim & & t_1' \arrow[dll, Rightarrow, "\beta"'] \arrow[dd, Rightarrow, "\varphi"] \\
 & \sim & c \arrow[d, Rightarrow, "\gamma"] & \sim & \\
n(t') \arrow[rr, equal] & & n(c) \arrow[rr, equal] & & n' = n(t_1').
\end{tikzcd}$$

Here $\alpha\colon t' \Rightarrow c$ and $\beta\colon t_1' \Rightarrow c$ arise from w-local confluence on edges $e\colon t \Rightarrow t'$ and $e'\colon t \Rightarrow t_1'$ of $\mathscr{G}_>$, with $(\alpha \bullet e) \sim (\beta \bullet e')$. Moreover, e' exhibits $w(t)\, R\, w(t_1')$ in W, so by induction, $n' = n(t_1')$ and φ is unique up to $\sim$ among morphisms $t_1' \Rightarrow n'$ in $F(\mathscr{G}[I_0^{-1}])$. Finally, writing $\beta = \beta_m \bullet \cdots \bullet \beta_1$ for edges $\beta_i\colon c_{i-1} \Rightarrow c_i$ of $\mathscr{G}[I_0^{-1}]$ with $t_1' = c_0$ and $c_m = c$, for each i either $w(c_{i-1}) = w(c_i)$ if $\beta_i \in E_0 \cup I_0^{-1}$ or $w(c_{i-1})\, R\, w(c_i)$ if $\beta_i \in E_>$, so $w(t)\, R\, w(t_1') = w(c_0)$ implies $w(t) >_R w(c_m) = w(c)$. So by strong induction, c has an N-normal form $n(c)$ with a morphism $\gamma\colon c \Rightarrow n(c)$ in $F(\mathscr{G}[I_0^{-1}])$. By the uniqueness conditions of N-normal forms, $n(t') = n(c) = n'$ as well as $\nu \sim (\gamma \bullet \alpha)$ and $\varphi \sim (\gamma \bullet \beta)$, so $(\nu \bullet e) \sim (\varphi \bullet e') \sim \nu'$. □

In the next section, we will see this result in action by translating the proof of Mac Lane's coherence theorem for monoidal categories to our formalism.

3.6 Coherence for monoidal categories

Building off of Section 3.3, we may now use Proposition 3.49 to prove Mac Lane's coherence theorem for monoidal categories from [16] using CNPs. We will first prove a coherence theorem for *lax* monoidal categories (as defined in Example 2.44), where our atomic transformations are not assumed to be invertible, then deduce from this a coherence theorem for standard monoidal categories, analogously to Example 2.44.

3.6.1 The coherence theorem for lax monoidal categories

We review the coherence schema $\mathfrak{C}_{\mathrm{LM}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing, \mathfrak{R}^{\mathrm{LM}})$ for lax monoidal categories that we constructed throughout the previous chapter, highlighting how it relates to the coherence schema $\mathfrak{C}_{\mathrm{A}} = (\mathfrak{S}_{\otimes}, \mathfrak{T}_{\mathrm{A}}, \varnothing, \mathfrak{R}^{\mathrm{A}})$ we constructed in Section 3.3.

- We use the signature for monoidal categories $\mathfrak{S}_{\mathrm{M}}$ from Example 2.2. Recall that it consists of:

  – one type, $T \coloneqq \{\mathtt{m}\}$;

  – one constant symbol, $C \coloneqq \{e\}$; and

  – one function symbol, $F \coloneqq \{\otimes\}$, with $\mathrm{ar}(\otimes) = 2$.

  In other words, $\mathfrak{S}_{\mathrm{M}}$ is $\mathfrak{S}_{\otimes}$ from Section 3.3 with an additional constant symbol e.

  We saw in Example 2.5 that $\mathfrak{S}_{\mathrm{M}}$ generates the language for monoidal categories $\mathcal{L}^{\mathrm{M}} = \mathcal{L}(\mathfrak{S}_{\mathrm{M}})$, consisting of:

  (1) the variable $x_i \coloneqq x_i^{\mathtt{m}}$ for $i > 0$;

  (2) the constant e; and

  (3) the term $\otimes(s, t)$ for $s, t \in \mathcal{L}^{\mathrm{M}}$.

  Note that the language $\mathcal{M}$ of $\mathfrak{S}_{\mathrm{M}}$, which is the free magma on the variables, is a proper subset of $\mathcal{L}^{\mathrm{M}}$, which is itself the free magma on the variables along with the constant e. That is, $\mathcal{M}$ is the submagma of $\mathcal{L}^{\mathrm{M}}$ consisting of all terms that do not contain the constant e.

- From Example 2.27, we use the edge labels $\mathcal{E}_{\mathrm{M}} \coloneqq \{\alpha, \lambda, \rho\}$ and the transformation schema for monoidal categories $\mathfrak{T}_{\mathrm{M}} \colon \mathcal{E}_{\mathrm{M}} \to \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$, with the domains and codomains of the edge labels given by

$$\alpha \colon \otimes(\otimes(x_1, x_2), x_3) \Rightarrow \otimes(x_1, \otimes(x_2, x_3)),$$
$$\lambda \colon \otimes(e, x_1) \Rightarrow x_1, \quad \text{and} \quad \rho \colon \otimes(x_1, e) \Rightarrow x_1,$$

  yielding as in Example 2.33 the graph $\mathscr{B}^{\mathrm{M}}$ whose vertices are terms in $\mathcal{L}^{\mathrm{M}}$ and whose edges are basic transformations in $\mathrm{Bs}^{\mathrm{M}} \coloneqq \mathrm{Bs}(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}})$. Explicitly, Bs^{M} is given as follows.

(1) For $s, t, u \in \mathcal{L}^{\mathrm{M}}$, there are atomic transformations in Bs^{M} given by

$$\alpha_{s,t,u} \colon \otimes(\otimes(s,t),u) \Rightarrow \otimes(s,\otimes(t,u)),$$
$$\lambda_s \colon \otimes(e,s) \Rightarrow s, \quad \text{and} \quad \rho_s \colon \otimes(s,e) \Rightarrow s.$$

(2) For $s, t, u \in \mathcal{L}^{\mathrm{M}}$ and $\varphi \colon s \Rightarrow t$ in Bs^{M}, there are basic transformations in Bs^{M} given by

$$\otimes(\varphi, \mathrm{id}_u) \colon \otimes(s,u) \Rightarrow \otimes(t,u) \quad \text{and} \quad \otimes(\mathrm{id}_u, \varphi) \colon \otimes(u,s) \Rightarrow \otimes(u,t).$$

Note that α has the same domain and codomain that it has in $\mathfrak{T}_{\mathrm{A}}$; that is, we may restrict $\mathfrak{T}_{\mathrm{M}} \colon \mathcal{E}_{\mathrm{M}} \to \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$ to the domain $\{\alpha\} \subset \mathcal{E}_{\mathrm{M}}$ and the codomain $\mathcal{M} \times \mathcal{M} \subset \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$ to obtain $\mathfrak{T}_{\mathrm{A}}$. Meanwhile, λ and ρ both have the constant e in their domains, so any basic transformation between terms in $\mathcal{M} \subseteq \mathcal{L}^{\mathrm{M}}$ must have label α and is in fact a basic transformation in $\mathrm{Bs}^{\mathrm{A}} = \mathrm{Bs}(\mathfrak{S}_{\otimes}, \mathfrak{T}_{\mathrm{A}})$. So $\mathscr{B}^{\mathrm{A}}$ is the full subgraph of $\mathscr{B}^{\mathrm{M}}$ spanned by $\mathcal{M}$.

- There are no marked edges.

- Following Example 2.44, the commutativity schema $\mathfrak{R}^{\mathrm{LM}}$ is the local binary relation on the free category $\mathscr{F}^{\mathrm{LM}} = F(\mathscr{B}^{\mathrm{M}})$ with five related pairs, corresponding to the following diagrams (subscripts suppressed):

$$\begin{tikzcd}
\otimes(\otimes(\otimes(x_1,x_2),x_3),x_4) \arrow[r, Rightarrow, "\otimes(\alpha,\mathrm{id})"] \arrow[dd, Rightarrow, "\alpha"'] & \otimes(\otimes(x_1,\otimes(x_2,x_3)),x_4) \arrow[d, Rightarrow, "\alpha"] \\
& \otimes(x_1,\otimes(\otimes(x_2,x_3),x_4)) \arrow[d, Rightarrow, "\otimes(\mathrm{id},\alpha)"] \\
\otimes(\otimes(x_1,x_2),\otimes(x_3,x_4)) \arrow[r, Rightarrow, "\alpha"] & \otimes(x_1,\otimes(x_2,\otimes(x_3,x_4))),
\end{tikzcd} \tag{3.51}$$

$$\begin{tikzcd}
\otimes(\otimes(x_1,e),x_2) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\otimes(\rho,\mathrm{id})"'] & & \otimes(x_1,\otimes(e,x_2)) \arrow[dl, Rightarrow, "\otimes(\mathrm{id},\lambda)"] \\
& \otimes(x_1,x_2). &
\end{tikzcd} \tag{3.52}$$

$$\begin{tikzcd} \otimes(\otimes(e,x_1),x_2) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\otimes(\lambda,\mathrm{id})"'] & & \otimes(e,\otimes(x_1,x_2)) \arrow[dl, Rightarrow, "\lambda"] \\ & \otimes(x_1,x_2); & \end{tikzcd} \tag{3.53}$$

$$\begin{tikzcd} \otimes(\otimes(x_1,x_2),e) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\rho"'] & & \otimes(x_1,\otimes(x_2,e)) \arrow[dl, Rightarrow, "\otimes(\mathrm{id},\rho)"] \\ & \otimes(x_1,x_2); & \end{tikzcd} \tag{3.54}$$

$$\begin{tikzcd} \otimes(e,e) \arrow[r, Rightarrow, bend left, "\lambda"] \arrow[r, Rightarrow, bend right, "\rho"'] & e. \end{tikzcd} \tag{3.55}$$

Note that the first diagram (3.51) is the same pentagon diagram from (3.22) defining $\mathfrak{R}^{\mathrm{A}}$, so $\mathfrak{R}^{\mathrm{A}} \subset \mathfrak{R}^{\mathrm{LM}}$ when $\mathfrak{R}^{\mathrm{LM}}$ is restricted to $\mathscr{F}^{\mathrm{A}} = F(\mathscr{B}^{\mathrm{A}})$.

So a model of this coherence schema $\mathfrak{C}_{\mathrm{LM}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing, \mathfrak{R}^{\mathrm{LM}})$ in $\underline{\mathbf{Cat}}$ is a *lax monoidal category* with a unit object, a binary operation, a lax associator, and lax unitors that obey the axioms (3.51) and (3.52).

We still have $V_k = \{x_1 < \cdots < x_k\}$, and we restrict our attention to the set of all terms $\mathcal{L}^{\mathrm{M}}_k \subset \mathcal{L}^{\mathrm{M}}$ with scope V_k; note that $\mathcal{L}^{\mathrm{M}}_k \cap \mathcal{M} = \mathcal{M}_k$. Then $\mathcal{L}^{\mathrm{M}}_k$ induces the full subgraph $\mathscr{B}^{\mathrm{M}}_k \subset \mathscr{B}^{\mathrm{M}}$ with edge-set $\mathrm{Bs}^{\mathrm{M}}_k \subset \mathrm{Bs}^{\mathrm{M}}$, with subgraph $\mathscr{B}^{\mathrm{A}}_k \subset \mathscr{B}^{\mathrm{M}}_k$ induced by $\mathcal{M}_k \subset \mathcal{L}^{\mathrm{M}}_k$. We take the free category $\mathscr{F}^{\mathrm{LM}}_k \coloneqq F(\mathscr{B}^{\mathrm{M}}_k)$.

So we will work with the restricted coherence schema $(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}^{\mathrm{M}}_k)$ for lax monoidal categories in k variables, as defined at the end of Example 2.44. For this restricted coherence schema, following Definition 2.46, we obtain a local binary relation $R^{\mathrm{LM}}_k \coloneqq R(\mathfrak{C}_{\mathrm{LM}})_{V_k}$ on $\mathscr{F}^{\mathrm{LM}}_k$. Then $\mathscr{P}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}^{\mathrm{M}}_k) = \mathscr{F}^{\mathrm{LM}}_k/\langle R^{\mathrm{LM}}_k\rangle$ is the category presented by $(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}^{\mathrm{M}}_k)$. We will also consider the monomorphic category $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}^{\mathrm{M}}_k) = \mathscr{F}^{\mathrm{LM}}_k/\langle R^{\mathrm{LM}}_k\rangle^{\mathrm{L}}$ presented by $(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}^{\mathrm{M}}_k)$. With $\mathfrak{R}^{\mathrm{A}} \subset \mathfrak{R}^{\mathrm{LM}}$ and $\mathscr{F}^{\mathrm{A}}_k \subset \mathscr{F}^{\mathrm{LM}}_k$, the congruence $\langle R^{\mathrm{A}}_k\rangle$ generated by the local binary relation $R^{\mathrm{A}}_k = R(\mathfrak{C}_{\mathrm{A}})_{V_k}$ on $\mathscr{F}^{\mathrm{A}}_k$ is contained in the restriction of $\langle R^{\mathrm{LM}}_k\rangle$ to $\mathscr{F}^{\mathrm{A}}_k$.

We can now combine Section 3.3 and Proposition 3.49 to solve our restricted coherence schema's coherent normalization problems $\mathrm{CN}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}^{\mathrm{M}}_k) = (\mathscr{B}^{\mathrm{M}}_k, \varnothing, \langle R^{\mathrm{LM}}_k\rangle)$ as well as $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}^{\mathrm{M}}_k) = (\mathscr{B}^{\mathrm{M}}_k, \varnothing, \langle R^{\mathrm{LM}}_k\rangle^{\mathrm{L}})$. First, we define a function $U\colon \mathcal{L}^{\mathrm{M}} \to \mathbb{N}$ by setting $U(t)$ equal to the number of times the constant e appears in t for all $t \in \mathcal{L}^{\mathrm{M}}$, so that

$\mathcal{M} = U^{-1}(0)$. We also denote the restriction of U to the domain $\mathcal{L}_k^{\mathrm{M}} \subset \mathcal{L}^{\mathrm{M}}$ by U. Then we have the following lemma, which will go towards showing that U is a weak weight.

Lemma 3.56. *For any terms $s, t \in \mathcal{L}^{\mathrm{M}}$ and basic transformation $\varphi\colon s \Rightarrow t$ in Bs^{M}, we have $U(s) \geq U(t)$. In particular,*

(1) if the label of φ is $L(\varphi) = \alpha$ (i.e. φ is α-basic), we have $U(s) = U(t)$;

(2) otherwise, if $L(\varphi) \in \{\lambda, \rho\}$ (i.e. φ is λ- or ρ-basic), we have $U(s) = U(t) + 1$.

Proof. Induct on φ. If φ is atomic, the result holds by inspection. Otherwise, φ is either $\otimes(\varphi', \mathrm{id}_u)\colon \otimes(s, u) \Rightarrow \otimes(t, u)$ or $\otimes(\mathrm{id}_u, \varphi')\colon \otimes(u, s) \Rightarrow \otimes(u, t)$ for $u \in \mathcal{L}^{\mathrm{M}}$ and $\varphi' \in \mathrm{Bs}^{\mathrm{M}}$ with $L(\varphi) = L(\varphi')$, so the result holds by induction. □

So we may partition the edge-set $\mathrm{Bs}_k^{\mathrm{M}}$ of $\mathscr{B}_k^{\mathrm{M}}$ into two subsets

$$\begin{aligned} (\mathrm{Bs}_k^{\mathrm{M}})_0 &= \{\varphi\colon s \Rightarrow t \text{ in } \mathscr{B}_k^{\mathrm{M}} \mid U(s) = U(t)\} = \{\varphi \in \mathrm{Bs}_k^{\mathrm{M}} \mid L(\varphi) = \alpha\} \\ \text{and} \quad (\mathrm{Bs}_k^{\mathrm{M}})_> &= \{\varphi\colon s \Rightarrow t \text{ in } \mathscr{B}_k^{\mathrm{M}} \mid U(s) > U(t)\} = \{\varphi \in \mathrm{Bs}_k^{\mathrm{M}} \mid L(\varphi) \in \{\lambda, \rho\}\}. \end{aligned} \tag{3.57}$$

We let $(\mathscr{B}_k^{\mathrm{M}})_0$ (resp. $(\mathscr{B}_k^{\mathrm{M}})_>$) denote the subgraph of $\mathscr{B}_k^{\mathrm{M}}$ with all vertices in $\mathcal{L}_k^{\mathrm{M}}$ but only the edges in $(\mathrm{Bs}_k^{\mathrm{M}})_0$ (resp. $(\mathrm{Bs}_k^{\mathrm{M}})_>$). That is, $(\mathscr{B}_k^{\mathrm{M}})_0$ (resp. $(\mathscr{B}_k^{\mathrm{M}})_>$) is the graph of terms in $\mathcal{L}_k^{\mathrm{M}}$ and the α-basic (resp. λ- and ρ-basic) transformations between them. We will need the following characterization of $(\mathscr{B}_k^{\mathrm{M}})_>$-minimality.

Lemma 3.58. *The set of $(\mathscr{B}_k^{\mathrm{M}})_>$-minimal terms in $\mathcal{L}_k^{\mathrm{M}}$ is $\{e\} \sqcup \mathcal{M}_k$,*

Proof. By the construction of $(\mathscr{B}_k^{\mathrm{M}})_>$, a term $t \in \mathcal{L}_k^{\mathrm{M}}$ is *not* $(\mathscr{B}_k^{\mathrm{M}})_>$-minimal if and only if it is the domain of some λ- or ρ-basic transformation. Such a domain must contain a subterm of the form $\mathrm{dom}(\lambda_s) = \otimes(e, s)$ or $\mathrm{dom}(\rho_s) = \otimes(s, e)$ for some $s \in \mathcal{L}_k^{\mathrm{M}}$, and conversely any term that contains such a subterm is the domain of a λ- or ρ-basic transformation by Lemma 2.37. Evidently neither e nor any element of $\mathcal{M}_k$ contains such a subterm.

Conversely, we must show that $t \in \mathcal{L}_k^{\mathrm{M}} \setminus (\{e\} \sqcup \mathcal{M}_k)$ always contains a subterm of the form $\otimes(e, s)$ or $\otimes(s, e)$. We do so by induction on the length of t. Every atomic term in $\mathcal{L}_k^{\mathrm{M}}$ is already in $\{e\} \sqcup \mathcal{M}_k$, so we may write $t = \otimes(u, v)$ for $u, v \in \mathcal{L}_k^{\mathrm{M}}$. If $u \notin \{e\} \sqcup \mathcal{M}_k$

or $v \notin \{e\} \sqcup \mathcal{M}_k$, we are done by induction, as any subterm of u or v is also a subterm of t. Otherwise, $u, v \in \{e\} \sqcup \mathcal{M}_k$. As $t \notin \mathcal{M}_k$, we cannot have $u, v \in \mathcal{M}_k$, so either $u = e$ or $v = e$, making t itself the subterm we seek. $\square$

We are now ready to show that U defines a weak weight on our CNP.

Lemma 3.59. *The function* $U\colon \mathcal{L}_k^{\mathrm{M}} \to \mathbb{N}$ *yields a weak weight* $(\mathscr{B}_k^{\mathrm{M}}, \varnothing, \langle R_k^{\mathrm{LM}} \rangle) \to (\mathbb{N}, >)$.

Proof. We follow Definition 3.47. By Lemma 3.56, every edge $\varphi\colon s \Rightarrow t$ in $\mathscr{B}_k^{\mathrm{M}}$ satisfies $U(s) \geq U(t)$, and we may partition the edge-set of $\mathscr{B}_k^{\mathrm{M}}$ into the α-basic transformations $(\mathrm{Bs}_k^{\mathrm{M}})_0$ and the λ- and ρ-basic transformations $(\mathrm{Bs}_k^{\mathrm{M}})_>$ as in (3.57), comprising the graphs $(\mathscr{B}_k^{\mathrm{M}})_0$ and $(\mathscr{B}_k^{\mathrm{M}})_>$, respectively. We must verify conditions (1) and (2) from Definition 3.47.

(1) Let $\varphi\colon t \Rightarrow u$ be an edge in $(\mathscr{B}_k^{\mathrm{M}})_0$; we must show that t is $(\mathscr{B}_k^{\mathrm{M}})_>$-minimal if and only if u is. We have $t, u \in \mathcal{L}_k^{\mathrm{M}}$ and φ an α-basic transformation, so $U(t) = U(u)$; and by Lemma 3.58, it suffices to show that t is in $\{e\} \sqcup \mathcal{M}_k$ if and only if u is. As the domain and codomain of an α-basic transformation, $t, u \notin \{e\}$. Then since $\mathcal{M}_k$ is precisely the set of terms in $\mathcal{L}_k^{\mathrm{M}}$ that U sends to 0, we have that $t \in \mathcal{M}_k$ if and only if $u \in \mathcal{M}_k$ because $U(t) = U(u)$.

(2) Let $\varphi\colon t \Rightarrow t'$ be an edge in $(\mathscr{B}_k^{\mathrm{M}})_0$ and $\psi'\colon t' \Rightarrow u'$ be an edge in $(\mathscr{B}_k^{\mathrm{M}})_>$. Then $t, t', u' \in \mathcal{L}_k^{\mathrm{M}}$; φ is an α-basic transformation; and ψ' is either a λ- or a ρ-basic transformation. We seek $u \in \mathcal{L}_k^{\mathrm{M}}$, a λ- or ρ-basic transformation $\psi\colon t \Rightarrow u$, and a morphism $\nu\colon u \Rightarrow u'$ of the free category $\mathscr{F}_k^{\mathrm{LM}}$ such that $(\psi' \bullet \varphi)\ \langle R_k^{\mathrm{LM}} \rangle\ (\nu \bullet \psi)$ as morphisms in $\mathscr{F}_k^{\mathrm{LM}}$:

$$\begin{tikzcd}
 & t \arrow[dl, Rightarrow, dashed, "\varphi"'] \arrow[dr, Rightarrow, "\exists\,\psi"] & \\
t' \arrow[dr, Rightarrow, "\psi'"'] & \langle R_k^{\mathrm{LM}} \rangle & u \arrow[dl, Rightarrow, "\exists\,\nu"] \\
 & u'. &
\end{tikzcd}$$

As usual, we induct on the length of t, which is equal to the length of t', and perform case analysis on the positions of the spans $S(\varphi)$ (with $\mathrm{dom}(A(\varphi))$ the $S(\varphi)$-subterm of t and $\mathrm{cod}(A(\varphi))$ the $S(\varphi)$-subterm of t') and $S(\psi')$ (with $\mathrm{dom}(A(\psi'))$ the $S(\psi')$-subterm of t') as subintervals of $[1, \mathrm{len}(t')]$.

There are no basic transformations with atomic domains, so assume $t' = \otimes(v', w')$. If $S(\varphi)$ and $S(\psi')$ are both in the left entry v' or both in the right entry w', the recursive definition of R_k^{LM} implies that we are done by induction. On the other hand, if one span is in the left entry and the other is in the right, the result follows from the construction of R^{fun}, exhibiting bifunctoriality of the interpretation of $\otimes$: we take ψ to be the basic transformation with the same atom and span as ψ', and we take ν to be the basic transformation with the same atom and span as φ.

By Corollaries 2.16 and 2.17, the remaining possibility to consider is if either φ or ψ' is an atomic transformation. That is, either $\varphi = \alpha_{p,q,r}$ for some $p, q, r \in \mathcal{L}_k^{\mathrm{M}}$, or ψ' is either λ_s or ρ_s for some $s \in \mathcal{L}_k^{\mathrm{M}}$.

First assume $\varphi = \alpha_{p,q,r}$, so $t' = \otimes(p, \otimes(q, r))$. If the span $S(\psi')$ is entirely in the left entry p, the middle entry q, or the right entry r, the result follows from the construction of R^{nat}, exhibiting naturality of the interpretation of α. Otherwise, by Corollary 2.17, ψ' is either $\lambda_{\otimes(q,r)}$ (with $p = e$), $\otimes(\mathrm{id}_p, \lambda_r)$ (with $q = e$), or $\otimes(\mathrm{id}_p, \rho_q)$ (with $r = e$). But then the result follows from R^{com} applied to the $\mathfrak{R}^{\mathrm{LM}}$-relations defined by (3.53), (3.52), or (3.54), respectively.

The remaining case is that ψ' is either λ_s or ρ_s, making t' either $\otimes(e, s)$ or $\otimes(s, e)$. If the span $S(\varphi)$ is entirely in the entry s, the result follows from the construction of R^{nat}, exhibiting the naturality of the interpretation of λ or ρ. Otherwise, by Corollary 2.17, φ must be atomic, a case we have already considered. $\square$

Next, we follow the remainder of Definition 3.47 to prove that our CNP is U-locally confluent. The structure of the proof is similar.

Lemma 3.60. *The coherent normalization problem* $\mathrm{CN}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}}) = (\mathscr{B}_k^{\mathrm{M}}, \varnothing, \langle R_k^{\mathrm{LM}} \rangle)$ *is U-locally confluent.*

Proof. Let $\varphi\colon t \Rightarrow a$ and $\psi\colon t \Rightarrow b$ be edges in $(\mathscr{B}_k^{\mathrm{M}})_>$, so $t, a, b \in \mathcal{L}_k^{\mathrm{M}}$ and each of φ and ψ is either a λ- or ρ-basic transformation. We seek $c \in \mathcal{L}_k^{\mathrm{M}}$ and morphisms $\alpha\colon a \Rightarrow c$ and

$\beta\colon b \Rightarrow c$ of the free category $\mathscr{F}_k^{\mathrm{LM}}$ such that $(\alpha \bullet \varphi)\,\langle R_k^{\mathrm{LM}}\rangle\,(\beta \bullet \psi)$. We induct on the length of t and perform case analysis on the spans $S(\varphi)$ and $S(\psi)$.

There are no basic transformations with atomic domains, so assume $t = \otimes(v,w)$. If $S(\varphi)$ and $S(\psi)$ are both in the left entry v or both in the right entry w, we are done by induction. If one span is in the left entry and the other is in the right, we are done by R^{fun}. Otherwise, either φ or ψ is atomic; assume without loss of generality that φ is, so $\varphi\colon t \Rightarrow a$ is either $\lambda_s\colon \otimes(e,s) \Rightarrow s$ or $\rho_s\colon \otimes(s,e) \Rightarrow s$ for some $s \in \mathcal{L}_k^{\mathrm{M}}$. If the span $S(\psi)$ is entirely in the entry s, we are done by R^{nat}. Otherwise, ψ is atomic as well. If φ and ψ have the same label, then $\varphi = \psi$, and we are done; so the only remaining case is when one of φ and ψ is $\lambda_e\colon \otimes(e,e) \Rightarrow e$ and the other is $\rho_e\colon \otimes(e,e) \Rightarrow e$. But these are $\mathfrak{R}^{\mathrm{LM}}$-related by (3.55). $\square$

We are now ready to apply Proposition 3.49 to solve $\mathrm{CN}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}}) = (\mathscr{B}_k^{\mathrm{M}}, \varnothing, \langle R_k^{\mathrm{LM}}\rangle)$. Following the hypotheses for that proposition, consider the full subgraph of $\mathscr{B}_k^{\mathrm{M}}$ spanned by the $(\mathscr{B}_k^{\mathrm{M}})_{>}$-minimal terms $\{e\} \sqcup \mathcal{M}_k$. As the subgraph of $\mathscr{B}_k^{\mathrm{M}}$ induced by $\mathcal{M}_k$ is simply $\mathscr{B}_k^{\mathrm{A}}$, while there are no basic transformations between e and the terms in $\mathcal{M}_k$, the full subgraph of $\mathscr{B}_k^{\mathrm{M}}$ induced by $\{e\} \sqcup \mathcal{M}_k$ is simply the disjoint union of $\mathscr{B}_k^{\mathrm{A}}$ with a graph containing one vertex $\{e\}$ and no edges; we denote this graph by $\{e\} \sqcup \mathscr{B}_k^{\mathrm{A}}$. Then we have a subproblem $(\{e\} \sqcup \mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{LM}}\rangle)$ of $(\mathscr{B}_k^{\mathrm{M}}, \varnothing, \langle R_k^{\mathrm{LM}}\rangle)$, where the former $\langle R_k^{\mathrm{LM}}\rangle$ is obtained by restricting the latter $\langle R_k^{\mathrm{LM}}\rangle$ to $F(\{e\} \sqcup \mathscr{B}_k^{\mathrm{A}})$. So Proposition 3.49 yields the following result.

Proposition 3.61. *A solution to* $(\{e\} \sqcup \mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{LM}}\rangle)$ *is also a solution to* $\mathrm{CN}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$ *and thus to* $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$.

Proof. The result follows from Lemmas 3.58 to 3.60 and Proposition 3.49. $\square$

It remains to solve $(\{e\} \sqcup \mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{LM}}\rangle)$. Recall from Definition 3.26 the definition of the set of right-normalized terms $\mathcal{N}_k \subset \mathcal{M}_k$ and from Corollary 3.28 that $\mathcal{N}_k$ is a solution to the coherent normalization problem $\mathrm{CN}(\mathfrak{C}_{\mathrm{A}}, \mathcal{M}_k) = (\mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{A}}\rangle)$.

Proposition 3.62. *The set* $\{e\} \sqcup \mathcal{N}_k$ *is a solution to* $(\{e\} \sqcup \mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{LM}}\rangle)$ *and thus to* $\mathrm{CN}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$ *and* $\mathrm{CN}_{\mathrm{L}}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$.

Proof. We have that $\mathcal{N}_k$ is a solution to $(\mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{A}}\rangle)$, and since $\langle R_k^{\mathrm{A}}\rangle \subset \langle R_k^{\mathrm{LM}}\rangle$ when the latter is restricted to $\mathscr{B}_k^{\mathrm{A}}$, it follows that $\mathcal{N}_k$ is a solution to $(\mathscr{B}_k^{\mathrm{A}}, \varnothing, \langle R_k^{\mathrm{LM}}\rangle)$ as well.

Meanwhile, if the graph of a CNP has vertex-set $\{e\}$ and no edges, then $\{e\}$ itself is a solution to that CNP. Then the result follows from Proposition 3.41. $\square$

Then applying Theorem 3.7 yields the following coherence theorem for our coherence schema $\mathfrak{C}_{\mathrm{LM}}$ (compare to Laplaza's [12, Theorem 5]).

Theorem 3.63. *For each $t \in \mathcal{L}_k^{\mathrm{M}}$, there is a unique $n(t) \in \{e\} \sqcup \mathcal{N}_k$ such that there is a morphism $t \Rightarrow n(t)$ in the universal parameter category $\mathscr{P}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$ for lax monoidal categories in k variables, and there is a unique morphism $t \Rightarrow n(t)$ in $\mathscr{P}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$. Moreover, the monomorphic universal parameter category $\mathscr{P}^{\mathrm{L}}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$ for lax monoidal categories is a preorder category: every diagram in it commutes.*

3.6.2 The coherence theorem for monoidal categories

It turns out the solution to our CNP for lax monoidal categories is also a solution to our CNP for monoidal categories. We review the definition of our coherence schema $\mathfrak{C}_{\mathrm{M}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{M}})$ for monoidal categories from Example 2.44 below, comparing it to our definition of the coherence schema for lax monoidal categories $\mathfrak{C}_{\mathrm{LM}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \varnothing, \mathfrak{R}^{\mathrm{LM}})$.

- We use the same signature for monoidal categories $\mathfrak{S}_{\mathrm{M}}$, generating the same language for monoidal categories $\mathcal{L}^{\mathrm{M}}$.

- We use the same set of edge labels $\mathcal{E}_{\mathrm{M}} = \{\alpha, \lambda, \rho\}$ and the same transformation schema for monoidal categories $\mathfrak{T}_{\mathrm{M}} \colon \mathcal{E}_{\mathrm{M}} \to \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$, yielding the graph $\mathscr{B}^{\mathrm{M}}$ whose vertices are terms in $\mathcal{L}^{\mathrm{M}}$ and whose edges are basic transformations in Bs^{M}.

- We mark every edge in $\mathcal{E}_{\mathrm{M}}$, thus marking every edge in the edge-set Bs^{M} of $\mathscr{B}^{\mathrm{M}}$.

- Following Example 2.44, the commutativity schema $\mathfrak{R}^{\mathrm{M}}$ is the local binary relation on the free category $\mathscr{F}^{\mathrm{M}} = F(\mathscr{B}^{\mathrm{M}}[(\mathrm{Bs}^{\mathrm{M}})^{-1}])$ with two related pairs, corresponding to the first two diagrams (3.51) and (3.52) that define $\mathfrak{R}^{\mathrm{LM}}$ on the subcategory $\mathscr{F}^{\mathrm{LM}} = F(\mathscr{B}^{\mathrm{M}})$.

A model of this coherence schema $\mathfrak{C}_{\mathrm{M}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{M}})$ in **Cat** is a *monoidal category*. Again we restrict to the set of all terms $\mathcal{L}_k^{\mathrm{M}}$ with scope $V_k = \{x_1 < \cdots < x_k\}$, inducing the full subgraph $\mathscr{B}_k^{\mathrm{M}}$ with edge-set $\mathrm{Bs}_k^{\mathrm{M}}$, yielding free category $\mathscr{F}_k^{\mathrm{M}} \coloneqq F(\mathscr{B}_k^{\mathrm{M}}[(\mathrm{Bs}_k^{\mathrm{M}})^{-1}])$. So we will work with the restricted coherence schema $(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}})$ for monoidal categories in k variables from the end of Example 2.44, with local binary relation $R_k^{\mathrm{M}} \coloneqq R(\mathfrak{C}_{\mathrm{M}})_{V_k}$ on $\mathscr{F}_k^{\mathrm{M}}$ as in Definition 2.46. Then $\mathscr{P}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}}) = \mathscr{F}_k^{\mathrm{M}} / \langle R_k^{\mathrm{M}} \rangle$ is the category presented by $(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}})$, and the corresponding coherent normalization problem is $\mathrm{CN}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}}) = (\mathscr{B}_k^{\mathrm{M}}, \mathrm{Bs}_k^{\mathrm{M}}, \langle R_k^{\mathrm{M}} \rangle)$ with all edges marked.

Having produced a solution $\{e\} \sqcup \mathcal{N}_k$ to $\mathrm{CN}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{M}_k)$, we may invoke Corollary 3.45 to show that $\{e\} \sqcup \mathcal{N}_k$ is also the solution to $\mathrm{CN}(\overline{\mathfrak{C}_{\mathrm{LM}}}, \mathcal{M}_k) = (\mathscr{B}_k^{\mathrm{M}}, \mathrm{Bs}_k^{\mathrm{M}}, \langle R(\overline{\mathfrak{C}_{\mathrm{LM}}})_{V_k} \rangle)$, where $\overline{\mathfrak{C}_{\mathrm{LM}}}$ is nearly the same coherence schema as $\mathfrak{C}_{\mathrm{LM}}$, but with every edge marked: $\overline{\mathfrak{C}_{\mathrm{LM}}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{LM}})$. We see that $\overline{\mathfrak{C}_{\mathrm{LM}}}$ differs from $\mathfrak{C}_{\mathrm{M}}$ only in its commutativity schema: $\mathfrak{R}^{\mathrm{LM}}$ contains $\mathfrak{R}^{\mathrm{M}}$ but also three additional relations. A standard argument due to Kelly [10], however, shows that when every edge in $\mathcal{E}_{\mathrm{M}}$ is marked, the additional relations in $\mathfrak{R}^{\mathrm{LM}}$ become redundant, with $\langle R_k^{\mathrm{M}} \rangle$ generated by $\mathfrak{R}^{\mathrm{M}}$ equal to $\langle R(\overline{\mathfrak{C}_{\mathrm{LM}}})_{V_k} \rangle$ generated by $\mathfrak{R}^{\mathrm{LM}}$ on $\mathscr{F}_k^{\mathrm{M}}$.

Lemma 3.64 (Kelly)**.** *We have an equality of congruences $\langle R_k^{\mathrm{M}} \rangle = \langle R(\overline{\mathfrak{C}_{\mathrm{LM}}})_{V_k} \rangle$ on $\mathscr{F}_k^{\mathrm{M}}$, so $\mathrm{CN}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}}) = \mathrm{CN}(\overline{\mathfrak{C}_{\mathrm{LM}}}, \mathcal{M}_k)$.*

Proof. By Lemma 2.50, it suffices to show $R^{\mathrm{com}}(\overline{\mathfrak{C}_{\mathrm{LM}}})_{V_k} \subset \langle R_k^{\mathrm{M}} \rangle$. The $R^{\mathrm{com}}(\overline{\mathfrak{C}_{\mathrm{LM}}})_{V_k}$-relations generated by the two $\mathfrak{R}^{\mathrm{M}}$-relations in $\mathfrak{R}^{\mathrm{LM}}$ are necessarily also in $\langle R_k^{\mathrm{M}} \rangle$, so it remains to show via diagram chase that the $R^{\mathrm{com}}(\overline{\mathfrak{C}_{\mathrm{LM}}})_{V_k}$-relations generated by each of the three $\mathfrak{R}^{\mathrm{LM}}$-relations not in $\mathfrak{R}^{\mathrm{M}}$ are in $\langle R_k^{\mathrm{M}} \rangle$. We consider each in turn for terms $s, t \in \mathcal{L}_k^{\mathrm{M}}$.

- We need to show that (3.53) with s substituted for x_1 and t for x_2 commutes up to $\langle R_k^{\mathrm{M}} \rangle$; that is,

$$(\lambda_{\otimes(s,t)} \bullet \alpha_{e,s,t}) \;\langle R_k^{\mathrm{M}} \rangle\; \otimes(\lambda_s, \mathrm{id}_t).$$

First, consider the following diagram (subscripts suppressed):

$$\begin{tikzcd}
\otimes(\otimes(\otimes(e,e),s),t) \arrow[rrrr, Rightarrow, "\otimes(\alpha,\mathrm{id})"] \arrow[drr, Rightarrow, "\otimes(\otimes(\rho,\mathrm{id}),\mathrm{id})"'] \arrow[ddd, Rightarrow, "\alpha"'] & & & & \otimes(\otimes(e,\otimes(e,s)),t) \arrow[dll, Rightarrow, "\otimes(\otimes(\mathrm{id},\lambda),\mathrm{id})"] \arrow[ddd, Rightarrow, "\alpha"] \\
& & \otimes(\otimes(e,s),t) \arrow[d, Rightarrow, "\alpha"] & & \\
& & \otimes(e,\otimes(s,t)) & & \\
\otimes(\otimes(e,e),\otimes(s,t)) \arrow[rr, Rightarrow, "\alpha"'] \arrow[urr, Rightarrow, "\otimes(\rho,\mathrm{id})"] & & \otimes(e,\otimes(e,\otimes(s,t))) \arrow[u, Rightarrow, "\otimes(\mathrm{id},\lambda)"] & & \otimes(e,\otimes(\otimes(e,s),t)). \arrow[ll, Rightarrow, "\otimes(\mathrm{id},\alpha)"] \arrow[ull, Rightarrow, "\otimes(\mathrm{id},\otimes(\lambda,\mathrm{id}))"']
\end{tikzcd}$$

The outer pentagon is obtained by substituting (e,e,s,t) for (x_1,x_2,x_3,x_4) in the $\mathfrak{R}^{\mathrm{M}}$-relation (3.51), while the upper triangle is obtained by applying $\otimes(\cdot,\mathrm{id}_t)$ to the $\mathfrak{R}^{\mathrm{M}}$-relation (3.52) with e substituted for x_1 and s for x_2. The left and right quadrangles commute up to $R^{\mathrm{nat}} \subset R_k^{\mathrm{M}}$. Finally, the lower left triangle is obtained by substituting e for x_1 and $\otimes(s,t)$ for x_2 in (3.52). Since every edge is marked, every morphism is invertible up to $R^{\mathrm{inv}} \subset R_k^{\mathrm{M}}$, so the lower right triangle must also commute up to $\langle R_k^{\mathrm{M}}\rangle$.

Observe that this lower right triangle is obtained by applying $\otimes(\mathrm{id}_e,\cdot)$ to the triangle we want to show commutes up to $\langle R_k^{\mathrm{M}}\rangle$. We have for $u,v \in \mathcal{L}_k^{\mathrm{M}}$ and basic transformation $\varphi\colon u \Rightarrow v$ that

$$\begin{tikzcd}
\otimes(e,u) \arrow[r, Rightarrow, "\lambda"] \arrow[d, Rightarrow, "\otimes(\mathrm{id},\varphi)"'] & u \arrow[d, Rightarrow, "\varphi"] \\
\otimes(e,v) \arrow[r, Rightarrow, "\lambda"] & v
\end{tikzcd}$$

commutes up to R^{nat}, and λ_u is invertible up to R^{inv}, so it follows that our desired triangle commutes up to $\langle R_k^{\mathrm{M}}\rangle$.

- The proof that (3.54) with s substituted for x_1 and t for x_2 commutes up to $\langle R_k^{\mathrm{M}}\rangle$ is analogous. We consider the diagram

$$
\begin{tikzcd}[arrows=Rightarrow]
\otimes(\otimes(s,\otimes(t,e)),e) \arrow[dr, "\otimes(\otimes(\mathrm{id},\rho),\mathrm{id})"'] \arrow[ddd, "\alpha"'] & \otimes(\otimes(\otimes(s,t),e),e) \arrow[l, "\otimes(\alpha,\mathrm{id})"'] \arrow[r, "\alpha"] \arrow[d, "\otimes(\rho,\mathrm{id})"] & \otimes(\otimes(s,t),\otimes(e,e)) \arrow[dl, "\otimes(\mathrm{id},\lambda)"] \arrow[ddd, "\alpha"] \\
 & \otimes(\otimes(s,t),e) \arrow[d, "\alpha"] & \\
 & \otimes(s,\otimes(t,e)) & \\
\otimes(s,\otimes(\otimes(t,e),e)) \arrow[ur, "\otimes(\mathrm{id},\otimes(\rho,\mathrm{id}))"] \arrow[rr, "\otimes(\mathrm{id},\alpha)"'] & & \otimes(s,\otimes(t,\otimes(e,e))). \arrow[ul, "\otimes(\mathrm{id},\otimes(\mathrm{id},\lambda))"']
\end{tikzcd}
$$

The outer pentagon commutes up to R^{com} from (3.51); the upper right triangle commutes up to R^{com} from (3.52); the left and right quadrangles commute up to R^{nat}; and the lower triangle is obtained by applying $\otimes(\mathrm{id}_s,\cdot)$ to an R^{com}-relation from (3.52). With every edge marked, every morphism is invertible up to R^{inv}, so the upper left triangle commutes up to $\langle R_k^{\mathrm{M}}\rangle$. This triangle is obtained by applying $\otimes(\cdot,\mathrm{id}_e)$ to the triangle we want to show commutes up to $\langle R_k^{\mathrm{M}}\rangle$, so by R^{nat} and the invertibility of ρ-basic transformations up to R^{inv}, our desired triangle commutes up to $\langle R_k^{\mathrm{M}}\rangle$.

- We need to show that (3.55) commutes up to $\langle R_k^{\mathrm{M}}\rangle$; that is, $\lambda_e \; \langle R_k^{\mathrm{M}}\rangle \; \rho_e$. We know

$$
\begin{tikzcd}[arrows=Rightarrow]
\otimes(e,\otimes(e,e)) \arrow[r, "\lambda"] \arrow[d, "\otimes(\mathrm{id},\lambda)"'] & \otimes(e,e) \arrow[d, "\lambda"] \\
\otimes(e,e) \arrow[r, "\lambda"] & e
\end{tikzcd}
$$

commutes up to R^{nat}, and λ is invertible up to R^{inv}, so the top arrow and the left arrow in the square above are $\langle R_k^{\mathrm{M}}\rangle$-related. These comprise the inner pair of parallel arrows below:

$$
\begin{tikzcd}[arrows=Rightarrow]
\otimes(\otimes(e,e),e) \arrow[r, "\alpha"] \arrow[rr, bend left=50, "\otimes(\lambda,\mathrm{id})"] \arrow[rr, bend right=50, "\otimes(\rho,\mathrm{id})"'] & \otimes(e,\otimes(e,e)) \arrow[r, bend left, "\lambda"] \arrow[r, bend right, "\otimes(\mathrm{id},\lambda)"'] & \otimes(e,e).
\end{tikzcd}
$$

The upper region in the preceding diagram is obtained by substituting e for both x_1 and x_2 in the $\mathfrak{R}^{\mathrm{LM}}$-relation given by (3.53), which we have already shown is also in $\langle R_k^{\mathrm{M}} \rangle$; while the lower region is obtained by substituting e for both x_1 and x_2 in the $\mathfrak{R}^{\mathrm{M}}$-relation given by (3.52). Therefore the outer pair of parallel arrows above, which is in turn the inner pair of parallel arrows below, are $\langle R_k^{\mathrm{M}} \rangle$-related:

$$
\begin{tikzcd}
\otimes(e,e) \arrow[rrr, Rightarrow, bend left=40, "\lambda"] \arrow[rrr, Rightarrow, bend right=40, "\rho"'] & \otimes(\otimes(e,e),e) \arrow[l, Rightarrow, "\rho"'] \arrow[r, Rightarrow, bend left, "\otimes(\lambda,\mathrm{id})"] \arrow[r, Rightarrow, bend right, "\otimes(\rho,\mathrm{id})"'] & \otimes(e,e) \arrow[r, Rightarrow, "\rho"] & e.
\end{tikzcd}
$$

The upper and lower regions both commute up to R^{nat}, and ρ is invertible up to R^{inv}, so the outer pair of parallel arrows are $\langle R_k^{\mathrm{M}} \rangle$-related, as desired. $\square$

Then applying Corollary 3.45 to Proposition 3.62 recovers Mac Lane's coherence theorem for monoidal categories, stated in our terminology as follows.

Theorem 3.65 (Mac Lane). *The set $\{e\} \sqcup \mathcal{N}_k$ is a solution to $\mathrm{CN}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}})$, and the inclusion functor from $\{e\} \sqcup \mathcal{N}_k$ viewed as a discrete category to the universal parameter category $\mathscr{P}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}})$ for monoidal categories in k variables is an equivalence.*

That is, for every term $t \in \mathcal{L}_k^{\mathrm{M}}$, there exists a unique term $n(t) \in \{e\} \sqcup \mathcal{N}_k$ such that there is a morphism $t \Rightarrow n(t)$ in $\mathscr{P}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}})$; and given terms $t, t' \in \mathcal{L}_k^{\mathrm{M}}$ with $n(t) = n(t')$, there exists a unique morphism $t \Rightarrow t'$ in $\mathscr{P}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}})$

So let $|\cdot| \colon \mathfrak{C}_{\mathrm{M}} \to \underline{\mathbf{K}}$ be a model for $\mathfrak{C}_{\mathrm{M}}$ in a cartesian monoidal 2-category $\underline{\mathbf{K}}$, giving rise to the functor

$$|\cdot|_{V_k} \colon \mathscr{P}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}}) \to \underline{\mathbf{K}}(|\mathtt{m}|^k, |\mathtt{m}|)$$

from (2.88). Then every diagram in $\underline{\mathbf{K}}(|\mathtt{m}|^k, |\mathtt{m}|)$ that factors through $|\cdot|_{V_k}$ is a commutative diagram of isomorphisms.

Proof. By Proposition 3.62, the set $\{e\} \sqcup \mathcal{N}_k$ is a solution to $\mathrm{CN}(\mathfrak{C}_{\mathrm{LM}}, \mathcal{L}_k^{\mathrm{M}})$, so by Corollary 3.45, $\{e\} \sqcup \mathcal{N}_k$ is also a solution to the coherent normalization problem $\mathrm{CN}(\overline{\mathfrak{C}_{\mathrm{LM}}}, \mathcal{L}_k^{\mathrm{M}})$

with all edges marked, which by Lemma 3.64 is the same CNP as $\mathrm{CN}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_k^{\mathrm{M}})$. The rest follows from Theorem 3.7 (2). $\square$

If we want to prove Mac Lane's coherence theorem for *symmetric* monoidal categories, however, we need to introduce one more construction on our CNPs.

3.7 Quotient CNPs

Previously, to solve our CNPs, we have assigned well-founded weights to our vertices in such a way that our edges always travel to vertices with equal or lower weight, allowing us to induct on our graphs. We have done so by exploiting the directionality of the edges we have encountered: e.g. λ- and ρ-basic transformations always reduce the number of unit constant symbols, while α-basic transformations always bring terms "closer" to becoming right normalized. Unfortunately, edges like β-basic transformations do not feature such directionality. Nevertheless, we can deal with them via another technique: by "quotienting out" everything else.

Informally, if $J \subseteq I$, the quotient $\mathfrak{G}/J$ is obtained by forcing the edges in J to be identities, "strictifying" out the morphisms generated by J. But just as we can only quotient congruences out of monoids and normal subgroups out of groups, we must impose some conditions on J to ensure that quotienting out by J results in a new CNP.

Definition 3.66 (Normality, quotient CNP)**.** Suppose $\mathscr{G} = [\mathrm{dom}, \mathrm{cod}\colon E \rightrightarrows V]$ is a graph and $\mathfrak{G} = (\mathscr{G}, I, \sim)$ is a CNP. Given $J \subseteq I$, let $\mathscr{J} \coloneqq [\mathrm{dom}, \mathrm{cod}\colon J \rightrightarrows V]$ denote the subgraph of $\mathscr{G}$ with the same vertex-set but with its edge-set restricted to $J \subseteq I \subseteq E$. Then $\mathscr{J}[J^{-1}]$ is a (wide) subgraph of $\mathscr{G}[I^{-1}]$, so $F(\mathscr{J}[J^{-1}])$ is a (wide) subcategory of $F(\mathscr{G}[I^{-1}])$, and we may restrict $\sim$ to a congruence relation on $F(\mathscr{J}[J^{-1}])$ that we also denote by $\sim$. We say that J is *normal* in $\mathfrak{G}$ and write $J \lhd \mathfrak{G}$ if the subproblem $\mathfrak{J} \coloneqq (\mathscr{J}, J, \sim)$ of $\mathfrak{G}$ has a solution. Then given $J \lhd \mathfrak{G}$, the *quotient coherent normalization problem* $\mathfrak{G}/J \coloneqq (\overline{\mathscr{G}}, \overline{I}, \approx)$ is defined as follows.

(1) The directed graph $\overline{\mathscr{G}}$ is given by the following.

(i) Its vertex-set is $V/\Leftrightarrow$, with $v \Leftrightarrow v'$ if and only if there exists a morphism $v \Rightarrow v'$ in $F(\mathscr{J}[J^{-1}])$.[6] Here $\Leftrightarrow$ is an equivalence relation: reflexivity is exhibited by identities, transitivity is exhibited by composition, and symmetry is satisfied because every edge in $\mathscr{J}[J^{-1}]$ has a formal inverse going in the opposite direction. We denote the $\Leftrightarrow$-equivalence class of $v \in V$ in $V/\Leftrightarrow$ by $[v]$. Note that $V/\Leftrightarrow$ is in bijection with any solution-set N of $\mathfrak{J}$, with $[v] = [v']$ if and only if they share an N-normal form.

(ii) Its edge-set is $(E \backslash J)/\equiv$, with $e \equiv e'$ for $e\colon u \Rightarrow v$ and $e'\colon u' \Rightarrow v'$ in $E \backslash J$ whenever there are morphisms $\sigma\colon u \Rightarrow u'$ and $\tau\colon v \Rightarrow v'$ in $F(\mathscr{J}[J^{-1}])$ such that $(\tau \bullet e) \sim (e' \bullet \sigma)$ in $F(\mathscr{G}[I^{-1}])$. Here $\equiv$ is an equivalence relation: reflexivity is exhibited by identities, symmetry is satisfied because every morphism in $F(\mathscr{J}[J^{-1}])$ is invertible up to $\sim$, and transitivity is satisfied because the following rectangle commutes up to $\sim$:

$$\begin{tikzcd} u \arrow[r, Rightarrow, "\sigma"] \arrow[d, Rightarrow, "\varphi"'] \arrow[dr, phantom, "\wr"] & u' \arrow[r, Rightarrow, "\tau"] \arrow[d, Rightarrow, "\varphi'"] \arrow[dr, phantom, "\wr"] & u'' \arrow[d, Rightarrow, "e''"] \\ v \arrow[r, Rightarrow, "\tau"'] & v' \arrow[r, Rightarrow, "\tau'"'] & v''. \end{tikzcd}$$

We also denote the $\equiv$-equivalence class of $e \in E \setminus J$ in $(E \setminus J)/\equiv$ by $[e]$.

(iii) The domain and codomain of an edge $[e] \in (E \setminus J)/\equiv$ with $e\colon u \Rightarrow v$ are $[u]$ and $[v]$. This is well-defined: if $[e] = [e']$ for $e'\colon u \Rightarrow v$, then there are morphisms $\sigma\colon u \Rightarrow u'$ and $\tau\colon v \Rightarrow v'$ in $F(\mathscr{J}[J^{-1}])$, so $u \Leftrightarrow u'$ and $v \Leftrightarrow v'$ and thus $[u] = [u']$ and $[v] = [v']$.

(2) We take $\overline{I} = \{[i] \mid i \in I \setminus J\}$ (recall that $[i]$ is the residue of i in our new edge-set $(E \setminus J)/\equiv$). In particular, if $J = I$, then $\overline{I} = \varnothing$.

We now define a local binary relation $\approx$ on $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ as follows. First, we may extend our function $[\cdot]$ on the vertices in V and the edges in $E \setminus J$ to the edges in J as well, giving rise to a graph morphism

$$[\cdot]\colon \mathscr{G} \to F(\overline{\mathscr{G}}[\overline{I}^{-1}])$$

[6] The idea is that we treat the edges in J as identities, so any pair of vertices connected by edges in J should be in the same equivalence class.

sending each vertex $v \in V$ to its residue $[v] \in V/\Leftrightarrow$, each edge $e \in E \setminus J$ to its residue $[e] \in (E \setminus J)/\equiv$, and each edge $j \in J$ to the identity on the residue $[\mathrm{dom}(j)] = [\mathrm{cod}(j)]$ (as j witnesses $\mathrm{dom}(j) \Leftrightarrow \mathrm{cod}(j)$). Then $[\cdot]$ further extends to a graph morphism

$$[\cdot]\colon \mathscr{G}[I^{-1}] \to F(\overline{\mathscr{G}}[\overline{I}^{-1}])$$

sending $i^{-1} \in (I \setminus J)^{-1}$ to $[i]^{-1} \in \overline{I}^{-1}$ and $j^{-1} \in J^{-1}$ to $\mathrm{id}_{[\mathrm{dom}(j)]} = \mathrm{id}_{[\mathrm{cod}(j)]}$. Extending freely gives a functor

$$[\cdot]\colon F(\mathscr{G}[I^{-1}]) \to F(\overline{\mathscr{G}}[\overline{I}^{-1}]).$$

Henceforth when we write $[\cdot]$, we refer to this functor, which subsumes all previous definitions of $[\cdot]$. We use the functor $[\cdot]$ to define $\approx$ as follows.

(3) Given parallel morphisms p, q in $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$, we declare $p \approx q$ whenever there are morphisms $\widetilde{p} \sim \widetilde{q}$ in $F(\mathscr{G}[I^{-1}])$ such that $[\widetilde{p}] = p$ and $[\widetilde{q}] = q$.

As suggested by the terminology, the quotient of a CNP by a normal set of marked edges is itself a CNP.

Lemma 3.67. *If $\mathfrak{G}$ is a CNP and $J \lhd \mathfrak{G}$ is normal, then $\mathfrak{G}/J$ is a CNP.*

Before we prove this, we list a number of other useful results about our above construction of quotient CNPs.

Lemma 3.68. *If $\mathfrak{G} = (\mathscr{G}, I, \sim)$ is a CNP and $J \lhd \mathfrak{G}$ is normal, the constructions in Definition 3.66 satisfy the following properties.*

(1) Every pair of parallel morphisms in $F(\mathscr{J}[J^{-1}])$ are $\sim$-related.

(2) Every morphism in $F(\mathscr{J}[J^{-1}]) \subseteq F(\mathscr{G}[I^{-1}])$ is mapped by $[\cdot]$ to an identity morphism in $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$.

(3) Every object and morphism in $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ is in the image of the functor $[\cdot]$.

(4) Consider a morphism φ of $F(\mathscr{G}[I^{-1}])$ such that, among the edges that comprise it, exactly n of them are neither in J nor J^{-1}. That is, we can write

$$\varphi = j_n \bullet e_n^{\pm 1} \bullet \cdots \bullet e_2^{\pm 1} \bullet j_1 \bullet e_1^{\pm 1} \bullet j_0$$

for morphisms j_i of $F(\mathscr{J}[J^{-1}])$ and edges $e_i \in E \setminus J$. Then $[\varphi]$ is a morphism in $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ comprised of n edges, namely

$$[\varphi] = [e_n]^{\pm 1} \bullet \cdots \bullet [e_1]^{\pm 1}.$$

So if ψ is another morphism of $F(\mathscr{G}[I^{-1}])$ for which $[\varphi] = [\psi]$, then among the edges that comprise ψ, it is also true that exactly n of them are neither in J nor J^{-1}.

(5) If $\varphi, \psi \colon u \Rightarrow v$ are parallel morphisms in $F(\mathscr{G}[I^{-1}])$ for which $[\varphi] = [\psi]$, then $\varphi \sim \psi$.

(6) If $\varphi \colon u \Rightarrow v$ and $\varphi' \colon u' \Rightarrow v'$ are morphisms in $F(\mathscr{G}[I^{-1}])$ for which $[\varphi] = [\varphi']$, then there exist morphisms $\sigma \colon u \Rightarrow u'$ and $\tau \colon v \Rightarrow v'$ in $F(\mathscr{J}[J^{-1}])$ such that $(\tau \bullet \varphi) \sim (\varphi' \bullet \sigma)$.

(7) If $\varphi, \psi \colon u \Rightarrow v$ are parallel morphisms in $F(\mathscr{G}[I^{-1}])$ for which $[\varphi] \approx [\psi]$, then $\varphi \sim \psi$.

Proof of (1). The CNP $\mathfrak{J} = (\mathscr{J}, J, \sim)$, with every edge marked, has a solution. So by Theorem 3.5, the category $F(\mathscr{J}[J^{-1}])/\sim$ it presents is a preorder category, and the result follows. □

Proof of (2). The edges in J and J^{-1} are all sent by $[\cdot]$ to identities. As $[\cdot]$ is a free extension, and identities compose to identities, all morphisms in $F(\mathscr{J}[J^{-1}])$ are sent to identities as well. □

Proof of (3). On objects, $[\cdot]$ is the quotient map $V \to V/\Leftrightarrow$ and therefore surjective. It follows that every identity in $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ is in the image of $[\cdot]$ as well. We can then proceed by induction on the length of the morphisms of $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ as paths. On edges in $E \setminus J$, the functor $[\cdot]$ is the quotient map to $(E \setminus J)/\equiv$, so every edge in this latter set is in the image of π. Similarly, every edge in $\overline{I}^{-1}$ can be written as $[i]^{-1} = [i^{-1}]$ for some $i \in I \setminus J$. Hence every edge in $\overline{\mathscr{G}}[\overline{I}^{-1}]$ is in the image of $[\cdot]$, completing the base case.

Now for $n > 0$, assume every morphism of $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ comprised of n edges of $\overline{\mathscr{G}}[\overline{I}^{-1}]$ is in the image of $[\cdot]$. We wish to show that every morphism of $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ comprised of $(n+1)$ edges of $\overline{\mathscr{G}}[\overline{I}^{-1}]$ is also in the image of $[\cdot]$. We may write such a morphism as $(f \bullet \nu)$ for a morphism ν of $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ comprised of n edges of $\overline{\mathscr{G}}[\overline{I}^{-1}]$ and an $(n+1)^{\text{th}}$ edge f of $\overline{\mathscr{G}}[\overline{I}^{-1}]$ with $\operatorname{cod}(\nu) = \operatorname{dom}(f)$. Then we may write $\nu = [\varphi]$ for some $\varphi\colon u \Rightarrow v$ in $F(\mathscr{G}[\overline{I}^{-1}])$, while f is either $[e]$ for some $e \in E \setminus J$ or $[i]^{-1}$ for some $i \in I \setminus J$.

In the first case, $e\colon v' \Rightarrow w$ is an edge in $E \setminus J$, making $[e]\colon [v'] \Rightarrow [w]$ an edge of $\mathscr{G}$ and a morphism of $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ with $[v] = [v']$. Hence there is a morphism $j\colon v \Rightarrow v'$ in $F(\mathscr{J}[J^{-1}]) \subseteq F(\mathscr{G}[I^{-1}])$, so by (2),

$$[e \bullet j \bullet \varphi] = [e] \bullet \mathrm{id}_{[v]} \bullet [\varphi] = [e] \bullet [\varphi] = f \bullet \nu.$$

Hence $(f \bullet \nu)$ is in the image of $[\cdot]$. Similarly, in the second case, $i\colon x \Rightarrow v''$ is an edge in $I \setminus J \subset I$, so that:

- $i^{-1}\colon v'' \Rightarrow x$ an element of I^{-1} and a morphism of $F(\mathscr{G}[I^{-1}])$;
- $[i]\colon [x] \Rightarrow [v'']$ is an element of $\overline{I}$ and a morphism of $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$; and
- $[i^{-1}] = [i]^{-1}\colon [v''] \Rightarrow [x]$ is an element of $\overline{I}^{-1}$ and a morphism of $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$;

with $[v] = [v'']$. So there is a morphism $j\colon v \Rightarrow v''$ in $F(\mathscr{J}[J^{-1}]) \subseteq F(\mathscr{G}[I^{-1}])$, and by (2),

$$[i^{-1} \bullet j \bullet \varphi] = [i]^{-1} \bullet \mathrm{id}_{[v]} \bullet [\varphi] = [i]^{-1} \bullet [\varphi] = f \bullet \nu.$$

Hence $(f \bullet \nu)$ in the image of $[\cdot]$. $\square$

Proof of (4). This follows from (2), how $[\cdot]$ is defined on $E \setminus J$ and $(I \setminus J)^{-1}$, and the freeness of both $F(\mathscr{G}[I^{-1}])$ and $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$. $\square$

Proof of (5). Say $\varphi, \psi\colon u \Rightarrow v$ are parallel morphisms in $F(\mathscr{G}[I^{-1}])$ for which $[\varphi] = [\psi]$. We induct on the length of $[\varphi] = [\psi]$ as a path. If $[\varphi]$ is an identity, by (4), φ and ψ must be morphisms in $F(\mathscr{J}[J^{-1}])$, so the result follows from (1).

Now for $n \geq 0$, say the result holds whenever $[\varphi]$ is comprised of n edges of $\overline{\mathscr{G}}[\overline{I}^{-1}]$, and assume instead $[\varphi]$ is comprised of $(n+1)$ edges. By (4), $[\varphi] = [\psi]$ implies that we can write

$$\varphi = j_{n+1} \bullet e_{n+1}^{\pm 1} \bullet j_n \bullet e_n^{\pm 1} \bullet \cdots \bullet e_1^{\pm 1} \bullet j_0 = j_{n+1} \bullet e_{n+1}^{\pm 1} \bullet \varphi'$$
$$\psi = k_{n+1} \bullet f_{n+1}^{\pm 1} \bullet k_n \bullet f_n^{\pm 1} \bullet \cdots \bullet f_1^{\pm 1} \bullet k_0 = k_{n+1} \bullet f_{n+1}^{\pm 1} \bullet \psi'.$$

for morphisms j_i, k_i of $F(\mathscr{J}[J^{-1}])$ and edges $e_i, f_i \in E \setminus J$, with each pair e_i, f_i bearing the same exponent and satisfying $[e_i] = [f_i]$, so $[\varphi'] = [\psi']$ as well. We have $\mathrm{dom}(\varphi') = u = \mathrm{dom}(\psi')$; we label the codomain of φ' by w and the codomain of ψ' by x; we label the domain of j_{n+1} by y and the domain of k_{n+1} by z; and $\mathrm{cod}(j_{n+1}) = v = \mathrm{cod}(k_{n+1})$, so that we have the following diagram:

$$\begin{tikzcd} u \arrow[r, Rightarrow, "\varphi'"] \arrow[d, equal] & w \arrow[r, Rightarrow, "e_{n+1}^{\pm 1}"] & y \arrow[r, Rightarrow, "j_{n+1}"] & v \arrow[d, equal] \\ u \arrow[r, Rightarrow, "\psi'"] & x \arrow[r, Rightarrow, "f_{n+1}^{\pm 1}"] & z \arrow[r, Rightarrow, "k_{n+1}"] & v \end{tikzcd}$$

With $[e_{n+1}] = [f_{n+1}]$, by construction, there exist $\sigma \colon w \Rightarrow x$ and $\tau \colon y \Rightarrow z$ in $F(\mathscr{J}[J^{-1}])$ making the middle square below commute up to $\sim$ (if the exponents are -1, we also use the fact that e_{n+1}^{-1} is the inverse of e_{n+1} and f_{n+1}^{-1} the inverse of f_{n+1} up to $\sim$):

$$\begin{tikzcd} u \arrow[r, Rightarrow, "\varphi'"] \arrow[d, equal] & w \arrow[r, Rightarrow, "e_{n+1}^{\pm 1}"] \arrow[d, Rightarrow, "\sigma"'] & y \arrow[r, Rightarrow, "j_{n+1}"] \arrow[d, Rightarrow, "\tau"] & v \arrow[d, equal] \\ u \arrow[r, Rightarrow, "\psi'"] & x \arrow[r, Rightarrow, "f_{n+1}^{\pm 1}"] & z \arrow[r, Rightarrow, "k_{n+1}"] & v \end{tikzcd}$$

Since the morphisms comprising the right square are all in $F(\mathscr{J}[J^{-1}])$, it, too, commutes up to $\sim$ by (1). Finally, since σ is in $F(\mathscr{J}[J^{-1}])$ as well, by (2),

$$[\sigma \bullet \varphi'] = [\sigma] \bullet [\varphi'] = [\varphi'] = [\psi']$$

of length n, so $(\sigma \bullet \varphi')$ and ψ' are parallel morphisms in $F(\mathscr{G}[I^{-1}])$ sent via $[\cdot]$ to the same morphism of length n. So by induction, $(\sigma \bullet \varphi') \sim \psi'$, and the result follows. □

Proof of (6). Let $\varphi \colon u \Rightarrow v$ and $\varphi' \colon u' \Rightarrow v'$ be morphisms in $F(\mathscr{G}[I^{-1}])$ for which $[\varphi] = [\varphi']$. Then $[\cdot]$ agrees on their domains and codomains as well: $[u] = [u']$ and $[v] = [v']$. It follows that there exist $\sigma \colon u \Rightarrow u'$ and $\tau \colon v \Rightarrow v'$ in $F(\mathscr{J}[J^{-1}])$. Then by (2),

$$[\tau \bullet \varphi] = [\tau] \bullet [\varphi] = [\varphi] = [\varphi'] = [\varphi'] \bullet [\sigma] = [\varphi' \bullet \sigma],$$

so $[\cdot]$ agrees on $(\tau \bullet \varphi), (\varphi' \bullet \sigma) \colon u \Rightarrow v'$ in $F(\mathscr{G}[I^{-1}])$. The result follows from (5). $\square$

Proof of (7). Let $\varphi, \psi \colon u \Rightarrow v$ be parallel morphisms in $F(\mathscr{G}[I^{-1}])$ for which $[\varphi] \approx [\psi]$. Then there exist $\varphi' \colon u' \Rightarrow v'$ and $\psi' \colon u' \Rightarrow v'$ for which $[\varphi] = [\varphi']$ and $[\psi] = [\psi']$ while $\varphi' \sim \psi'$. By (6), there also exist $\sigma, \sigma' \colon u \Rightarrow u'$ and $\tau, \tau' \colon v \Rightarrow v'$ in $F(\mathscr{J}[J^{-1}])$ such that the top and bottom regions below commute up to $\sim$:

$$\begin{tikzcd}
u \arrow[r, Rightarrow, bend left, "\sigma"] \arrow[r, Rightarrow, bend right, "\sigma'"'] \arrow[rrr, Rightarrow, bend left=60, "\varphi"] \arrow[rrr, Rightarrow, bend right=60, "\psi"'] & u' \arrow[r, Rightarrow, bend left, "\varphi'"] \arrow[r, Rightarrow, bend right, "\psi'"'] & v' & v \arrow[l, Rightarrow, bend right, "\tau"'] \arrow[l, Rightarrow, bend left, "\tau'"]
\end{tikzcd}$$

Also $\sigma \sim \sigma'$ and $\tau \sim \tau'$ by (1), and all four are invertible up to $\sim$, so the result follows. $\square$

Proof of Lemma 3.67. We have our marked directed graph $(\overline{\mathscr{G}}, \overline{I})$, so it remains to show that $\approx$ is a congruence relation on $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ satisfying Definition 3.1 (c).

First we show that $\approx$ is an equivalence relation on hom-sets. Reflexivity follows from Lemma 3.68 (3) and the reflexivity of $\sim$, and symmetry follows from symmetry of $\sim$. For transitivity, take parallel morphisms p, q, r in $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ along with morphisms $\widetilde{p}, \widetilde{q}, \widetilde{q'}, \widetilde{r}$ in $F(\mathscr{G}[I^{-1}])$ with $[\widetilde{p}] = p, [\widetilde{q}] = [\widetilde{q'}] = q$, and $[\widetilde{r}] = r$ such that $\widetilde{p} \sim \widetilde{q}$ and $\widetilde{q'} \sim \widetilde{r}$. By Lemma 3.68 (6), $[\widetilde{q}] = [\widetilde{q'}]$ implies that there exist morphisms σ and τ in $F(\mathscr{J}[J^{-1}])$ for which $(\sigma \bullet \widetilde{q}) \sim (\widetilde{q'} \bullet \tau)$, so $(\sigma \bullet \widetilde{p}) \sim (\widetilde{r} \bullet \tau)$. By Lemma 3.68 (2), $[\tau_0]$ and $[\sigma_n]$ are identities, so $[\sigma_n \bullet \widetilde{p}] = p$ and $[\widetilde{r} \bullet \tau_0] = r$. Hence $p \approx r$, proving transitivity.

It remains to show that in $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$, given composable morphisms p, q, r and morphism $q' \approx q$, we have $p \bullet q \bullet r \approx p \bullet q' \bullet r$. We must have morphisms $\widetilde{q} \sim \widetilde{q'}$, say $t \Rightarrow u$, in $F(\mathscr{G}[I^{-1}])$ with $[\widetilde{q}] = q$ and $[\widetilde{q'}] = q'$. Then by Lemma 3.68 (3), there exist $\widetilde{p} \colon u' \Rightarrow v$ and $\widetilde{r} \colon s \Rightarrow t'$ with $[u] = [u']$ and $[t'] = [t]$ such that $[\widetilde{p}] = p$ and $[\widetilde{r}] = r$. So we have morphisms $i \colon u \Rightarrow u'$ and $j \colon t' \Rightarrow t$ in $F(\mathscr{J}[J^{-1}])$, and by Lemma 3.68 (2),

$$[\widetilde{p} \bullet i \bullet \widetilde{q} \bullet j \bullet \widetilde{r}] = p \bullet q \bullet r \quad \text{and} \quad [\widetilde{p} \bullet i \bullet \widetilde{q'} \bullet j \bullet \widetilde{r}] = p \bullet q' \bullet r.$$

As $\sim$ is a congruence, $\widetilde{p} \bullet i \bullet \widetilde{q} \bullet j \bullet \widetilde{r} \sim \widetilde{p} \bullet i \bullet \widetilde{q'} \bullet j \bullet \widetilde{r}$, so $p \bullet q \bullet r \approx p \bullet q' \bullet r$.

Finally, we show that $\approx$ satisfies Definition 3.1 (c): for every edge $i\colon x \Rightarrow y$ in $I \setminus J$, we have $([i]^{-1} \bullet [i]) \approx \mathrm{id}_{[x]}$ and $([i] \bullet [i]^{-1}) \approx \mathrm{id}_{[y]}$ (recall that so far, the superscript -1's indicate formal symbols rather than actual categorical inverses). As $(\mathscr{G}, I, \sim)$ is a CNP, we have $(i^{-1} \bullet i) \approx \mathrm{id}_x$ and $(i \bullet i^{-1}) \approx \mathrm{id}_y$; moreover, $[i]^{-1} = [i^{-1}]$. Hence

$$([i]^{-1} \bullet [i]) = [i^{-1} \bullet i] \approx [\mathrm{id}_x] = \mathrm{id}_{[x]} \quad \text{and} \quad ([i] \bullet [i]^{-1}) = [i \bullet i^{-1}] \approx [\mathrm{id}_y] = \mathrm{id}_{[y]},$$

so $(\overline{\mathscr{G}}, \overline{I}, \approx)$ is a CNP. $\square$

We may now lift any solution of a quotient CNP to a solution of the original CNP.

Proposition 3.69. *Suppose $\mathfrak{G} = (\mathscr{G}, I, \sim)$ is a CNP and $J \lhd \mathfrak{G}$ is normal. Let $\mathscr{J} = [\mathrm{dom}, \mathrm{cod}\colon J \rightrightarrows V]$ denote the subgraph of $\mathscr{G}$ with edge-set restricted to J, and let $\mathfrak{J}$ denote the CNP $(\mathscr{J}, J, \sim)$ with $\sim$ restricted to $F(\mathscr{J}[J^{-1}])$. If $\widetilde{N}$ is a solution to $\mathfrak{J}$ and N is a solution to $\mathfrak{G}/J$, then $M = \{v \in \widetilde{N} \mid [v] \in N\}$ is a solution to $\mathfrak{G}$.*

Proof. Fix a vertex t of $\mathscr{G}$ with residue $[t]$, a vertex of $\overline{\mathscr{G}}$. As N is a solution to $\mathfrak{G}/J$, there is a unique $n[t] \in N$ with a morphism $\overline{\nu}\colon [t] \Rightarrow n[t]$ of $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$, with $\overline{\nu}$ unique up to $\approx$. We can write $\overline{\nu} = [e_m]^{\pm 1} \bullet \cdots \bullet [e_1]^{\pm 1}$ so that each $[e_k]^{\pm 1}\colon [t_{k-1}] \Rightarrow [t_k]$ is an edge of $\overline{\mathscr{G}}[\overline{I}^{-1}]$. In particular, for each $1 \le k \le m$, either $[e_k]$ has a positive exponent, so there is an edge $e_k\colon t'_{k-1} \Rightarrow t_k$ in $E \setminus J$, or $[e_k]$ has a negative exponent, so there is an edge $e_k\colon t_k \Rightarrow t'_{k-1}$ in $I \setminus J$; and we have that $t = t_0$, that $[t_k] = [t'_k]$ as exhibited by some morphism $j_k\colon t_k \Rightarrow t'_k$ in $F(\mathscr{J}[J^{-1}])$, and that $[t_m] = n[t]$. Finally, since $\widetilde{N}$ is a solution to $\mathfrak{J}$, there is a unique $\widetilde{n}(t_m) \in \widetilde{N}$ with a morphism $j\colon t_m \Rightarrow n(t_m)$ in $F(\mathscr{J}[J^{-1}])$, with j unique up to $\sim$. Composing yields a morphism

$$t = t_0 \overset{j_0}{\Longrightarrow} t'_0 \overset{e_1^{\pm 1}}{\Longrightarrow} t_1 \overset{j_1}{\Longrightarrow} t'_1 \overset{e_2^{\pm 1}}{\Longrightarrow} t_2 \overset{j_2}{\Longrightarrow} t'_2 \Rightarrow \cdots \Rightarrow t_{m-1} \overset{j_{m-1}}{\Longrightarrow} t'_{m-1} \overset{e_m^{\pm 1}}{\Longrightarrow} t_m \overset{j}{\Rightarrow} \widetilde{n}(t_m)$$

in $F(\mathscr{G}[I^{-1}])$ we call ν for which $[\nu] = \overline{\nu}$ by Lemma 3.68 (2), with $[\widetilde{n}(t_m)] = [t_m] = n[t] \in N$ and thus $\widetilde{n}(t_m) \in M$.

Now suppose we have a morphism $\nu'\colon t \Rightarrow n' \in M$ in $F(\mathscr{G}[I^{-1}])$; we want to show $n' = \widetilde{n}(t_m)$ and $\nu' \sim \nu$. We have that $[\cdot]$ sends ν' to a morphism $[\nu']\colon [t] \Rightarrow [n']$ in $F(\overline{\mathscr{G}}[\overline{I}^{-1}])$ with $[n'] \in N$; as N is a solution to $(\overline{\mathscr{G}}, \overline{I}, \approx)$, we have $[n'] = n[t] = [t_m]$ and $[\nu'] \approx \overline{\nu}\colon [t] \Rightarrow n[t]$. In particular, there is a morphism $t_m \Rightarrow n'$ in $F(\mathscr{J}[J^{-1}])$, but as $n' \in M \subseteq \widetilde{N}$ and $\widetilde{N}$ is a solution to $(\mathscr{J}, J, \sim)$, it follows that $n' = \widetilde{n}(t_m)$, making ν' and ν parallel morphisms satisfying $[\nu'] \approx \overline{\nu} = [\nu]$. Thus $\nu' \sim \nu$ by Lemma 3.68 (7). $\square$

We are now ready to apply our theory of quotient CNPs to proving Mac Lane's coherence theorem for symmetric monoidal categories.

3.8 Coherence for symmetric monoidal categories

Continuing from Section 3.6, we invoke Proposition 3.69 to prove Mac Lane's coherence theorem for symmetric monoidal categories from [16], using the coherence theorem for monoidal categories from Theorem 3.65 as well as the CNP from Proposition 3.32.

First, we review our construction throughout the previous chapter of the coherence schema $\mathfrak{C}_{\mathrm{SM}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{SM}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{LM}})$ for symmetric monoidal categories, highlighting how it relates to our coherence schema $\mathfrak{C}_{\mathrm{M}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{M}}, \mathcal{E}_{\mathrm{M}}, \mathfrak{R}^{\mathrm{M}})$ for monoidal categories.

- We use the same signature for monoidal categories $\mathfrak{S}_{\mathrm{M}}$ from Example 2.2, generating the language for monoidal categories $\mathcal{L}^{\mathrm{M}}$.

- From Example 2.27, we use the edge labels $\mathcal{E}_{\mathrm{SM}} \coloneqq \{\alpha, \beta, \lambda, \rho\}$ and the transformation schema for symmetric monoidal categories $\mathfrak{T}_{\mathrm{SM}} \colon \mathcal{E}_{\mathrm{SM}} \to \mathcal{L}^{\mathrm{M}} \times \mathcal{L}^{\mathrm{M}}$, with the domains and codomains of the edge labels given by

$$\alpha \colon \otimes(\otimes(x_1, x_2), x_3) \Rightarrow \otimes(x_1, \otimes(x_2, x_3)) \qquad \beta \colon \otimes(x_1, x_2) \Rightarrow \otimes(x_2, x_1)$$
$$\lambda \colon \otimes(e, x_1) \Rightarrow x_1 \qquad \rho \colon \otimes(x_1, e) \Rightarrow x_1$$

  yielding as in Example 2.33 the graph $\mathscr{B}^{\mathrm{SM}}$ whose vertices are terms in $\mathcal{L}^{\mathrm{M}}$ and whose edges are basic transformations in $\mathrm{Bs}^{\mathrm{SM}} \coloneqq \mathrm{Bs}(\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{SM}})$. Explicitly, $\mathrm{Bs}^{\mathrm{SM}}$ is given as follows.

  (1) For $s, t, u \in \mathcal{L}^{\mathrm{M}}$, there are atomic transformations in $\mathrm{Bs}^{\mathrm{SM}}$ given by

$$\alpha \colon \otimes(\otimes(s, t), u) \Rightarrow \otimes(s, \otimes(t, u)) \qquad \beta \colon \otimes(s, t) \Rightarrow \otimes(t, s)$$
$$\lambda \colon \otimes(e, s) \Rightarrow s \qquad \rho \colon \otimes(s, e) \Rightarrow s.$$

(2) For $s, t, u \in \mathcal{L}^{\mathrm{M}}$ and $\varphi\colon s \Rightarrow t$ in $\mathrm{Bs}^{\mathrm{SM}}$, there are basic transformations in $\mathrm{Bs}^{\mathrm{SM}}$ given by

$$\otimes(\varphi, \mathrm{id}_u)\colon \otimes(s,u) \Rightarrow \otimes(t,u) \quad \text{and} \quad \otimes(\mathrm{id}_u, \varphi)\colon \otimes(u,s) \Rightarrow \otimes(u,t).$$

Recall that $\mathfrak{T}_{\mathrm{M}}$ is the restriction of $\mathfrak{T}_{\mathrm{SM}}$ to $\mathcal{E}_{\mathrm{M}} = \{\alpha, \lambda, \rho\} \subset \{\alpha, \beta, \lambda, \rho\} = \mathcal{E}_{\mathrm{SM}}$, so $\mathrm{Bs}^{\mathrm{M}} \subset \mathrm{Bs}^{\mathrm{SM}}$, where Bs^{M} consists of all α-, λ-, and ρ-basic transformations in $\mathrm{Bs}^{\mathrm{SM}}$, and $\mathscr{B}^{\mathrm{M}}$ is the corresponding wide subgraph of $\mathrm{Bs}^{\mathrm{SM}}$.

- The marked edges are $\mathcal{E}_{\mathrm{M}} = \{\alpha, \lambda, \rho\}$, so every basic transformation in $\mathrm{Bs}^{\mathrm{M}} \subset \mathrm{Bs}^{\mathrm{SM}}$ is marked, giving rise to the graph $\mathscr{B}^{\mathrm{SM}}[(\mathrm{Bs}^{\mathrm{M}})^{-1}]$ and free category $\mathscr{F}^{\mathrm{SM}} = F(\mathscr{B}^{\mathrm{SM}}[(\mathrm{Bs}^{\mathrm{M}})^{-1}])$.

- Following Example 2.44, the commutativity schema $\mathfrak{R}^{\mathrm{SM}}$ is the local binary relation on the free category $\mathscr{F}^{\mathrm{SM}}$ with five related pairs, corresponding to the following diagrams (subscripts suppressed):

$$\begin{tikzcd}
\otimes(\otimes(\otimes(x_1,x_2),x_3),x_4) \arrow[r, Rightarrow, "\otimes(\alpha,\mathrm{id})"] \arrow[dd, Rightarrow, "\alpha"'] & \otimes(\otimes(x_1,\otimes(x_2,x_3)),x_4) \arrow[d, Rightarrow, "\alpha"] \\
 & \otimes(x_1,\otimes(\otimes(x_2,x_3),x_4)) \arrow[d, Rightarrow, "\otimes(\mathrm{id},\alpha)"] \\
\otimes(\otimes(x_1,x_2),\otimes(x_3,x_4)) \arrow[r, Rightarrow, "\alpha"] & \otimes(x_1,\otimes(x_2,\otimes(x_3,x_4)));
\end{tikzcd} \tag{3.70}$$

$$\begin{tikzcd}
\otimes(\otimes(x_1,e),x_2) \arrow[rr, Rightarrow, "\alpha"] \arrow[dr, Rightarrow, "\otimes(\rho,\mathrm{id})"'] & & \otimes(x_1,\otimes(e,x_2)) \arrow[dl, Rightarrow, "\otimes(\mathrm{id},\lambda)"] \\
 & \otimes(x_1,x_2); &
\end{tikzcd} \tag{3.71}$$

$$\begin{tikzcd}
\otimes(x_1,x_2) \arrow[rr, equal, "\mathrm{id}"] \arrow[dr, Rightarrow, "\beta"'] & & \otimes(x_1,x_2) \\
 & \otimes(x_2,x_1); \arrow[ur, Rightarrow, "\beta"'] &
\end{tikzcd} \tag{3.72}$$

$$\begin{tikzcd}
\otimes(x_1,e) \arrow[rr, Rightarrow, "\beta"] \arrow[dr, Rightarrow, "\rho"'] & & \otimes(e,x_1) \arrow[dl, Rightarrow, "\lambda"] \\
 & x_1; &
\end{tikzcd} \tag{3.73}$$

$$\begin{tikzcd} \otimes(\otimes(x_1,x_2),x_3) \arrow[r, Rightarrow, "\alpha"] \arrow[d, Rightarrow, "\otimes(\beta,\mathrm{id})"'] & \otimes(x_1,\otimes(x_2,x_3)) \arrow[r, Rightarrow, "\beta"] & \otimes(\otimes(x_2,x_3),x_1) \arrow[d, Rightarrow, "\alpha"] \\ \otimes(\otimes(x_2,x_1),x_3) \arrow[r, Rightarrow, "\alpha"] & \otimes(x_2,\otimes(x_1,x_3)) \arrow[r, Rightarrow, "\otimes(\mathrm{id},\beta)"] & \otimes(x_2,\otimes(x_3,x_1)). \end{tikzcd} \tag{3.74}$$

Here the first two diagrams (3.70) and (3.71) are the same as those that define $\mathfrak{R}^{\mathrm{M}}$.

So a model of this coherence schema $\mathfrak{C}_{\mathrm{SM}} = (\mathfrak{S}_{\mathrm{M}}, \mathfrak{T}_{\mathrm{SM}}, \{\alpha, \lambda, \rho\}, \mathfrak{R}^{\mathrm{SM}})$ in $\underline{\mathbf{Cat}}$ is precisely a *symmetric monoidal category.*

We still have $V_k = \{x_1 < \cdots < x_k\}$, but we shall restrict our attention to the set of all *regular* terms $\mathcal{L}_k^{\mathrm{M,reg}} \subset \mathcal{L}_k^{\mathrm{M}} \subset \mathcal{L}^{\mathrm{M}}$ with scope V_k, i.e. no repeating variables (repeating constants are allowed). Then $\mathcal{L}_k^{\mathrm{M,reg}}$ induces the full subgraph $\mathscr{B}_k^{\mathrm{SM}} \subset \mathscr{B}^{\mathrm{SM}}$ with edge-set $\mathrm{Bs}_k^{\mathrm{SM}} \subset \mathrm{Bs}^{\mathrm{SM}}$ and marked edge-set $\mathrm{Bs}_k^{\mathrm{M,reg}} \subset \mathrm{Bs}^{\mathrm{M}}$, the edge-set of the wide subgraph $\mathscr{B}_k^{\mathrm{M,reg}} \subset \mathscr{B}_k^{\mathrm{SM}}$. We take the free category $\mathscr{F}_k^{\mathrm{SM}} := F(\mathscr{B}_k^{\mathrm{SM}}[(\mathrm{Bs}_k^{\mathrm{M,reg}})^{-1}])$.

So we will work with the restricted regular coherence schema $(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}_k^{\mathrm{M,reg}})$ for symmetric monoidal categories in k variables, as defined at the end of Example 2.44. For this restricted coherence schema, following Definition 2.46, we obtain a local binary relation $R_k^{\mathrm{SM}} := R(\mathfrak{C}_{\mathrm{SM}})_{V_k}$ restricted to $\mathscr{F}_k^{\mathrm{SM}}$. Then $\mathscr{P}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}_k^{\mathrm{M,reg}}) = \mathscr{F}_k^{\mathrm{SM}}/\langle R_k^{\mathrm{SM}}\rangle$ is the category presented by $(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}_k^{\mathrm{M,reg}})$. We seek a solution to the corresponding CNP given by $\mathrm{CN}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}_k^{\mathrm{M,reg}}) = (\mathscr{B}_k^{\mathrm{SM}}, \mathrm{Bs}_k^{\mathrm{M,reg}}, \langle R_k^{\mathrm{SM}}\rangle)$. First, we decompose $\mathrm{CN}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}_k^{\mathrm{M,reg}})$ into a coproduct of CNPs as follows, writing $\underline{k} = \{1, 2, \ldots, k\}$.

Proposition 3.75. *We have an equality of CNPs*

$$\mathrm{CN}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}_k^{\mathrm{M,reg}}) = \coprod_{S \subseteq \underline{k}} (\mathscr{B}_S^{\mathrm{SM}}, \mathrm{Bs}_S^{\mathrm{M}}, \langle R_k^{\mathrm{SM}}\rangle),$$

where for each $S \subseteq \underline{k}$,

- $\mathscr{B}_S^{\mathrm{SM}}$ *is the directed graph given as follows. Its vertex-set $\mathcal{L}_S^{\mathrm{M}}$ consists of all terms containing each of the variables in $\{x_i \mid i \in S\}$ exactly once, and no other variables; while its edge-set $\mathrm{Bs}_S^{\mathrm{M}}$ consists of all basic transformations between such terms. That is, $\mathscr{B}_S^{\mathrm{SM}}$ is the full subgraph of $\mathscr{B}_k^{\mathrm{SM}}$ spanned by $\mathcal{L}_S^{\mathrm{M}}$.*

- $\mathrm{Bs}_S^{\mathrm{M}}$ *is the set of all α-, λ-, and ρ-basic transformations between terms in $\mathcal{L}_S^{\mathrm{M}}$.*

- $\langle R^{\mathrm{SM}}_k \rangle$ *is appropriately restricted to* $F(\mathscr{B}^{\mathrm{SM}}_S[(\mathrm{Bs}^{\mathrm{M}}_S)^{-1}])$.

Proof. We may partition the set $\mathcal{L}^{\mathrm{M,reg}}_k$ of regular terms of scope V_k as follows. Each term in $\mathcal{L}^{\mathrm{M,reg}}_k$ contains exactly the variables in a subset $\{x_i \mid i \in S\} \subseteq V_k$, where S could be any subset of $\underline{k}$. So we may write

$$\mathcal{L}^{\mathrm{M,reg}}_k = \coprod_{S \subseteq \underline{k}} \mathcal{L}^{\mathrm{M}}_S$$

with each $\mathcal{L}^{\mathrm{M}}_S$ defined as above. For our transformation schema, a variable is in the domain of a basic transformation if and only if it is in the codomain, so there are no edges in $\mathscr{B}^{\mathrm{SM}}_k$ between vertices in $\mathcal{L}^{\mathrm{M}}_S$ and $\mathcal{L}^{\mathrm{M}}_T$ for $S \neq T$. Hence the graph $\mathscr{B}^{\mathrm{SM}}_k$ can likewise be decomposed as

$$\mathscr{B}^{\mathrm{SM}}_k = \coprod_{S \subseteq \underline{k}} \mathscr{B}^{\mathrm{SM}}_S$$

with each $\mathscr{B}^{\mathrm{SM}}_S$ defined as above, and the result follows. $\square$

Notice that the structure to each CNP $(\mathscr{B}^{\mathrm{SM}}_S, \mathrm{Bs}^{\mathrm{M}}_S, \langle R^{\mathrm{SM}}_k \rangle)$ for $S \subseteq \underline{k}$ defined above depends only on the cardinality of S, say $r = |S|$. After all, any bijection $\iota \colon \underline{r} \xrightarrow{\sim} S$ extends to a bijection $\mathcal{L}^{\mathrm{M}}_{\underline{r}} \to \mathcal{L}^{\mathrm{M}}_S$ replacing each x_i in a term with $x_{\iota(i)}$. This bijection in turn extends to a graph isomorphism $\mathscr{B}^{\mathrm{SM}}_{\underline{r}} \cong \mathscr{B}^{\mathrm{SM}}_S$, which further extends to a CNP isomorphism $(\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_k \rangle) \xrightarrow{\sim} (\mathscr{B}^{\mathrm{SM}}_S, \mathrm{Bs}^{\mathrm{M}}_S, \langle R^{\mathrm{SM}}_k \rangle)$. We can therefore rephrase Proposition 3.75 as follows.

Proposition 3.76. *We have an isomorphism of CNPs*

$$\mathrm{CN}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}^{\mathrm{M,reg}}_k) \cong \coprod_{0 \le r \le k} \coprod_{i=1}^{\binom{k}{r}} (\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_k \rangle).$$

So it remains to solve the CNP $(\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_k \rangle)$ for $r \le k$. By Corollary 3.11, it suffices to solve the CNP $(\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_r \rangle)$, since $\langle R^{\mathrm{SM}}_r \rangle \subseteq \langle R^{\mathrm{SM}}_k \rangle$. To do so, we shall apply Proposition 3.69. First, we leverage our solution to the CNP given by $\mathrm{CN}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}^{\mathrm{M}}_r) = (\mathscr{B}^{\mathrm{M}}_r, \mathrm{Bs}^{\mathrm{M}}_r, \langle R^{\mathrm{M}}_r \rangle)$ from Theorem 3.65 (i.e. Mac Lane's coherence theorem for monoidal categories) to obtain a solution to the subproblem $(\mathscr{B}^{\mathrm{M}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_r \rangle)$ of $(\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_r \rangle)$, where $\mathscr{B}^{\mathrm{M}}_{\underline{r}}$ is the wide subgraph of $\mathscr{B}^{\mathrm{SM}}_{\underline{r}}$ with edge-set $\mathrm{Bs}^{\mathrm{M}}_{\underline{r}}$. We shall denote $\mathscr{F}^{\mathrm{M,reg}}_r := F(\mathscr{B}^{\mathrm{M,reg}}_r[(\mathrm{Bs}^{\mathrm{M,reg}}_r)^{-1}])$ and $\mathscr{F}^{\mathrm{M}}_{\underline{r}} := F(\mathscr{B}^{\mathrm{M}}_{\underline{r}}[(\mathrm{Bs}^{\mathrm{M}}_{\underline{r}})^{-1}])$.

Proposition 3.77. *Recall from Definition 3.26 the set of right-normalized terms $\mathcal{N}_r \subset \mathcal{M}_r$ in r variables with no units. Let $\mathcal{N}_r^{\mathrm{reg}} \coloneqq \mathcal{N}_r \cap \mathcal{L}_r^{\mathrm{M,reg}}$ denote the set of all* regular *right-normalized terms with no units, and let $\mathcal{N}_{\underline{r}} \coloneqq \mathcal{N}_r \cap \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ denote the set of all right-normalized terms with no units in which* each variable in $V_r = \{x_1 < \cdots < x_r\}$ appears exactly once *and* no other variables appear*.*

Then $\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}}$ is a solution to the CNP $(\mathscr{B}_r^{\mathrm{M,reg}}, \mathrm{Bs}_r^{\mathrm{M,reg}}, \langle R_r^{\mathrm{SM}} \rangle)$ (here $\langle R_r^{\mathrm{SM}} \rangle$ is restricted to $\mathscr{F}_r^{\mathrm{M,reg}} \subset \mathscr{F}_r^{\mathrm{SM}}$).

Meanwhile, a solution to the CNP $(\mathscr{B}_{\underline{r}}^{\mathrm{M}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}} \rangle)$ (here $\langle R_r^{\mathrm{SM}} \rangle$ is restricted to $\mathscr{F}_{\underline{r}}^{\mathrm{M}} \subset \mathscr{F}_r^{\mathrm{SM}}$) is either $\{e\}$ if $r = 0$ or $\mathcal{N}_{\underline{r}}$ if $r \geq 1$. So $\mathrm{Bs}_{\underline{r}}^{\mathrm{M}} \lhd (\mathscr{B}_{\underline{r}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}} \rangle)$.

Proof. By Theorem 3.65, the set $\{e\} \sqcup \mathcal{N}_r$ is a solution to the CNP given by $\mathrm{CN}(\mathfrak{C}_{\mathrm{M}}, \mathcal{L}_r^{\mathrm{M}}) = (\mathscr{B}_r^{\mathrm{M}}, \mathrm{Bs}_r^{\mathrm{M}}, \langle R_r^{\mathrm{M}} \rangle)$ for monoidal categories in r variables. The edge-set $\mathrm{Bs}_r^{\mathrm{M}}$ of the graph $\mathscr{B}_r^{\mathrm{M}}$ consists of α-, λ-, and ρ-basic transformations, and for each, every variable appears the same number of times in the domain as in the codomain. It follows that in the graph $\mathscr{B}_k^{\mathrm{M}}$, there are no edges between terms in $\mathcal{L}_r^{\mathrm{M,reg}}$ and not in $\mathcal{L}_r^{\mathrm{M,reg}}$, and there are no edges between terms in $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$ and not in $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$. Thus $\mathscr{B}_r^{\mathrm{M}}$ can be written as a coproduct in which one of the summands is the full subgraph $\mathscr{B}_r^{\mathrm{M,reg}}$ spanned by terms in $\mathcal{L}_r^{\mathrm{M,reg}}$, and $\mathscr{B}_r^{\mathrm{M}}$ can also be written as a coproduct in which one of the summands is the full subgraph $\mathscr{B}_{\underline{r}}^{\mathrm{M}}$ spanned by terms in $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$. So by Proposition 3.41, the subproblem $(\mathscr{B}_r^{\mathrm{M,reg}}, \mathrm{Bs}_r^{\mathrm{M,reg}}, \langle R_r^{\mathrm{M}} \rangle)$ has solution

$$(\{e\} \sqcup \mathcal{N}_r) \cap \mathcal{L}_r^{\mathrm{M,reg}} = \{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}},$$

as the atomic term e is itself in $\mathcal{L}_r^{\mathrm{M,reg}}$, while the subproblem $(\mathscr{B}_{\underline{r}}^{\mathrm{M}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{M}} \rangle)$ has solution

$$(\{e\} \sqcup \mathcal{N}_r) \cap \mathcal{L}_{\underline{r}}^{\mathrm{M}} = \begin{cases} \{e\} & \text{if } r = 0 \\ \mathcal{N}_{\underline{r}} & \text{otherwise,} \end{cases}$$

as $e \in \mathcal{L}_{\underline{0}}^{\mathrm{M}}$ but no terms in $\mathcal{L}_{\underline{0}}^{\mathrm{M}}$ have variables, while $e \notin \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ for $r > 0$. As shorthand, we denote this solution set by $\overline{\mathcal{N}_{\underline{r}}}$ for all r.

Finally, since the two $\mathfrak{R}^{\mathrm{M}}$-relations defined on $\mathscr{F}^{\mathrm{M}}$ are contained in the restriction of $\mathfrak{R}^{\mathrm{SM}}$ to $\mathscr{F}^{\mathrm{M}} \subset \mathscr{F}^{\mathrm{SM}}$, the local binary relation R_r^{M} defined on $\mathscr{F}_r^{\mathrm{M}}$ is contained in the restriction of R_r^{SM} to $\mathscr{F}_r^{\mathrm{M}}$, so the restriction of $\langle R_r^{\mathrm{M}} \rangle$ to $\mathscr{F}_r^{\mathrm{M,reg}}$ or to $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ is in turn contained in the restriction of $\langle R_r^{\mathrm{SM}} \rangle$ to the same. It follows from Corollary 3.11 that $\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}}$ is a solution to $(\mathscr{B}_r^{\mathrm{M,reg}}, \mathrm{Bs}_r^{\mathrm{M,reg}}, \langle R_r^{\mathrm{SM}} \rangle)$ and $\overline{\mathcal{N}_{\underline{r}}}$ is a solution to $(\mathscr{B}_{\underline{r}}^{\mathrm{M}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}} \rangle)$. $\square$

We may therefore construct the quotient CNP denoted $(\mathscr{B}_{\underline{r}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}} \rangle)/\mathrm{Bs}_{\underline{r}}^{\mathrm{M}}$, following Definition 3.66, effectively quotienting out all the α-, λ-, and ρ-basic transformations and leaving only equivalence classes of the β-basic transformations. Explicitly, we may define the quotient $(\mathscr{B}_{\underline{r}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}} \rangle)/\mathrm{Bs}_{\underline{r}}^{\mathrm{M}} = (\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}, \varnothing, \approx)$ as follows, writing $\mathscr{F}_{\underline{r}}^{\mathrm{SM}} := F(\mathscr{B}_{\underline{r}}^{\mathrm{SM}}[(\mathrm{Bs}_{\underline{r}}^{\mathrm{M}})^{-1}])$.

(1) The graph $\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}$ is given by the following.

- Its vertex-set is $\mathcal{L}_{\underline{r}}^{\mathrm{M}}/\Leftrightarrow$, with $v \Leftrightarrow v'$ if and only if there exists a morphism $v \Rightarrow v'$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$. Note that $\mathcal{L}_{\underline{r}}^{\mathrm{M}}/\Leftrightarrow$ is in bijection with the solution set $\{e\} \sqcup \mathcal{N}_{\underline{r}}$. We denote the $\Leftrightarrow$-equivalence class of $t \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ by $[t]$.
- Set $\mathrm{Bs}_{\underline{r}}^{\beta} := \mathrm{Bs}_{\underline{r}}^{\mathrm{SM}} \setminus \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}$, so that it consists of exactly the β-basic transformations between terms in $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$. Then our edge-set is $\mathrm{Bs}_{\underline{r}}^{\beta}/\equiv$, with $\varphi \equiv \varphi'$ for β-basic transformations $\varphi \colon u \Rightarrow v$ and $\varphi' \colon u' \Rightarrow v'$ between regular terms $u, v, u', v' \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ whenever there are morphisms $\kappa \colon u \Rightarrow u'$ and $\nu \colon v \Rightarrow v'$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ such that $(\nu \bullet \varphi)\ \langle R_r^{\mathrm{SM}} \rangle\ (\varphi' \bullet \kappa)$ in $\mathscr{F}_{\underline{r}}^{\mathrm{SM}}$: the square

$$\begin{tikzcd} u \arrow[r, Rightarrow, "\kappa"] \arrow[d, Rightarrow, "\varphi"'] & u' \arrow[d, Rightarrow, "\varphi'"] \\ v \arrow[r, Rightarrow, "\nu"] & v' \end{tikzcd}$$

commutes up to $\langle R_r^{\mathrm{SM}} \rangle$.

Given β-basic transformation $\varphi \colon u \Rightarrow v$ in $\mathrm{Bs}_{\underline{r}}^{\beta}$, we denote the $\equiv$-equivalence class of φ in $\mathrm{Bs}_{\underline{r}}^{\beta}/\equiv$ by $[\varphi]$, with domain $[u]$ and codomain $[v]$ in our graph $\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}$.

(2) Since we are quotienting out by $\mathrm{Bs}_{\underline{r}}^{\mathrm{M}}$, the set of all marked edges of our original CNP, no marked edges remain.

(3) Now define a graph morphism $[\cdot] \colon \mathscr{B}_{\underline{r}}^{\mathrm{SM}}[(\mathrm{Bs}_{\underline{r}}^{\mathrm{M}})^{-1}] \to F(\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}})$ by sending $t \mapsto [t]$ on vertices; $\varphi \mapsto [\varphi]$ on edges $\varphi \in \mathrm{Bs}_{\underline{r}}^{\beta}$; and $\psi^{\pm 1} \mapsto \mathrm{id}_{[\mathrm{dom}(\psi)]}$ on edges $\psi \in \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}$. Freely

extend $[\cdot]$ to a functor $[\cdot]\colon \mathscr{F}_{\underline{r}}^{\mathrm{SM}} \to F(\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}})$. Then for each pair of $\langle R_r^{\mathrm{SM}}\rangle$-related parallel morphisms ν, ν' in $\mathscr{F}_{\underline{r}}^{\mathrm{SM}}$, set $[\nu] \approx [\nu']$.

To understand this quotient, we need to better understand the solution-set $\{e\} \sqcup \mathcal{N}_{\underline{r}}$ of our CNP $(\mathscr{B}_{\underline{r}}^{\mathrm{M}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}}\rangle)$. To that end, recall our left Σ_r-set

$$\mathcal{S}_r = \{\sigma^* \mid \sigma \in \Sigma_r\}$$

with

$$\sigma^* = x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(r)};$$

we set $\mathcal{S}_0 \coloneqq \{e\}$. The set $\mathcal{S}_r$ is contained in a larger set

$$\mathcal{P}_r = \{e\} \sqcup \{x_{\iota(1)} \cdots x_{\iota(m)} \mid 1 \leq m \leq r,\ \iota\colon \underline{m} \to \underline{r} \text{ injective}\}.$$

We have a function $[\![\cdot]\!]\colon \mathcal{L}_r^{\mathrm{M,reg}} \to \mathcal{P}_r$ given by removing all parentheses, commas, function symbols, and constant symbols from the terms in $\mathcal{L}_r^{\mathrm{M,reg}}$, leaving only the variables in the order in which they appear—except that we send terms with no variables to e. Then the image of $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$ under $[\![\cdot]\!]$ is precisely $\mathcal{S}_r$, so we may restrict $[\![\cdot]\!]$ to a function $[\![\cdot]\!]\colon \mathcal{L}_{\underline{r}}^{\mathrm{M}} \to \mathcal{S}_r$.

Note that $[\![\cdot]\!]$ agrees on the domain and codomain of every α-, λ-, and ρ-basic transformation between terms in $\mathcal{L}_r^{\mathrm{M,reg}}$. This yields an alternate formulation of the solution-sets $\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}}$ to $(\mathscr{B}_r^{\mathrm{M,reg}}, \mathrm{Bs}_r^{\mathrm{M,reg}}, \langle R_r^{\mathrm{SM}}\rangle)$ and $\{e\} \sqcup \mathcal{N}_{\underline{r}}$ to $(\mathscr{B}_{\underline{r}}^{\mathrm{M}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}}\rangle)$. Recall that each term $t \in \mathcal{L}_r^{\mathrm{M,reg}}$ must have a unique $(\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}})$-normal form $n(t) \in \{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}}$ for which there exists a morphism $t \Rightarrow n(t)$ in $\mathscr{F}_r^{\mathrm{M,reg}} \subset \mathscr{F}_r^{\mathrm{M}}$, yielding a function given by $n\colon \mathcal{L}_r^{\mathrm{M,reg}} \to (\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}})$ that is the identity when its domain is restricted to $(\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}})$. Furthermore, if $t \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$, then $n(t) \in \overline{\mathcal{N}_{\underline{r}}}$ and the morphism $t \Rightarrow n(t)$ is in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$; so we may restrict n to a function $\mathcal{L}_{\underline{r}}^{\mathrm{M}} \to \overline{\mathcal{N}_{\underline{r}}}$ that is the identity when its domain is restricted to $\overline{\mathcal{N}_{\underline{r}}}$. The fact that $[\![\cdot]\!]$ agrees on the domain and codomain of every edge in $\mathscr{B}_r^{\mathrm{M,reg}}$ implies that $[\![t]\!] = [\![n(t)]\!]$ for every $t \in \mathcal{L}_r^{\mathrm{M,reg}}$. In fact, we can say something stronger.

Proposition 3.78. *Restricting $[\![\cdot]\!]\colon \mathcal{L}_r^{\mathrm{M,reg}} \to \mathcal{P}_r$ to the domain $(\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}}) \subset \mathcal{L}_r^{\mathrm{M,reg}}$ yields a bijection $[\![\cdot]\!]\colon (\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}}) \xrightarrow{\sim} \mathcal{P}_r$ so that $[\![\cdot]\!] = [\![\cdot]\!] \circ n$:*

$$\begin{tikzcd} & \mathcal{L}_r^{\mathrm{reg}} \arrow[dl, "n"'] \arrow[dr, "{[\![\cdot]\!]}"] & \\ \{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}} \arrow[rr, "{[\![\cdot]\!]}"] & & \mathcal{P}_r \end{tikzcd}$$

commutes. Restricting $[\![\cdot]\!]$ *further to the domain* $\overline{\mathcal{N}_{\underline{r}}}$ *yields a bijection given by* $[\![\cdot]\!]\colon \overline{\mathcal{N}_{\underline{r}}} \xrightarrow{\sim} \mathcal{S}_r$*.*

Proof. We have already seen that the given diagram commutes: $[\![t]\!] = [\![n(t)]\!]$ for $t \in \mathcal{L}_r^{\mathrm{M,reg}}$. It remains to show that $[\![\cdot]\!]\colon (\{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}}) \to \mathcal{P}_r$ is bijective and that the image of $\{e\} \sqcup \mathcal{N}_{\underline{r}}$ under $[\![\cdot]\!]$ is precisely $\mathcal{S}_r$. Note that $[\![e]\!] = e \in \mathcal{S}_r$; while $t \in \{e\} \sqcup \mathcal{N}_r^{\mathrm{reg}}$ satisfying $[\![t]\!] = e$ implies that t has no variables, so $t \notin \mathcal{N}_r^{\mathrm{reg}}$ and thus $t = e$. So it suffices to show that $[\![\cdot]\!]\colon \mathcal{N}_r^{\mathrm{reg}} \to \mathcal{P}_r \setminus \{e\}$ is bijective and that the image of $\overline{\mathcal{N}_{\underline{r}}}$ under $[\![\cdot]\!]$ is $\mathcal{S}_r$.

Given $t \in \mathcal{N}_r^{\mathrm{reg}}$, the value of $[\![t]\!]$ tells us the order in which the variables of t appear. Then since (2) implies (4) in Proposition 3.27, given the order of variables appearing in the right-normalized term $t \in \mathcal{N}_r^{\mathrm{reg}}$, we may uniquely recover t. It follows that $[\![\cdot]\!]\colon \mathcal{N}_r^{\mathrm{reg}} \to \mathcal{P}_r \setminus \{e\}$ is injective.

Conversely, for all injections $\iota\colon \underline{m} \to \underline{r}$ for $1 \le m \le r$, we construct a term $t \in \mathcal{L}_r^{\mathrm{M,reg}}$ by

$$t = \otimes(x_{\iota(1)}, \otimes(x_{\iota(2)}, \cdots, \otimes(x_{\iota(m-1)}, x_{\iota(m)}) \cdots))$$

that is right-normalized because (4) implies (2) in Proposition 3.27. So $t \in \mathcal{N}_r^{\mathrm{reg}}$ and

$$[\![t]\!] = [\otimes(x_{\iota(1)}, \otimes(x_{\iota(2)}, \cdots, \otimes(x_{\iota(m-1)}, x_{\iota(m)}) \cdots))] = x_{\iota(1)} x_{\iota(2)} \cdots x_{\iota(m)}.$$

Hence $[\![\cdot]\!]\colon \mathcal{N}_r^{\mathrm{reg}} \to \mathcal{P}_r \setminus \{e\}$ is surjective.

For $r = 0$, the image of $\overline{\mathcal{N}_{\underline{r}}} = \{e\}$ under $[\![\cdot]\!]$ is $\mathcal{S}_0 = \{e\}$. Otherwise, if $r > 0$, we have that $t \in \overline{\mathcal{N}_{\underline{r}}} = \mathcal{N}_{\underline{r}}$, i.e.

$$t = \otimes(x_{\sigma^{-1}(1)}, \otimes(x_{\sigma^{-1}(2)}, \cdots, \otimes(x_{\sigma^{-1}(r-1)}, x_{\sigma^{-1}(r)}) \cdots))$$

for some $\sigma \in \Sigma_r$, if and only if we can write $[\![t]\!]$ in the form

$$[\![t]\!] = [\otimes(x_{\sigma^{-1}(1)}, \otimes(x_{\sigma^{-1}(2)}, \cdots, \otimes(x_{\sigma^{-1}(r-1)}, x_{\sigma^{-1}(r)}) \cdots))] = x_{\sigma^{-1}(1)} x_{\sigma^{-1}(2)} \cdots x_{\sigma^{-1}(r)},$$

i.e. $[\![t]\!] \in \mathcal{S}_r$. So we have a bijection $[\![\cdot]\!]\colon \overline{\mathcal{N}_{\underline{r}}} \xrightarrow{\sim} \mathcal{S}_r$. $\square$

Corollary 3.79. *For* $t, t' \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$*, we have* $[t] = [t']$ *in* $\mathcal{L}_{\underline{r}}^{\mathrm{M}}/\!\Leftrightarrow$ *if and only if* $[\![t]\!] = [\![t']\!]$ *in* $\mathcal{S}_r$*. As* $\mathcal{L}_{\underline{r}}^{\mathrm{M}}/\!\Leftrightarrow$ *is also in bijection with the solution set* $\overline{\mathcal{N}_{\underline{r}}}$*, it follows that* $\mathcal{L}_{\underline{r}}^{\mathrm{M}}/\!\Leftrightarrow$ *is in bijection with* $\mathcal{S}_r$ *via* $[t] \mapsto [\![t]\!]$*.*

Proof. Given $t, t' \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ for which $[t] = [t']$, there exists a morphism $t \Rightarrow t'$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$, so by Lemma 3.2 we have $n(t) = n(t')$ and thus $[\![t]\!] = [\![n(t)]\!] = [\![n(t')]\!] = [\![t']\!]$.

Conversely, given $t, t' \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}/\Leftrightarrow$ for which $[\![t]\!] = [\![t']\!]$, we have $[\![n(t)]\!] = [\![t]\!] = [\![t']\!] = [\![n(t')]\!]$, so $n(t) = n(t')$. So there exist $\nu\colon t \Rightarrow n(t)$ and $\nu'\colon t' \Rightarrow n(t') = n(t)$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$. Formally inverting every edge in ν' yields a third morphism $\overline{\nu}\colon n(t) \Rightarrow t'$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$, so there is a morphism $(\overline{\nu} \bullet \nu)\colon t \Rightarrow t'$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ witnessing $[t] = [t']$.

It follows that the function $[\![\cdot]\!]\colon \mathcal{L}_{\underline{r}}^{\mathrm{M}} \to \mathcal{S}_r$ descends to a function $(\mathcal{L}_{\underline{r}}^{\mathrm{M}}/\Leftrightarrow) \to \mathcal{S}_r$ sending $[t] \mapsto [\![t]\!]$ for $t \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$. This function is a bijection: it is a composite of the bijection $(\mathcal{L}_{\underline{r}}^{\mathrm{M}}) \xrightarrow{\sim} \overline{\mathcal{N}_{\underline{r}}}$ sending $[t] \mapsto n(t)$ and the bijection $\overline{\mathcal{N}_{\underline{r}}} \xrightarrow{\sim} \mathcal{S}_r$ sending $n \in \mathcal{N}_{\underline{r}}$ to $[\![n]\!]$, as $[\![t]\!] = [\![n(t)]\!]$. ☐

We now extend the function $[\![\cdot]\!]$ on the set $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$ to the set of β-basic transformations $\mathrm{Bs}_{\underline{r}}^{\beta}$. That is, given β-basic transformation $\varphi\colon u \Rightarrow v$ in $\mathrm{Bs}_{\underline{r}}^{\beta}$, we know the $(S_1(\varphi), S_2(\varphi))$-subterm of u can be written as $\otimes(w, z)$ for some $w, z \in \mathcal{L}_r^{\mathrm{M,reg}}$, while the $(S_1(\varphi), S_2(\varphi))$-subterm of v can be written as $\otimes(z, w)$ (so the atom of φ is $A(\varphi) = \beta_{w,z}$); meanwhile, $\mathrm{len}(u) = \mathrm{len}(v)$, and the $(1, S_1(\varphi) - 1)$- and $(S_2(\varphi) + 1, \mathrm{len}(u))$-subterms of u and v coincide. Then we have two cases. If neither w nor z is equal to the constant symbol e, there exist $\sigma, \rho \in \Sigma_r$ for which

$$[\![u]\!] = \sigma^* \coloneqq x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(i)} [\![w]\!] [\![z]\!] x_{\sigma^{-1}(k+1)} \cdots x_{\sigma^{-1}(r)}$$
$$[\![v]\!] = \rho^* \coloneqq x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(i)} [\![z]\!] [\![w]\!] x_{\sigma^{-1}(k+1)} \cdots x_{\sigma^{-1}(r)}$$

for $[\![w]\!] = x_{\sigma^{-1}(i+1)} \cdots x_{\sigma^{-1}(j)}$, $[\![z]\!] = x_{\sigma^{-1}(j+1)} \cdots x_{\sigma^{-1}(k)}$, and $0 \le i < j < k \le r$. It follows that the block transposition $\tau[i,j,k] \in B_r$ satisfies $\tau[i,j,k]\sigma = \rho$. We define $[\![\varphi]\!] \coloneqq \tau[i,j,k]$. On the other hand, if either w or z is equal to e, then $[\![u]\!] = [\![v]\!]$, and we define $[\![\varphi]\!] \coloneqq \mathrm{id} \in \Sigma_r$. Putting this together, we have defined a function $[\![\cdot]\!]\colon \mathrm{Bs}_{\underline{r}}^{\beta} \to \{\mathrm{id}\} \sqcup B_r \subset \Sigma_r$ such that if $[\![\mathrm{dom}(\varphi)]\!] = \sigma^*$ and $[\![\mathrm{cod}(\varphi)]\!] = \rho^*$ for $\sigma, \rho \in \Sigma_r$, we have $[\![\varphi]\!]\sigma = \rho$.

Now consider the CNP $\mathfrak{G}_{\Sigma_r} = (\mathscr{S}_r, \varnothing, \langle R_{\Sigma_r} \rangle)$ we defined in Proposition 3.32, whose graph $\mathscr{S}_r$ has vertex set $\mathcal{S}_r$ and edges

$$\tau[i,j,k]\colon \sigma^* \Rightarrow \tau[i,j,k]\sigma^*$$

for $\sigma \in \Sigma_r$ and $0 \leq i < j < k \leq r$ as well as

$$\mathrm{id}\colon \sigma^* \Rightarrow \sigma^*$$

for $\sigma \in \Sigma_r$. If we define $\mathscr{B}_{\underline{r}}^{\beta}$ to be the (wide) subgraph of $\mathscr{B}_{\underline{r}}^{\mathrm{SM}}$ with vertex-set $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$ and edge-set $\mathrm{Bs}_{\underline{r}}^{\beta}$, we then have a graph morphism $[\![\cdot]\!]\colon \mathscr{B}_{\underline{r}}^{\beta} \to \mathscr{S}_r$ sending vertex $v \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ to $[\![v]\!] \in \mathcal{S}_r$ and edge $\varphi\colon u \Rightarrow v$ in $\mathrm{Bs}_{\underline{r}}^{\beta}$ to edge $[\![\varphi]\!]\colon [\![u]\!] \Rightarrow [\![v]\!]$ of $\mathscr{S}_r$.

We also have our graph morphism $[\cdot]\colon \mathscr{B}_{\underline{r}}^{\beta} \to \overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}$ from defining the quotient CNP $(\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}, \varnothing, \approx)$. We claim that the graph morphism $[\![\cdot]\!]$ descends via $[\cdot]$ to a graph morphism $\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}} \to \mathscr{S}_r$.

Proposition 3.80. *There is a graph morphism $\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}} \to \mathscr{S}_r$ sending $[\varphi]\colon [u] \Rightarrow [v]$ for $\varphi\colon u \Rightarrow v$ in $\mathscr{B}_{\underline{r}}^{\beta}$ to $[\![\varphi]\!]\colon [\![u]\!] \to [\![v]\!]$.*

Proof. We have already demonstrated a bijection on vertices $(\mathcal{L}_{\underline{r}}^{\mathrm{M}}/\Leftrightarrow) \xrightarrow{\sim} \mathcal{S}_r$ that is given by $[t] \mapsto [\![t]\!]$ for $t \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$, and we have a graph morphism $[\![\cdot]\!]\colon \mathscr{B}_{\underline{r}}^{\beta} \to \mathscr{S}_r$. To show that the latter descends to a well-defined graph morphism $\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}} \to \mathscr{S}_r$, we need to show for β-basic transformations $\varphi\colon u \Rightarrow v$ and $\varphi'\colon u' \Rightarrow v'$ with $u, v, u', v' \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ for which there exist morphisms $\kappa\colon u \Rightarrow u'$ and $\nu\colon v \Rightarrow v'$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ such that the square

$$\begin{tikzcd}
u \arrow[r, Rightarrow, "\kappa"] \arrow[d, Rightarrow, "\varphi"'] & u' \arrow[d, Rightarrow, "\varphi'"] \\
v \arrow[r, Rightarrow, "\nu"] & v'
\end{tikzcd}$$

commutes up to $\langle R_r^{\mathrm{SM}} \rangle$, we have $[\![\varphi]\!] = [\![\varphi']\!]$. We already know $[u] = [u']$ and $[v] = [v']$, so $[\![u]\!] = [\![u']\!]$ and $[\![v]\!] = [\![v']\!]$. But the edge $[\![\varphi]\!]\colon [\![u]\!] \Rightarrow [\![v]\!]$ in $\mathscr{S}_r$ as we have defined is in fact entirely determined by its domain $[\![u]\!]$ and codomain $[\![v]\!]$, so we necessarily have $[\![\varphi]\!] = [\![\varphi']\!]$. $\square$

Moreover, we show that the graph morphism exhibited above is in fact an isomorphism.

Proposition 3.81. *The graph morphism $\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}} \to \mathscr{S}_r$ sending $[\varphi]\colon [u] \Rightarrow [v]$ for $\varphi\colon u \Rightarrow v$ in $\mathscr{B}_{\underline{r}}^{\beta}$ to $[\![\varphi]\!]\colon [\![u]\!] \to [\![v]\!]$ is a graph isomorphism.*

Proof. Again, we have already showed that this graph morphism is isomorphic on vertices, so it remains to show that it is isomorphic on edges. For surjectivity, fix $\sigma, \rho \in \Sigma_r$ and an edge $\sigma^* \Rightarrow \rho^*$ in $\mathscr{S}_r$. If $\sigma = \rho$, this edge is $\mathrm{id}\colon \sigma^* \Rightarrow \sigma^*$. Take $u \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ with $\llbracket u \rrbracket = \sigma^*$. Then $\otimes(u, e)$ and $\otimes(e, u)$ are also in $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$, with $\llbracket \otimes(u, e) \rrbracket = \sigma^* = \llbracket \otimes(e, u) \rrbracket$, and there is an edge $\beta_{u,e}\colon \otimes(u, e) \Rightarrow \otimes(e, u)$ of $\mathscr{B}_{\underline{r}}^{\beta}$ with $\llbracket \beta_{u,e} \rrbracket = \mathrm{id}\colon \sigma^* \Rightarrow \sigma^*$. Otherwise, our target edge is $\tau[i, j, k]\colon \sigma^* \Rightarrow \tau[i, j, k]\sigma^*$ for $0 \leq i < j < k \leq r$. Writing

$$w \coloneqq \begin{cases} e & \text{if } i = 0 \\ \otimes(x_{\sigma^{-1}(1)}, \otimes(x_{\sigma^{-1}(2)}, \cdots \otimes(x_{\sigma^{-1}(i-1)}, x_{\sigma^{-1}(i)}) \cdots)) & \text{otherwise,} \end{cases}$$

$$\begin{aligned} s &\coloneqq \otimes(x_{\sigma^{-1}(i+1)}, \otimes(x_{\sigma^{-1}(i+2)}, \cdots \otimes(x_{\sigma^{-1}(j-1)}, x_{\sigma^{-1}(j)}) \cdots)), \\ t &\coloneqq \otimes(x_{\sigma^{-1}(j+1)}, \otimes(x_{\sigma^{-1}(j+2)}, \cdots \otimes(x_{\sigma^{-1}(k-1)}, x_{\sigma^{-1}(k)}) \cdots)), \end{aligned}$$

$$z \coloneqq \begin{cases} e & \text{if } k = r \\ \otimes(x_{\sigma^{-1}(k+1)}, \otimes(x_{\sigma^{-1}(k+2)}, \cdots \otimes(x_{\sigma^{-1}(r-1)}, x_{\sigma^{-1}(r)}) \cdots)) & \text{otherwise,} \end{cases}$$

we have

$$\begin{aligned} & \llbracket \otimes(\otimes(w, \otimes(s, t)), z) \rrbracket = \llbracket w \rrbracket \llbracket s \rrbracket \llbracket t \rrbracket \llbracket z \rrbracket = \sigma^* \\ \text{and} \quad & \llbracket \otimes(\otimes(w, \otimes(t, s)), z) \rrbracket = \llbracket w \rrbracket \llbracket t \rrbracket \llbracket s \rrbracket \llbracket z \rrbracket = \tau[i, j, k]\sigma^*, \end{aligned}$$

so $\otimes(\otimes(\mathrm{id}_w, \beta_{s,t}), \mathrm{id}_z)$ is an edge in $\mathrm{Bs}_{\underline{r}}^{\beta}$ that $\llbracket \cdot \rrbracket$ sends to $\tau[i, j, k]\colon \sigma^* \Rightarrow \tau[i, j, k]\sigma^*$.

For injectivity, fix β-basic transformations $\varphi\colon u \Rightarrow v$ and $\varphi'\colon u' \Rightarrow v'$ for some terms $u, v, u', v' \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$ with $[u] = [u']$ and $[v] = [v']$ such that $\llbracket \varphi \rrbracket = \llbracket \varphi' \rrbracket$. We want to show that $[\varphi] = [\varphi']$; that is, there exist morphisms $\kappa\colon u \Rightarrow u'$ and $\nu\colon v \Rightarrow v'$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ for which $(\nu \bullet \varphi) \langle R_r^{\mathrm{SM}} \rangle (\varphi' \bullet \kappa)$. We have two cases: either we have $\llbracket \varphi \rrbracket = \mathrm{id}$ or $\llbracket \varphi \rrbracket = \tau[i, j, k]$ for $0 \leq i < j < k \leq r$. We rely on Proposition 3.77, in which we showed that $(\mathscr{B}_{\underline{r}}^{\mathrm{M}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}} \rangle)$ has a solution, so any two parallel morphisms in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ are $\langle R_r^{\mathrm{SM}} \rangle$-related.

If $\llbracket \varphi \rrbracket = \mathrm{id} = \llbracket \varphi' \rrbracket$, then $[u'] = [u] = [v] = [v']$, so there exists a square of morphisms

$$\begin{tikzcd} u \arrow[r, Rightarrow, "\kappa"] \arrow[d, Rightarrow, "\psi"'] & u' \arrow[d, Rightarrow, "\psi'"] \\ v \arrow[r, Rightarrow, "\nu"] & v' \end{tikzcd}$$

in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ that commutes up to $\langle R_r^{\mathrm{SM}}\rangle$. Moreover $A(\varphi)$ is either $\beta_{s,e}$ or $\beta_{e,s}$ for some $s \in \mathcal{L}_r^{\mathrm{M}}$ and $A(\varphi')$ is either $\beta_{t,e}$ or $\beta_{e,t}$ for some $t \in \mathcal{L}_r^{\mathrm{M}}$. By the $\mathfrak{R}^{\mathrm{SM}}$-relations (3.73) and (3.72),

$$\begin{tikzcd} \otimes(s,e) \arrow[rr, Rightarrow, "\beta_{s,e}"] \arrow[dr, Rightarrow, "\rho"'] & & \otimes(e,s) \arrow[dl, Rightarrow, "\lambda"] \\ & s & \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} \otimes(e,s) \arrow[rr, Rightarrow, "\beta_{e,s}"] \arrow[dr, Rightarrow, "\lambda"'] & & \otimes(s,e) \arrow[dl, Rightarrow, "\rho"] \\ & s & \end{tikzcd}$$

are commutative up to $\langle R_r^{\mathrm{SM}}\rangle$, so $\beta_{s,e}$ and $\beta_{e,s}$ are $\langle R_r^{\mathrm{SM}}\rangle$-related to morphisms in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ (generated by a λ-basic edge and a ρ-basic edge); and the same with t in place of s. Then Lemma 2.49 implies that $\varphi\colon u \Rightarrow v$ and $\varphi'\colon u' \Rightarrow v'$ are also $\langle R_r^{\mathrm{SM}}\rangle$-related to morphisms in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$. So they are necessarily $\langle R_r^{\mathrm{SM}}\rangle$-related to the morphisms ψ and ψ' in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$, and the result follows: $[\varphi] = [\varphi']$.

Otherwise, $[\![\varphi]\!] = \tau[i,j,k] = [\![\varphi']\!]$ for $0 \le i < j < k \le r$. Here we will construct morphisms $\kappa\colon u \Rightarrow u'$ and $\nu\colon v \Rightarrow v'$ in $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$ for which $(\nu \bullet \varphi)\ \langle R_r^{\mathrm{SM}}\rangle\ (\varphi' \bullet \kappa)$ in stages. Taking $\sigma \in \Sigma_r$ for which $[\![u]\!] = \sigma^* = [\![u']\!]$, we have $[\![v]\!] = \tau[i,j,k]\sigma^* = [\![v']\!]$. We have $A(\varphi) = \beta_{s,t}$ and $A(\varphi') = \beta_{s',t'}$ for nonconstant terms $s, t, s', t' \in \mathcal{L}_r^{\mathrm{M,reg}}$ for which

$$[\![s]\!] = x_{\sigma^{-1}(i+1)} \cdots x_{\sigma^{-1}(j)} = [\![s']\!]$$
$$[\![t]\!] = x_{\sigma^{-1}(j+1)} \cdots x_{\sigma^{-1}(k)} = [\![t']\!].$$

So $n(s) = n(s')$ and $n(t) = n(t')$ in $\mathcal{N}_r^{\mathrm{reg}}$, yielding morphisms $\nu_s\colon s \Rightarrow s'$ and $\nu_t\colon t \Rightarrow t'$ in $\mathscr{F}^{\mathrm{M,reg}}$. Then κ and ν will each be the composite of three morphisms: informally, one that does not "affect" s or t, one that "changes" s to s' using ν_s, and one that "changes" t to t' using ν_t. We formalize this as follows.

Setting $\ell \coloneqq \sigma^{-1}(k)$, we may form terms $y, y' \in \mathcal{L}_r^{\mathrm{M,reg}}$ each containing exactly one instance of x_ℓ and no instances of the variables $x_{\sigma^{-1}(i+1)}, \ldots, x_{\sigma^{-1}(k-1)}$ with

$$[\![y]\!] = x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(i)}\, x_\ell\, x_{\sigma^{-1}(k+1)} \cdots x_{\sigma^{-1}(r)} = [\![y']\!]$$

as well as

$$\begin{aligned} y[x_\ell/\otimes(s,t)] &= u, \\ y[x_\ell/\otimes(t,s)] &= v, \\ y'[x_\ell/\otimes(s',t')] &= u', \\ \text{and} \quad y'[x_\ell/\otimes(t',s')] &= v', \end{aligned}$$

where we write $y[x_\ell/z]$ (resp. $y'[x_\ell/z]$) to denote the substitution of a term $z \in \mathcal{L}_r^{\mathrm{M,reg}}$ for x_ℓ in y (resp. y') while keeping every other variable the same. Then $n(y) = n(y')$ in $\mathcal{N}_r^{\mathrm{reg}}$, so there exist edges $\kappa_1 \colon y_0 \Rightarrow y_1, \ldots, \kappa_m \colon y_{m-1} \Rightarrow y_m$ of $\mathscr{B}_r^{\mathrm{M}}$ with $y_0 = y$, $y_m = y'$, and each $y_i \in \mathcal{L}_r^{\mathrm{M,reg}}$ with $[\![y_i]\!] = x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(i)} x_\ell x_{\sigma^{-1}(k+1)} \cdots x_{\sigma^{-1}(r)} = [\![y]\!]$, such that

$$\kappa' \coloneqq (\kappa_m^{\pm 1} \bullet \cdots \bullet \kappa_1^{\pm 1})$$

forms a path $y \Rightarrow y'$, and therefore

$$\kappa'[x_\ell/z] \coloneqq (\kappa_m^{\pm 1}[x_\ell/z] \bullet \cdots \bullet \kappa_1^{\pm 1}[x_\ell/z])$$

forms a path $y[x_\ell/z] \Rightarrow y'[x_\ell/z]$. In particular, we may form a square

$$\begin{tikzcd}
u = y[x_\ell/\otimes(s,t)] \arrow[r, Rightarrow, "{\kappa'[x_\ell/\otimes(s,t)]}"] \arrow[d, Rightarrow, "\varphi"'] & y'[x_\ell/\otimes(s,t)] \arrow[d, Rightarrow] \\
v = y[x_\ell/\otimes(t,s)] \arrow[r, Rightarrow, "{\kappa'[x_\ell/\otimes(t,s)]}"] & y'[x_\ell/\otimes(t,s)]
\end{tikzcd}$$

comprised of squares

$$\begin{tikzcd}
y_{i-1}[x_\ell/\otimes(s,t)] \arrow[r, Rightarrow, "{\kappa_i[x_\ell/\otimes(s,t)]^{\pm 1}}"] \arrow[d, Rightarrow] & y_i[x_\ell/\otimes(s,t)] \arrow[d, Rightarrow] \\
y_{i-1}[x_\ell/\otimes(t,s)] \arrow[r, Rightarrow, "{\kappa_i[x_\ell/\otimes(t,s)]^{\pm 1}}"] & y_i[x_\ell/\otimes(t,s)]
\end{tikzcd}$$

in which every object is a term in $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$, every vertical arrow is a β-basic transformation in $\mathrm{Bs}_{\underline{r}}^{\beta}$, and every horizontal arrow is in $(\mathrm{Bs}_{\underline{r}}^{\mathrm{M}})^{\pm 1}$. These commute up to $\langle R_r^{\mathrm{SM}} \rangle$ because either $\mathrm{dom}(A(\kappa_i))$ does not contain x_ℓ, in which case the square commutes up to $\langle R_r^{\mathrm{SM}} \rangle$ due to R^{fun} and possibly R^{inv}; or $\mathrm{dom}(A(\kappa_i))$ contains x_ℓ, in which case the square above commutes up to $\langle R_r^{\mathrm{SM}} \rangle$ due to R^{nat} and possibly R^{inv}. It remains to show that we may fill in the horizontal arrows in the diagram below with morphisms from $\mathscr{F}_{\underline{r}}^{\mathrm{M}}$:

$$\begin{tikzcd}
y'[x_\ell/\otimes(s,t)] \arrow[r, Rightarrow] \arrow[d, Rightarrow] & y'[x_\ell/\otimes(s',t)] \arrow[r, Rightarrow] \arrow[d, Rightarrow] & y'[x_\ell/\otimes(s',t')] = u' \arrow[d, Rightarrow, "\varphi'"] \\
y'[x_\ell/\otimes(t,s)] \arrow[r, Rightarrow] & y'[x_\ell/\otimes(t,s')] \arrow[r, Rightarrow] & y'[x_\ell/\otimes(t',s')] = v'.
\end{tikzcd}$$

Here every object is a term in $\mathcal{L}_{\underline{r}}^{\mathrm{M}}$ and every vertical arrow is a β-basic transformation in $\mathrm{Bs}_{\underline{r}}^{\beta}$. But the left pair of horizontal arrows may be constructed from $\nu_s \colon s \Rightarrow s'$, while the right pair of horizontal arrows may be constructed from $\nu_t \colon t \Rightarrow t'$, and commutativity up to $\langle R_r^{\mathrm{SM}} \rangle$ of either square follows from R^{nat} and possibly R^{inv}. $\square$

So we have a graph isomorphism $\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}} \cong \mathscr{S}_r$. Then if every pair of R_{Σ_r}-related morphisms in $F(\mathscr{S}_r)$ corresponds to a pair of $\approx$-related morphisms in $F(\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}})$, we may invoke Corollary 3.11 to conclude that every solution to the CNP $\mathfrak{G}_{\Sigma_r} = (\mathscr{S}_r, \varnothing, \langle R_{\Sigma_r}\rangle)$ yields a corresponding solution to our quotient CNP $(\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}}, \varnothing, \approx)$.

Proposition 3.82. *Extend the graph isomorphism $\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}} \xrightarrow{\sim} \mathscr{S}_r$ sending $[\varphi]\colon [u] \Rightarrow [v]$ for $\varphi\colon u \Rightarrow v$ in $\mathscr{B}^{\beta}_{\underline{r}}$ to $[\![\varphi]\!]\colon [\![u]\!] \to [\![v]\!]$ to an isomorphism of free categories $F(\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}}) \xrightarrow{\sim} F(\mathscr{S}_r)$. Lifting a pair of R_{Σ_r}-related morphisms in $F(\mathscr{S}_r)$ through this isomorphism yields a pair of $\approx$-related morphisms in $F(\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}})$.*

Proof. For all $\sigma \in \Sigma_r$, we consider each pair of R_{Σ_r}-related morphisms in $F(\mathscr{S}_r)$ with domain σ^* defined in Proposition 3.32, showing that the pair lifts to a pair of $\approx$-related morphisms in $F(\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}})$, i.e. the image under $[\cdot]$ of a pair of $\langle R^{\mathrm{SM}}_r\rangle$-related morphisms in $\mathscr{F}^{\mathrm{SM}}_{\underline{r}}$. We will use the fact that $[\cdot]\colon \mathscr{F}^{\mathrm{SM}}_{\underline{r}} \to F(\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}})$ sends β-basic transformations to their residue edges in $\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}}$ while sending α-, λ-, and ρ-basic transformations and their formal inverses to identities.

(0) Select $s \in \mathcal{L}^{\mathrm{M}}_{\underline{r}}$ for which $[\![s]\!] = \sigma^*$. Then the image under $[\cdot]$ of the diagram

$$\begin{tikzcd} \otimes(s,e) \arrow[r, Rightarrow, bend left, "\beta_{s,e}"] \arrow[r, Rightarrow, bend right, "\lambda^{-1}\bullet\rho"'] & \otimes(e,s) \end{tikzcd}$$

in $\mathscr{F}^{\mathrm{SM}}_{\underline{r}}$, which commutes up to $\langle R^{\mathrm{SM}}_r\rangle$ by (3.73) defining $\mathfrak{R}^{\mathrm{SM}}$, is the diagram

$$\begin{tikzcd} {[s]} \arrow[r, Rightarrow, bend left, "{[\beta_{s,e}]}"] \arrow[r, equal, bend right] & {[s]} \end{tikzcd}$$

in $F(\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}})$ that commutes up to $\approx$, where the bottom arrow is the identity on $[s]$, i.e. the empty path at $[s]$ in $\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}}$. Then our isomorphism sends $[s] \mapsto [\![s]\!] = \sigma^*$ and $[\beta_{s,e}] \mapsto [\![\beta_{s,e}]\!] = \mathrm{id}\colon \sigma^* \Rightarrow \sigma^*$, the edge of $\mathscr{S}_r$, yielding the diagram (3.33) from our definition of R_{Σ_r}.

(1) For $0 \leq i < j < k \leq r$, select $w, s, t, z \in \mathcal{L}_r^{\mathrm{M,reg}}$ so that

$$[\![w]\!] \coloneqq \begin{cases} e & \text{if } i = 0 \\ x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(i)} & \text{otherwise,} \end{cases}$$

$$[\![s]\!] \coloneqq x_{\sigma^{-1}(i+1)} \cdots x_{\sigma^{-1}(j)},$$
$$[\![t]\!] \coloneqq x_{\sigma^{-1}(j+1)} \cdots x_{\sigma^{-1}(k)}, \quad \text{and}$$

$$[\![z]\!] \coloneqq \begin{cases} e & \text{if } k = r \\ x_{\sigma^{-1}(k+1)} \cdots x_{\sigma^{-1}(r)} & \text{otherwise,} \end{cases}$$

so that

$$[\![\otimes(\otimes(w, \otimes(s,t)), z)]\!] = [\![w]\!][\![s]\!][\![t]\!][\![z]\!] = \sigma^*$$
$$\text{and} \quad [\![\otimes(\otimes(w, \otimes(t,s)), z)]\!] = [\![w]\!][\![t]\!][\![s]\!][\![z]\!] = \tau[i,j,k]\sigma^*.$$

Then the diagram

$$\begin{tikzcd} & \otimes(\otimes(w,\otimes(t,s)),z) \arrow[dr, Rightarrow, "\otimes(\otimes(\mathrm{id},\beta),\mathrm{id})"] & \\ \otimes(\otimes(w,\otimes(s,t)),z) \arrow[ur, Rightarrow, "\otimes(\otimes(\mathrm{id},\beta),\mathrm{id})"] \arrow[rr, equal] & & \otimes(\otimes(w,\otimes(s,t)),z) \end{tikzcd}$$

in $\mathscr{F}_{\underline{r}}^{\mathrm{SM}}$ commutes up to R_r^{SM} by (3.72) defining $\mathfrak{R}^{\mathrm{SM}}$. Applying $[\cdot]$ to the above then yields a diagram in $F(\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}})$ that commutes up to $\approx$; then applying our isomorphism yields the diagram

$$\begin{tikzcd} & {[\![w]\!][\![t]\!][\![s]\!][\![z]\!] = \tau[i,j,k]\sigma^*} \arrow[dr, Rightarrow, "\tau[i,i-j+k,k]"] & \\ {\sigma^* = [\![w]\!][\![s]\!][\![t]\!][\![z]\!]} \arrow[ur, Rightarrow, "\tau[i,j,k]"] \arrow[rr, equal] & & {[\![w]\!][\![s]\!][\![t]\!][\![z]\!] = \sigma^*,} \end{tikzcd}$$

which is precisely the diagram (3.34) from our definition of R_{Σ_r}.

(2) For $0 \leq i < j < k \leq i' < j' < k' \leq r$, we may select $w, s, t, y, s', t', z \in \mathcal{L}_r^{\mathrm{M,reg}}$ so that

$$\sigma^* = [\![w]\!][\![s]\!][\![t]\!][\![y]\!][\![s']\!][\![t']\!][\![z]\!],$$
$$\tau[i,j,k]\sigma^* = [\![w]\!][\![t]\!][\![s]\!][\![y]\!][\![s']\!][\![t']\!][\![z]\!],$$
$$\text{and} \quad \tau[i',j',k']\sigma^* = [\![w]\!][\![s]\!][\![t]\!][\![y]\!][\![t']\!][\![s']\!][\![z]\!],$$

then construct a term $u \in \mathcal{L}^{\mathrm{M}}_{\underline{r}}$ with $[\![u]\!] = \sigma^*$, parenthesized appropriately so that $\otimes(s,t)$ and $\otimes(s',t')$ are both subterms of u. Then we may construct four β-basic transformations in a square

$$\begin{tikzcd} u \arrow[r, Rightarrow, "\varphi"] \arrow[d, Rightarrow, "\psi"'] & v \arrow[d, Rightarrow, "\psi'"] \\ u' \arrow[r, Rightarrow, "\varphi'"] & v' \end{tikzcd}$$

with atoms $A(\varphi) = A(\varphi') = \beta_{s,t}$ and $A(\psi) = A(\psi') = \beta_{s',t'}$, and this square commutes up to $R^{\mathrm{com}} \subset R^{\mathrm{SM}}_r$. Applying $[\cdot]$ followed by our isomorphism yields the square

$$\begin{tikzcd} \sigma^* \arrow[r, Rightarrow, "{\tau[i,j,k]}"] \arrow[d, Rightarrow, "{\tau[i',j',k']}"'] & \tau[i,j,k]\sigma^* \arrow[d, Rightarrow, "{\tau[i',j',k']}"] \\ \tau[i',j',k']\sigma^* \arrow[r, Rightarrow, "{\tau[i,j,k]}"] & \tau[i',j',k']\tau[i,j,k]\sigma^*; \end{tikzcd}$$

from (3.35) defining R_{Σ_r}.

(3) Similarly, the square (3.36) arises from applying $[\cdot]$ followed by our isomorphism to a square commuting up to R^{SM}_r of β-basic transformations, here arising from R^{nat}.

(4) Finally, for $0 \le i < j < k < \ell \le r$, we may select $w, p, q, s, z \in \mathcal{L}^{\mathrm{M,reg}}_r$ so that

$$\begin{aligned} \sigma^* &= [\![w]\!][\![p]\!][\![q]\!][\![s]\!][\![z]\!], \\ \tau[i,j,k]\sigma^* &= [\![w]\!][\![q]\!][\![p]\!][\![s]\!][\![z]\!], \\ \text{and} \quad \tau[i,j,\ell]\sigma^* &= [\![w]\!][\![q]\!][\![s]\!][\![p]\!][\![z]\!]. \end{aligned}$$

Then we let $v \coloneqq \otimes(\otimes(p,q),s)$ to construct a term $u \coloneqq \otimes(\otimes(w,v),z)$ in $\mathcal{L}^{\mathrm{M}}_{\underline{r}}$ with $[\![u]\!] = \sigma^*$. By (3.74) defining $\mathfrak{R}^{\mathrm{SM}}$, the following hexagon commutes up to $R^{\mathrm{com}} \subset R^{\mathrm{SM}}_r$ in $\mathscr{F}^{\mathrm{SM}}_{\underline{r}}$:

$$\begin{tikzcd} \otimes(\otimes(p,q),s) \arrow[r, Rightarrow, "\alpha"] \arrow[d, Rightarrow, "{\otimes(\beta,\mathrm{id})}"'] & \otimes(p,\otimes(q,s)) \arrow[r, Rightarrow, "\beta"] & \otimes(\otimes(q,s),p) \arrow[d, Rightarrow, "\alpha"] \\ \otimes(\otimes(q,p),s) \arrow[r, Rightarrow, "\alpha"] & \otimes(q,\otimes(p,s)) \arrow[r, Rightarrow, "{\otimes(\mathrm{id},\beta)}"] & \otimes(q,\otimes(s,p)). \end{tikzcd}$$

Applying $\otimes(\otimes(w,\cdot),z)$ to each term and $\otimes(\otimes(\mathrm{id}_w,\cdot),\mathrm{id}_z)$ to each basic transformation above yields a hexagon with u in the upper left hand corner that also commutes up to

R_r^{SM}. Then applying $[\cdot]$ to the resulting hexagon sends each α-basic transformation to an identity, yielding a triangle that commutes up to $\approx$. Applying $[\![\cdot]\!]$ to that triangle then yields (flipped for readability)

$$\begin{tikzcd}
{[\![w]\!][\![p]\!][\![q]\!][\![s]\!][\![z]\!] = \sigma^*} \arrow[d, Rightarrow, "{\tau[i,j,k]}"'] \arrow[dr, Rightarrow, "{\tau[i,j,\ell]}"] & \\
& {[\![w]\!][\![q]\!][\![s]\!][\![p]\!][\![z]\!] = \tau[i,j,\ell]\sigma^*} \\
{[\![w]\!][\![q]\!][\![p]\!][\![s]\!][\![z]\!] = \tau[i,j,k]\sigma^*} \arrow[ur, Rightarrow, "{\tau[i-j+k,k,\ell]}"'] &
\end{tikzcd} \tag{3.83}$$

from (3.37) defining R_{Σ_r}. $\square$

Corollary 3.84. *For all $t \in \mathcal{L}_{\underline{r}}^{\mathrm{M}}$, the singleton set $\{[t]\}$ is a solution to the quotient CNP $(\mathscr{B}_{\underline{r}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}}\rangle)/\mathrm{Bs}_{\underline{r}}^{\mathrm{M}} = (\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}, \varnothing, \approx)$.*

Proof. In Proposition 3.32, we saw that any singleton subset of the vertex-set of $\mathscr{S}_r$ is a solution to the CNP $(\mathscr{S}_r, \varnothing, \langle R_{\Sigma_r}\rangle)$. The graph isomorphism $\mathscr{S}_r \xrightarrow{\sim} \overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}$ extends to a category isomorphism $F(\mathscr{S}_r) \xrightarrow{\sim} F(\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}})$ sending morphisms related by the local binary relation R_{Σ_r} to morphisms related by the congruence $\approx$, so it necessarily sends morphisms related by the congruence $\langle R_{\Sigma_r}\rangle$ generated by R_{Σ_r} to morphisms related by the congruence $\approx$ as well. So by Proposition 3.10, the image under this isomorphism of any solution to $(\mathscr{S}_r, \varnothing, \langle R_{\Sigma_r}\rangle)$ is a solution to $(\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}, \varnothing, \approx)$, and the result follows. $\square$

So we may apply Proposition 3.69 to deduce a solution to our CNP $(\mathscr{B}_{\underline{r}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}}\rangle)$ from the solution $\overline{\mathcal{N}_{\underline{r}}}$ to the CNP $(\mathscr{B}_{\underline{r}}^{\mathrm{M}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}}\rangle)$ and any singleton solution to the quotient CNP $(\mathscr{B}_{\underline{r}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}}\rangle)/\mathrm{Bs}_{\underline{r}}^{\mathrm{M}} = (\overline{\mathscr{B}_{\underline{r}}^{\mathrm{SM}}}, \varnothing, \approx)$.

Proposition 3.85. *Any term $t \in \overline{\mathcal{N}_{\underline{r}}}$ comprises a singleton solution $\{t\}$ to the CNP $(\mathscr{B}_{\underline{r}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}}\rangle)$.*

In particular, a solution for $(\mathscr{B}_{\underline{0}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{0}}^{\mathrm{M}}, \langle R_0^{\mathrm{SM}}\rangle)$ is $\{e\}$, while a possible solution for $(\mathscr{B}_{\underline{r}}^{\mathrm{SM}}, \mathrm{Bs}_{\underline{r}}^{\mathrm{M}}, \langle R_r^{\mathrm{SM}}\rangle)$ for $r \geq 1$ is the singleton set

$$\{\otimes(x_1, \otimes(x_2, \cdots, \otimes(x_{r-1}, x_r)\cdots))\}.$$

Proof. We have $\mathrm{Bs}^{\mathrm{M}}_{\underline{r}} \lhd (\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_{r}\rangle)$, with the set $\overline{\mathcal{N}_{\underline{r}}}$ a solution to $(\mathscr{B}^{\mathrm{M}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_{r}\rangle)$. Moreover, $\{[t]\}$ is a solution to the quotient CNP $(\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_{r}\rangle)/\mathrm{Bs}^{\mathrm{M}}_{\underline{r}} = (\overline{\mathscr{B}^{\mathrm{SM}}_{\underline{r}}}, \varnothing, \approx)$. It follows from Proposition 3.69 that the set

$$M \coloneqq \{t' \in \overline{\mathcal{N}_{\underline{r}}} \mid [t'] = [t]\}$$

is a solution to $\mathrm{Bs}^{\mathrm{M}}_{\underline{r}} \lhd (\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_{r}\rangle)$. But if $[t'] = [t]$, then there is a morphism $t \Rightarrow t'$ in $\mathscr{F}^{\mathrm{M}}_{\underline{r}}$, which is only possible for $t, t' \in \overline{\mathcal{N}_{\underline{r}}}$ if $t = t'$. So in fact our solution set is just $M = \{t\}$. ☐

Finally, we may apply Proposition 3.41 to Proposition 3.76 to construct a solution to $\mathrm{CN}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}^{\mathrm{M,reg}}_{k})$ as a coproduct, yielding our version of Mac Lane's coherence theorem for symmetric monoidal categories.

Theorem 3.86 (Mac Lane)**.** *The set of regular terms*

$$\{e\} \sqcup \{\otimes(x_{\iota(1)}, \otimes(x_{\iota(2)}, \cdots \otimes(x_{\iota(r-1)}, x_{\iota(r)}) \cdots)) \mid 1 \le r \le k,\ \iota\colon \underline{r} \to \underline{k} \text{ injective}\}$$

is a solution to $\mathrm{CN}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}^{\mathrm{M,reg}}_{k})$*, and the inclusion functor from this set viewed as a discrete category to the universal parameter category* $\mathscr{P}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}^{\mathrm{M,reg}}_{k})$ *for monoidal categories in* k *variables is an equivalence.*

That is, for every regular term $t \in \mathcal{L}^{\mathrm{M,reg}}_{k}$*, there exists a unique term* $n(t)$ *in the given solution set such that there is a morphism* $t \Rightarrow n(t)$ *in* $\mathscr{P}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}^{\mathrm{M,reg}}_{k})$*; and given terms* $t, t' \in \mathcal{L}^{\mathrm{M,reg}}_{k}$ *with* $n(t) = n(t')$*, there exists a unique morphism* $t \Rightarrow t'$ *in* $\mathscr{P}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}^{\mathrm{M,reg}}_{k})$*.*

So let $|\cdot|\colon \mathfrak{C}_{\mathrm{SM}} \to \underline{\mathbf{K}}$ *be a model for* $\mathfrak{C}_{\mathrm{SM}}$ *in a cartesian monoidal* 2*-category* $\underline{\mathbf{K}}$*, giving rise to the functor*

$$|\cdot|_{V_k}\colon \mathscr{P}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}^{\mathrm{M,reg}}_{k}) \to \underline{\mathbf{K}}(|\mathtt{m}|^{k}, |\mathtt{m}|)$$

from (2.88)*. Then every diagram in* $\underline{\mathbf{K}}(|\mathtt{m}|^{k}, |\mathtt{m}|)$ *that factors through* $|\cdot|_{V_k}$ *is a commutative diagram of isomorphisms.*

Proof. We constructed an isomorphism between the CNP in question and a coproduct of CNPs in Proposition 3.76, and we have solved each summand $(\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_{k}\rangle)$ by solving

$(\mathscr{B}^{\mathrm{SM}}_{\underline{r}}, \mathrm{Bs}^{\mathrm{M}}_{\underline{r}}, \langle R^{\mathrm{SM}}_{r}\rangle)$ and applying Corollary 3.11; so by Proposition 3.41, the coproduct CNP has the coproduct

$$\{e\} \sqcup \coprod_{1\leq r\leq k} \coprod_{i=1}^{\binom{k}{r}} \{\otimes(x_1, \otimes(x_2, \cdots, \otimes(x_{r-1}, x_r)\cdots))\}$$

as a solution. Passing through the isomorphism, adjusting indices accordingly, yields the desired solution.

Note that while we did not explicitly mark the edge label β in the coherence schema $\mathfrak{C}_{\mathrm{SM}}$, we included a relation in $\mathfrak{R}^{\mathrm{SM}}$ that gave each atomic transformation $\beta_{s,t}$ for $s, t \in \mathcal{L}^{\mathrm{M,reg}}_k$ an inverse $\beta_{t,s}$ up to $\langle R^{\mathrm{SM}}_k\rangle$, so our universal parameter category $\mathscr{P}(\mathfrak{C}_{\mathrm{SM}}, \mathcal{L}^{\mathrm{M,reg}}_k)$ nevertheless must be a groupoid. The rest follows. $\square$

Appendix A
CATEGORY THEORY PRELIMINARIES

Here we review the standard categorical concepts we use.

A.1 Directed graphs and free categories

We use the usual category-theoretic definition of *graphs* and *graph morphisms*, following [16, Section II.7]. Thus our graphs are directed, allowing self-loops and parallel edges.

Definition A.1 (Graph)**.** A *(directed) graph* $\mathscr{G} = [E \rightrightarrows V]$ consists of:

- a set V of *vertices*;
- a set E of *edges*;
- a *domain function* $\mathrm{dom}\colon E \to V$; and
- a *codomain function* $\mathrm{cod}\colon E \to V$.

For each $e \in E$, we write $e\colon \mathrm{dom}(e) \Rightarrow \mathrm{cod}(e)$.[1]

Definition A.2 (Graph morphism, subgraph, graph isomorphism)**.** If we are given graphs $\mathscr{G}_i = [E_i \rightrightarrows V_i]$ for $i = 1, 2$, then a *graph morphism* $f\colon \mathscr{G}_1 \to \mathscr{G}_2$ consists of:

- an *on-vertices function* $f\colon V_1 \to V_2$ and
- an *on-edges function* $f\colon E_1 \to E_2$;

[1] We use the double-arrow $\Rightarrow$ instead of the single-arrow $\to$ because our edges will come to represent 2-cells (natural transformations) between the 1-cells (functors) represented by our vertices.

such that every edge $e\colon v \Rightarrow w$ in $\mathscr{G}_1$ is sent via f to an edge $f(e)\colon f(v) \Rightarrow f(w)$ in $\mathscr{G}_2$. If both functions are inclusions, $\mathscr{G}_1$ is a *subgraph* of $\mathscr{G}_2$, and if further E_1 contains every edge in E_2 between vertices in V_1, then $\mathscr{G}_1$ is a *full subgraph* of $\mathscr{G}_2$. A *graph isomorphism* is a graph morphism that is bijective on both vertices and edges.

Every (small) category has an underlying graph.

Definition A.3 (Underlying graph)**.** Given a small category $\mathscr{C}$, the *underlying graph of* $\mathscr{C}$ is the graph whose vertices are the objects of $\mathscr{C}$, whose edges are the morphisms of $\mathscr{C}$, and whose domain and codomain functions coincide with that of $\mathscr{C}$. Typically we denote this underlying graph by $\mathscr{C}$ as well, relying on context to interpret it as a graph or as a category.

Conversely, every graph gives rise to a free category.

Definition A.4 (Free category)**.** Given a graph $\mathscr{G} = [E \rightrightarrows V]$, the *free category on* $\mathscr{G}$, denoted $F(\mathscr{G})$, is the category whose objects are vertices in V and whose morphisms are finite (possibly empty) paths of composable edges in E. More precisely, a morphism in $F(\mathscr{G})$ from a vertex $v_0 \in V$ to a vertex $v \in V$ is a sequence

$$(v, e_n, v_{n-1}, \ldots, v_2, e_2, v_1, e_1, v_0)$$

for $n \geq 0$ and $v_1, \ldots, v_{n-1}, v_n \coloneqq v \in V$, $e_1, \ldots, e_n \in E$, and $e_i\colon v_{i-1} \Rightarrow v_i$ for all $1 \leq i \leq n$. The sequence represents a (directed) path in $\mathscr{G}$ of length n. Composition is given by path concatenation (i.e. concatenating sequences but keeping only one copy of the intermediate duplicate vertex), which we denote by $\bullet$. The sequence $1_{v_0} \coloneqq (v_0)$, i.e. the empty path, is the identity at v_0.

There is a canonical graph morphism $\mathscr{G} \to F(\mathscr{G})$ from $\mathscr{G}$ to the underlying graph of its free category, sending each vertex of $\mathscr{G}$ to itself as an object and each edge $e\colon v \Rightarrow w$ of $\mathscr{G}$ to the morphism (w, e, v), the length-1 path. Typically we abuse notation by writing just e for its image (w, e, v) in $F(\mathscr{G})$ and denote the morphism $(v_n, e_n, \ldots, e_1, v_0)$ by $e_n \bullet \cdots \bullet e_1$.

A.2 Congruences and quotient categories

We may designate certain classes of morphisms in a category as equivalent using a *congruence* or as equal by taking a *quotient*, following [16, Section II.8] and [1, Definition 4.6].

Throughout this section, let $\mathscr{C}$ be a category.

Definition A.5 (Local binary relation, congruence)**.** A *local binary relation R on $\mathscr{C}$* is a tuple $(R_{x,y})_{x,y\in\mathscr{C}}$ where each $R_{x,y}$ is a binary relation on the hom-set $\mathscr{C}(x,y)$ (we may omit the subscripts when they are clear from context) A *congruence $\sim$ on $\mathscr{C}$* is a local binary relation $\sim$ on $\mathscr{C}$ such that $\sim_{x,y}$ is an equivalence relation on $\mathscr{C}(x,y)$ for all objects $x,y\in\mathscr{C}$, and for any objects $x,y,z\in\mathscr{C}$ and morphisms $f,f'\colon x\rightrightarrows y$ and $g,g'\colon y\rightrightarrows z$, if $f\sim_{x,y} f'$ and $g\sim_{y,z} g'$, then $(g\circ f)\sim_{x,z}(g'\circ f')$.[2]

Definition A.6 (Quotient category)**.** Let $\sim$ be a congruence on $\mathscr{C}$. The *quotient category* $\overline{\mathscr{C}} := \mathscr{C}/{\sim}$ has the same objects as $\mathscr{C}$ and the hom-sets

$$\overline{\mathscr{C}}(x,y) := \mathscr{C}(x,y)/{\sim_{x,y}}\ .$$

Identities and composition in $\overline{\mathscr{C}}$ are given by $\mathrm{id}_x := [\mathrm{id}_x]$ and $[g]\circ[f] := [g\circ f]$ (which the congruence condition ensures is well-defined). There is a canonical projection functor

$$\pi\colon \mathscr{C}\to\overline{\mathscr{C}}$$

that is the identity on objects and sends $f\colon x\to y$ in $\mathscr{C}$ to its congruence class $[f]\colon x\to y$ in $\overline{\mathscr{C}}$.

Given a congruence $\sim$ on $\mathscr{C}$, the quotient category $\overline{\mathscr{C}} = \mathscr{C}/{\sim}$ equipped with its projection functor $\pi\colon\mathscr{C}\to\overline{\mathscr{C}}$ is universal in the following sense: π has the property

$$\text{for any } x,y\in\mathscr{C} \text{ and } f,f'\in\mathscr{C}(x,y)\text{: if } f\sim_{x,y} f'\text{, then } \pi(f)=\pi(f'), \tag{$*$}$$

and any functor $F\colon\mathscr{C}\to\mathscr{D}$ that also has the property $(*)$ factors uniquely through π as a functor $\overline{F}\colon\overline{\mathscr{C}}\to\mathscr{D}$. The functor $\overline{F}$ is necessarily given by $\overline{F}x = Fx$ on objects $x\in\mathscr{C}$ and $\overline{F}[f] = Ff$ on morphisms f in $\mathscr{C}$.

Example A.7 (Common congruences and quotient categories)**.** (1) We can see that the identity relations $=_{x,y}$ on each hom-set $\mathscr{C}(x,y)$ assemble into a congruence. It is the smallest congruence on $\mathscr{C}$, and $\overline{\mathscr{C}}\cong\mathscr{C}$.

---

[2]Equivalently, $f\sim f'$ implies $(k\circ f\circ h)\sim(k\circ f'\circ h)$ whenever these composites are defined.

(2) The codiscrete relations $\sim_{x,y}$ given by $f \sim_{x,y} f'$ for all $f, f' \in \mathscr{C}(x,y)$ assemble into a congruence. It is the largest congruence on $\mathscr{C}$, and $\overline{\mathscr{C}}$ is the *preorder reflection of* $\mathscr{C}$.

(3) If $\{\sim^i\}_{i\in I}$ is a nonempty indexed set of congruences on $\mathscr{C}$, then we have a congruence given by their intersection $\sim \coloneqq \bigcap_{i\in I} \sim^i$, defined explicitly by

$$f \sim_{x,y} f' \quad \text{if and only if} \quad f \sim^i_{x,y} f' \quad \text{for all } i \in I.$$

(4) In particular, if $\mathscr{C}$ is small and has a local binary relation R, there is a nonempty set

$$\{\sim^i\}_{i\in I} = \{\sim \text{ is a congruence on } \mathscr{C} \mid R_{x,y} \subseteq \sim_{x,y} \text{ for all } x, y \in \mathscr{C}\}$$

of all congruences on $\mathscr{C}$ that contain R. In this case, we denote the intersection $\bigcap_{i\in I} \sim^i$ by $\langle R\rangle$ and call it the *congruence generated by* R. It is the smallest congruence on $\mathscr{C}$ that contains R. Explicitly, it can also be constructed as follows: define a local binary relation $\overline{R}$ from R by $(k \circ f \circ h)\ \overline{R}_{t,w}\ (k \circ f' \circ h)$ whenever $f\ R_{u,v}\ f'$ for $f, f' \colon u \rightrightarrows v, h \colon t \to u$, and $k \colon v \to w$ in $\mathscr{C}$ Then for $x, y \in \mathscr{C}$, we let $\langle R\rangle_{x,y}$ be the reflexive, symmetric, and transitive closure of $\overline{R}_{x,y}$.

(5) If $F \colon \mathscr{C} \to \mathscr{D}$ is a functor and $\approx$ is a congruence on $\mathscr{D}$, the *inverse image of* $\approx$ *along* F is the congruence $\sim$ on $\mathscr{C}$ given by

$$f \sim f' \quad \text{if and only if} \quad Ff \approx Ff'.$$

for parallel morphisms f, f' in $\mathscr{C}$. So F descends uniquely to quotients, yielding a new functor $\overline{F} \colon (\mathscr{C}/\sim) \to (\mathscr{D}/\approx)$ that coincides with F on objects and sends morphism $[f]$ in $(\mathscr{C}/\sim)$ to $[Ff]$ in $(\mathscr{D}/\approx)$. We have the following special cases of this construction.

(i) If $F \colon \mathscr{C} \to \mathscr{D}$ is a subcategory embedding and $\approx$ is any congruence on $\mathscr{D}$, then the inverse image $\sim$ of $\approx$ along F is the congruence given by

$$f \sim f' \quad \text{if and only if} \quad f \approx f'$$

for parallel morphisms f, f' in $\mathscr{C} \subseteq \mathscr{D}$. We call $\sim$ the *restriction of* $\approx$ *to* $\mathscr{C}$. Then $\overline{F}$ is a subcategory embedding as well: it coincides with F on objects, and on morphisms, given parallel morphisms f, f' in $\mathscr{C}$ for which $\overline{F}[f] = \overline{F}[f']$, we have $[Ff] = [Ff']$ and thus $[f] = [f']$. Abusing notation, we generally denote $\sim$ also by $\approx$.

(ii) If $\approx$ is the identity relation and $F\colon \mathscr{C} \to \mathscr{D}$ is any functor, then the inverse image $\sim_F$ of the identity relation along F is given by

$$f \sim_F f' \quad \text{if and only if} \quad Ff = Ff'$$

for parallel morphisms f, f' in $\mathscr{C}$. We call $\sim_F$ the *kernel of* F. By the universal property of the quotient category, F factors uniquely through the projection functor to yield a new functor $\overline{F}\colon (\mathscr{C}/{\sim_F}) \to \mathscr{D}$.

Example A.8 (Monoid congruence, quotient monoid)**.** Suppose that M is a monoid. A *congruence* $\sim$ *on* M is a congruence on its corresponding 1-object category, namely an equivalence relation on M that respects its multiplication. We can quotient out by $\sim$ to obtain the *quotient monoid* $M/{\sim}$. Given a binary relation R on M, the congruence $\langle R\rangle$ it generates is the reflexive, symmetric, and transitive closure of the relation $\approx$ on M given by $sut \approx svt$ for $s, t, u, v \in M$ when $u\, R\, v$.

In practice, the local binary relation R that we wish to impose on a category $\mathscr{C}$ may not be a congruence, so we must first close R up to $\langle R\rangle$ before taking a quotient. Fortunately, the universal property of $\mathscr{C}/\langle R\rangle$ is determined by R alone (see [16, Section II.8]).

Proposition A.9. *Suppose $\mathscr{C}$ is small with local binary relation R. The quotient projection*

$$\pi\colon \mathscr{C} \to \mathscr{C}/\langle R\rangle$$

has the property

$$\text{for any } x, y \in \mathscr{C} \text{ and } f, f' \in \mathscr{C}(x, y)\text{: if } f\ R_{x,y}\ f'\text{, then } \pi(f) = \pi(f'), \tag{$**$}$$

and any functor $F\colon \mathscr{C} \to \mathscr{D}$ with this property factors uniquely through π.

Proof. If $f\ R_{x,y}\ f'$, then $f\ \langle R\rangle_{x,y}\ f'$ and hence $\pi(f) = \pi(f')$.

On the other hand, if F has property $(**)$, then R is contained in the kernel of F. Therefore $\langle R\rangle$ is also contained in the kernel, i.e. for any objects $x, y \in \mathscr{C}$ and $f, f' \in \mathscr{C}(x, y)$, if $f \langle R\rangle f'$ then $Ff = Ff'$. Therefore $F\colon \mathscr{C} \to \mathscr{D}$ factors uniquely through $\pi\colon \mathscr{C} \to \mathscr{C}/\langle R\rangle$ by the universal property of the quotient projection. $\square$

The considerations above show that congruences on a category $\mathscr{C}$ are the same thing as kernels of functors out of $\mathscr{C}$. However, if we restrict attention to functors $F\colon \mathscr{C} \to \mathscr{D}$ with additional conditions on $\mathscr{D}$ (e.g. $\mathscr{D}$ is a groupoid), then there will also be additional conditions on our kernels. We highlight one useful condition.

Definition A.10 (Left-cancelability)**.** Let $\sim$ be a congruence on $\mathscr{C}$. We say that $\sim$ is *left-cancelable* if for any objects $x, y, z \in \mathscr{C}$ and morphisms $f, f'\colon x \rightrightarrows y$ and $g\colon y \to z$, if $(g \circ f) \sim_{x,z} (g \circ f')$, then $f \sim_{x,y} f'$.

Quotienting by a left-cancelable congruence makes every morphism a monomorphism.

Lemma A.11. *Let $\sim$ be a left-cancelable congruence on $\mathscr{C}$. Then every morphism in the quotient category $\overline{\mathscr{C}} = \mathscr{C}/\sim$ is a monomorphism.*

Proof. Given morphisms $[f], [f']\colon x \rightrightarrows y$ and $[g]\colon y \to z$ in $\overline{\mathscr{C}}$ that are residues of morphisms $f, f'\colon x \rightrightarrows y$ and $g\colon y \to z$ in $\mathscr{C}$, assume $[g] \circ [f] = [g] \circ [f']$. Then $[g \circ f] = [g \circ f']$, so $(g \circ f) \sim_{x,z} (g \circ f')$. By left-cancelability, $f \sim_{x,y} f'$, so $[f] = [f']$. $\square$

As with ordinary congruences,

- the maximum congruence on $\mathscr{C}$ is (trivially) left-cancelable, and
- if $\{\sim^i\}_{i \in I}$ is a nonempty indexed set of left-cancelable congruences, then their intersection $\bigcap_{i \in I} \sim^i$ is also left-cancelable.

It follows that if $\mathscr{C}$ is a small category and R is a local binary relation, then there is a smallest left-cancelable congruence that contains R, which we denote

$$\langle R\rangle^{\mathrm{L}} := \bigcap \left\{ \begin{array}{c} \sim \text{ is a left-cancelable} \\ \text{congruence on } \mathscr{C} \end{array} \,\middle|\, R_{x,y} \subseteq \sim_{x,y} \text{ for all } x, y \in \mathscr{C} \right\}, \tag{A.12}$$

and the quotient $\mathscr{C}/\langle R\rangle^{\mathrm{L}}$ has the following universal property.

Proposition A.13. *Suppose $\mathscr{C}$ is small with local binary relation R, and consider the quotient projection*

$$\pi\colon \mathscr{C} \to \mathscr{C}/\langle R\rangle^{\mathrm{L}}.$$

Then every morphism in $\mathscr{C}/\langle R\rangle^{\mathrm{L}}$ is a monomorphism and

$$\text{for any } x, y \in \mathscr{C} \text{ and } f, f' \in \mathscr{C}(x,y)\text{: if } f\ R_{x,y}\ f'\text{, then } \pi(f) = \pi(f'). \tag{$***$}$$

Moreover, if $\mathscr{D}$ is a category in which every morphism is a monomorphism, and $F\colon \mathscr{C} \to \mathscr{D}$ is a functor that satisfies $(***)$, then F factors uniquely through π.

Proof. By Lemma A.11, every morphism in $\mathscr{C}/\langle R\rangle^{\mathrm{L}}$ is a monomorphism; and if $f\ R_{x,y}\ f'$, then by construction $f\ \langle R\rangle^{\mathrm{L}}\ f'$, so $\pi(f) = \pi(f')$, satisfying $(***)$.

On the other hand, if F has property $(***)$, then R is contained in the kernel $\sim_F$ of F. We claim that $\sim_F$ is left-cancelable. Fix parallel morphisms f, f' in $\mathscr{C}$ whose common codomain is the domain of some g in $\mathscr{C}$ with $(g \circ f) \sim_F (g \circ f')$, so

$$Fg \circ Ff = F(g \circ f) = F(g \circ f') = Fg \circ Ff'$$

in $\mathscr{D}$. As Fg is monic, $Ff = Ff'$, so $f \sim_F f'$. Hence $\sim_F$ is a left-cancelable congruence containing R. Thus $\sim_F$ also contains $\langle R\rangle^{\mathrm{L}}$, so F factors uniquely through π. $\square$

We conclude by noting that the inverse image of a left-cancelable congruence is still left-cancelable.

Lemma A.14. *Let $F\colon \mathscr{C} \to \mathscr{D}$ be a functor and $\approx$ be a left-cancelable congruence on $\mathscr{C}$. Then the congruence $\sim$ on $\mathscr{C}$ given by the inverse image of $\approx$ along F, as in Example A.7 (5), is also left-cancelable.*

Proof. Fix objects $x, y, z \in \mathscr{C}$ and morphisms $f, f'\colon x \rightrightarrows y$ and $g\colon y \to z$ for which $(g \circ f) \sim (g \circ f')$. Then

$$Fg \circ Ff = F(g \circ f) \approx F(g \circ f') = Fg \circ Ff'.$$

Since $\approx$ is left-cancelable, it follows that $Ff \approx Ff'$, so $f \sim f'$, as desired. $\square$

Appendix B
MONOID & GROUP THEORY PRELIMINARIES

Here we review the standard monoid- and group-theoretic constructions we use. Unless otherwise stated, we denote monoid and group multiplication by juxtaposition.

B.1 Translation categories for actions

First, we define *translation groupoids* for group actions and, more generally, *translation categories* for monoid actions. We denote monoid and group multiplication by juxtaposition, and we will use left actions by convention.

Definition B.1 (Translation category, translation groupoid)**.** Let M be a monoid whose unit is e, and let X be a left M-set. The *translation category of* X is the category $\mathbb{T}_M X$ defined as follows.

(1) Its objects are the elements of X: we have $\mathrm{Ob}(\mathbb{T}_M X) \coloneqq X$.

(2) Its hom-sets for $x, y \in X$ are given by

$$(\mathbb{T}_M X)(x, y) \coloneqq \{\mu \in M \mid \mu x = y\}.$$

That is, for each $\mu \in M$ and $x \in X$, there is a morphism $\mu \colon x \to \mu x$.

(3) The identity morphism on $x \in X$ is e: we have $\mathrm{id}_x \coloneqq e \colon x \to x$.

(4) Composition is given by monoid multiplication: the composite

$$x \xrightarrow{\mu} \mu x \xrightarrow{\mu'} \mu' \mu x$$

is given by $\mu' \mu \colon x \to \mu' \mu x$.

Unitality and associativity of the monoid imply unitality and associativity of the category.

If M is a group, $\mathbb{T}_M X$ is a *translation groupoid*, with inverses given by group inverses: the inverse of $\gamma\colon x \to \gamma x$ is $\gamma^{-1}\colon \gamma x \to x$.

Note that this construction is functorial in both M and X. Given a monoid map $f\colon N \to M$, we can also view X as a left N-set via f, and we have a functor of translation categories $\mathbb{T}_f X\colon \mathbb{T}_N X \to \mathbb{T}_M X$ that is the identity on objects and sends each morphism $n\colon x \to y$ in $\mathbb{T}_N X$ to $f(n)\colon x \to y$ in $\mathbb{T}_M X$. Meanwhile, given an M-equivariant map $h\colon X \to Y$, we have a functor of translation categories $\mathbb{T}_M h\colon \mathbb{T}_M X \to \mathbb{T}_M Y$ that coincides with h on objects and sends each morphism $m\colon x \to mx$ in $\mathbb{T}_M X$ to $m\colon h(x) \to h(mx) = mh(x)$.

B.2 Presentations

Next, we fix notation and terminology for free monoids and free groups as well as monoid and group representations by generators and relations.

Definition B.2 (Free monoid/group, monoid/group presentation)**.** (1) Given a set S, let $F_{\mathrm{mon}}S$ denote the *free monoid* on S. Each $s \in S$ is itself an element of $F_{\mathrm{mon}}S$.

(2) A monoid M is *presented (as a monoid) by generators G and relations R* if $G \subseteq M$, if R is a relation on $F_{\mathrm{mon}}G$, and if the monoid homomorphism $F_{\mathrm{mon}}G \to M$ induced by freely extending the set inclusion $G \hookrightarrow M$ descends to a monoid isomorphism $F_{\mathrm{mon}}G/\langle R\rangle \xrightarrow{\sim} M$, where $\langle R\rangle$ is the congruence relation on $F_{\mathrm{mon}}G$ generated by R (see Example A.8). Typically we denote the residue in $F_{\mathrm{mon}}G/\langle R\rangle$ of a free product of generators $x \in F_{\mathrm{mon}}G$ by $[x]$.

(3) Given a set S, let $F_{\mathrm{grp}}S$ denote the *free group* on S. Each $s \in S$ is itself an element of $F_{\mathrm{grp}}S$ with inverse s^{-1}.

(4) A group Γ is *presented (as a group) by generators G and relations R* if $G \subseteq \Gamma$, if R is a relation on $F_{\mathrm{grp}}G$, and if the group homomorphism $F_{\mathrm{grp}}G \to \Gamma$ induced by freely extending the set inclusion $G \hookrightarrow \Gamma$ descends to a group isomorphism $F_{\mathrm{grp}}G/\langle\langle R\rangle\rangle \xrightarrow{\sim} \Gamma$, where $\langle\langle R\rangle\rangle$ is the normal subgroup of $F_{\mathrm{grp}}G$ generated by $\{y^{-1}x \mid x, y \in F_{\mathrm{grp}}G, xRy\}$.

Typically we denote the residue in $F_{\mathrm{grp}}G/\langle\langle R\rangle\rangle$ of a free product of generators and their inverses $y \in F_{\mathrm{grp}}G$ by $[[y]]$.

Note that a set G is a subset of the free monoid $F_{\mathrm{mon}}G$, which in turn embeds as a submonoid of the free group $F_{\mathrm{grp}}G$ viewed as a monoid.

The following lemma characterizes when monoid and group presentations coincide.

Lemma B.3. *Suppose G is a set of generators and R is a binary relation on $F_{\mathrm{mon}}G \subset F_{\mathrm{grp}}G$; together, they present the monoid $M \coloneqq F_{\mathrm{mon}}G/\langle R\rangle$ and the group $\Gamma \coloneqq F_{\mathrm{grp}}G/\langle\langle R\rangle\rangle$. Then there is a commutative triangle of functions*

$$\begin{tikzcd} & G \arrow[dl, "\iota"'] \arrow[dr, "\iota'"] & \\ M \arrow[rr, "\kappa"] & & \Gamma, \end{tikzcd}$$

where the functions ι and ι' send generators to their own residues, and the horizontal comparison monoid map κ is induced by the universal property of M. Then κ is an isomorphism if and only if M is a group (for instance, if for every $g \in G$ there exists $n(g) \in \mathbb{N}$ for which $g^{n(g)}\, R\, e$).

Proof. The function ι' freely extends to a monoid map $\varphi\colon F_{\mathrm{mon}}G \to \Gamma$ sending $x \in F_{\mathrm{mon}}G \subset F_{\mathrm{grp}}G$ to its residue $[[x]] \in \Gamma$. Then given $x\, R\, y$ for $x, y \in F_{\mathrm{mon}}G$, we have $y^{-1}x \in \langle\langle R\rangle\rangle$, so its residue $[[y^{-1}x]]$ in Γ is the identity. Thus $[[x]] = [[y]]$, ensuring $\varphi(x) = \varphi(y)$. So φ descends to the quotient M, inducing a monoid map $\kappa\colon M \to \Gamma$ sending the residue $[x] \in M$ to the residue $[[x]] \in \Gamma$ for all $x \in F_{\mathrm{mon}}G$, In particular, for $g \in G \subset F_{\mathrm{mon}}G$, we have

$$\kappa(\iota(g)) = \kappa([g]) = [[g]] = \iota'(g),$$

making the given triangle commute.

If κ is an isomorphism of monoids, then its domain is a group because its codomain is, so we are done. Conversely, assume M is a group, making κ a group homomorphism; we seek its inverse. The function ι freely extends to a group homomorphism $\varphi'\colon F_{\mathrm{grp}}G \to M$, which sends $x \in F_{\mathrm{mon}}G \subset F_{\mathrm{grp}}G$ to its residue $[x] \in M$, and thus sends $y^{-1} \in F_{\mathrm{grp}}G$ for

$y \in F_{\text{mon}}G$ to $[y]^{-1} \in M$. In particular, there exists $y' \in F_{\text{mon}}G$ whose residue in M is $[y'] = [y]^{-1}$, so $[y'y]$ is the identity of M.

Then if $x\,R\,y$, we have $y'x \,\langle R\rangle\, y'y \,\langle R\rangle\, e$, the identity of $F_{\text{mon}}G$. Therefore in M we have

$$\varphi'(y^{-1}x) = [y]^{-1}[x] = [y'][x] = [y'x] = [e],$$

the identity of M, so $y^{-1}x$ is in the kernel of φ' for all $x, y \in F_{\text{mon}}G$ with $x\,R\,y$. Thus φ' descends to the quotient Γ, inducing a group homomorphism $\lambda\colon \Gamma \to M$ that sends $[[x]] \mapsto [x]$ as well as $[[y^{-1}]] \mapsto [y]^{-1}$ for $x, y \in F_{\text{mon}}G$. Plainly $\lambda \circ \kappa$ is the identity on M, while

$$\kappa(\lambda([[y^{-1}]])) = \kappa([y]^{-1}) = \kappa([y])^{-1} = [[y]]^{-1} = [[y^{-1}]].$$

so $\kappa \circ \lambda$ is the identity on Γ. $\square$

We will make use of this lemma in the following section, when we give group presentations that are also monoid presentations for the symmetric group.

B.3 The symmetric group

We review classical properties of the symmetric group before considering its presentation by *block transpositions*, which feature in our coherence proofs for symmetric structures. Unless otherwise stated, fix a positive integer r.

Definition B.4 (Symmetric group, permutation)**.** Let $\underline{r} \coloneqq \{1, \ldots, r\}$ denote the set of the smallest r positive integers. Then the r^{th} *symmetric group* Σ_r is the automorphism group of $\underline{r}$; its elements are called *permutations*. Multiplication of permutations is denoted by juxtaposition, written in compositional order: for $\sigma, \tau \in \Sigma_r$, we have $\sigma\tau \coloneqq \sigma \circ \tau$, with $(\sigma\tau)(k) = \sigma(\tau(k))$ for $k \in \underline{r}$. We denote the identity permutation by id.

Given $\sigma \in \Sigma_r$, contrary to the usual convention, we sometimes denote σ by

$$[\sigma^{-1}(1), \sigma^{-1}(2), \ldots, \sigma^{-1}(n)] \quad \text{or} \quad \Big[\sigma^{-1}(1) \quad \sigma^{-1}(2) \quad \cdots \quad \sigma^{-1}(r).\Big]$$

This is because we primarily think of permutations in terms of their action on ordered tuples. Throughout this section, we let $V_r \coloneqq \{x_1, \ldots, x_r\}$ denote a set of n distinct formal symbols x_k for $k \in \underline{r}$, and we let $\mathcal{S}_r$ denote the set of length-r formal words (i.e. associative products, in which we may insert parentheses only for clarity) comprised of these r symbols with no repeats. Equivalently,

$$\mathcal{S}_r \coloneqq \{x_{\sigma^{-1}(1)}x_{\sigma^{-1}(2)} \cdots x_{\sigma^{-1}(r)} \mid \sigma \in \Sigma_r\}. \tag{B.5}$$

Then there is a standard (left) action of Σ_r on $\mathcal{S}_r$ that is isomorphic to the left action of Σ_r on itself (i.e. the free orbit of Σ_r), defined as follows. For $\sigma \in \Sigma_r$, we have

$$\sigma \cdot (x_1 \cdots x_r) \coloneqq x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(r)}.$$

That is, σ moves the k^{th} variable to the $\sigma(k)^{\text{th}}$ position. More generally, given $\rho \in \Sigma_r$,

$$\begin{aligned} \rho \cdot (x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(r)}) &= x_{(\sigma^{-1}\rho^{-1})(1)} \cdots x_{(\sigma^{-1}\rho^{-1})(r)} \\ &= x_{(\rho\sigma)^{-1}(1)} \cdots x_{(\rho\sigma)^{-1}(r)} \\ &= (\rho\sigma) \cdot (x_1 \cdots x_r), \end{aligned}$$

with ρ moving the k^{th} variable $x_{\sigma^{-1}(k)}$ to the $\rho(k)^{\text{th}}$ position. So we may interpret each permutation $\sigma \in \Sigma_r$ in terms of its action on $\mathcal{S}_r$, and writing out $\sigma^{-1}(k)$ for each $k \in \underline{r}$ shows how σ reorders the factors of each product in $\mathcal{S}_r$.

As shorthand, we write $\sigma^* \coloneqq x_{\sigma^{-1}(1)} \cdots x_{\sigma^{-1}(r)}$, so that $\rho \cdot \sigma^* = (\rho\sigma)^*$, which we may denote simply by $\rho\sigma^*$.

One special type of permutation is an *adjacent transposition*, which swaps consecutive numbers.

Definition B.6 (Adjacent transposition)**.** For $1 \leq i \leq r-1$, define $\tau_i \in \Sigma_r$ by

$$\tau_i = [1, \ldots, i-1, i+1, i, i+2, \ldots, r]$$

or more precisely by

$$\tau_i(k) \coloneqq \begin{cases} i+1 & \text{if } k = i; \\ i & \text{if } k = i+1; \\ k & \text{otherwise.} \end{cases}$$

Then τ_i is an *adjacent transposition*, and we write $T_r \coloneqq \{\tau_1, \ldots, \tau_{r-1}\}$.

The adjacent transposition τ_i acts on a product in $\mathcal{S}_r$ by swapping the i^{th} and $(i+1)^{\text{th}}$ variables.

We have the following standard presentation of the symmetric group with adjacent transpositions as generators.

Proposition B.7. *The symmetric group Σ_r is presented as a group by generators $T_r = \{\tau_1, \dots, \tau_{r-1}\}$, denoting adjacent transpositions, and relations R given by*

$$\begin{aligned} \tau_i^2 \, R \, \mathrm{id} \quad & \text{for } 1 \leq i \leq r-1, \\ \tau_i \tau_j \, R \, \tau_j \tau_i \quad & \text{for } 1 \leq i < i+2 \leq j \leq r-1, \\ \text{and} \quad \tau_i \tau_{i+1} \tau_i \, R \, \tau_{i+1} \tau_i \tau_{i+1} \quad & \text{for } 1 \leq i \leq r-2, \end{aligned}$$

where id *is the identity, so $\Sigma_r \cong F_{\mathrm{grp}} T_r / \langle\langle R \rangle\rangle$.*

By Lemma B.3, the symmetric group Σ_r is presented as a monoid by generators T_r and relations R as well: $\Sigma_r \cong F_{\mathrm{mon}} T_r / \langle R \rangle$ as monoids and thus also as groups.

For our purposes, we would like to consider not only adjacent transpositions that swap consecutive variables, but *block transpositions* that swap consecutive *substrings* of variables. There are more thorough reviews of the theory of block transpositions at [5] or [11], although our notation differs.

Definition B.8 (Block transposition)**.** For $0 \leq i < j < k \leq r$, define $\tau[i,j,k] \in \Sigma_r$ by

$$\tau[i,j,k] \coloneqq [1, \dots, i, j+1, \dots, k, i+1, \dots, j, k+1, \dots, r]$$

or more precisely by

$$\tau[i,j,k](\ell) \coloneqq \begin{cases} \ell + (k-j) & \text{if } i+1 \leq \ell \leq j; \\ \ell - (j-i) & \text{if } j+1 \leq \ell \leq k; \\ \ell & \text{otherwise.} \end{cases}$$

Then $\tau[i,j,k]$ is a *block transposition*, and we write

$$B_r \coloneqq \{\tau[i,j,k] \mid 0 \leq i < j < k \leq r\}.$$

Consider a block transposition $\tau[i, j, k] \in B_r$. The idea is that the *block* $[i+1, j]$ is swapped with the *block* $[j+1, k]$. More precisely, $\tau[i, j, k]$ acts on $\mathcal{S}_r$ as follows. Given $t \in \mathcal{S}_r$, write $t = t_1 t_2 t_3 t_4$ so that:

- t_1 consists of the first i variables of s;
- t_2 consists of the $(i+1)^{\text{th}}$ through j^{th} variables of s;
- t_3 consists of the $(j+1)^{\text{th}}$ through k^{th} variables of s; and
- t_4 consists of the last $(r-k)$ variables of s.

Then $\tau[i, j, k] \cdot s = t_1 t_3 t_2 t_4$, swapping subwords t_2 and t_3.

Note that every adjacent transposition is a block transposition: $\tau_i = \tau[i-1, i, i+1]$.

Block transpositions obey the following relations. First, inverses of block transpositions are themselves block transpositions that may be computed as follows (the proof is a straightforward computation).

Lemma B.9. *For $0 \leq i < j < k \leq r$, we have $\tau[i, j, k]^{-1} = \tau[i, i-j+k, k]$.*

The idea is that

$$\tau[i, j, k] \cdot (x_1 \cdots x_r) = x_1 \cdots x_i (x_{j+1} \cdots x_k)(x_{i+1} \cdots x_j) x_{k+1} \cdots x_r,$$

and since on the right hand side there are $(k-j)$ variables between the left pair of parentheses, to recover $x_1 \cdots x_r$ we must swap the $(i+1)^{\text{th}}$ through $(i-j+k)^{\text{th}}$ variables with the $(i-j+k+1)^{\text{th}}$ through k^{th} variables:

$$\tau[i, i-j+k, k] \cdot (x_1 \cdots x_i x_{j+1} \cdots x_k x_{i+1} \cdots x_j x_{k+1} \cdots x_r) = x_1 \cdots x_r.$$

Next, if the values not fixed by one block transposition are disjoint from the values not fixed by another, i.e. if two block transpositions are "non-colliding," then those two block transpositions should commute.

Lemma B.10. *Given $0 \leq i < j < k \leq i' < j' < k' \leq r$, we have*

$$\tau[i, j, k]\tau[i', j', k'] = \tau[i', j', k']\tau[i, j, k].$$

Either side of the above equation will send $x_1 \cdots x_r$ to

$$x_1 \cdots x_i(x_{j+1} \cdots x_k)(x_{i+1} \cdots x_j)x_{k+1} \cdots x_{i'}(x_{j'+1} \cdots x_{k'})(x_{i'+1} \cdots x_{j'})x_{k'+1} \cdots x_r.$$

There is another relation that we may depict as follows. After applying a block transposition $\tau[j,k,\ell]$ to $x_1 \cdots x_r$, we may then swap a block consisting of all the entries that were just swapped with an adjacent block. In particular, if $0 \le i < j < k < \ell \le r$, then

$$\begin{aligned} \tau[i,j,\ell]\tau[j,k,\ell] \cdot (x_1 \cdots x_r) &= \tau[i,j,\ell] \cdot (x_1 \cdots x_j x_{k+1} \cdots x_\ell x_{j+1} \cdots x_k x_{\ell+1} \cdots x_r) \\ &= x_1 \cdots x_i(x_{k+1} \cdots x_\ell x_{j+1} \cdots x_k)(x_{i+1} \cdots x_j)x_{\ell+1} \cdots x_r. \end{aligned}$$

But the same result could be obtained by applying $\tau[i,j,\ell]$ first, then applying another block transposition to swap $x_{j+1} \cdots x_k$ with $x_{k+1} \cdots x_\ell$ in their new positions. So in the product $\tau[i,j,\ell] \cdot (x_1 \cdots x_r)$, which is

$$x_1 \cdots x_i(x_{j+1} \cdots x_k x_{k+1} \cdots x_\ell)(x_{i+1} \cdots x_j)x_{\ell+1} \cdots x_r,$$

the variables $x_{j+1} \cdots x_k$ comprise the $(j+1-(j-i))^{\text{th}}$ (or simply $(i+1)^{\text{th}}$) through $(k-(j-i))^{\text{th}}$ positions, while the variables $x_{k+1} \cdots x_\ell$ comprise the $(k+1-(j-i))^{\text{th}}$ through $(\ell-(j-i))^{\text{th}}$ positions. We therefore have the following relation.

Lemma B.11. *Given $0 \le i < j < k < \ell \le r$, we have*

$$\tau[i,j,\ell]\tau[j,k,\ell] = \tau([j,k,\ell]-(j-i))\tau[i,j,\ell],$$

where we write $\tau([j,k,\ell]-(j-i)) \coloneqq \tau[j-(j-i), k-(j-i), \ell-(j-i)] = \tau[i, i-j+k, i-j+\ell]$.

Finally, transposing a block over two consecutive blocks can be written as a single block transposition. That is, given $0 \le i < j < k < \ell \le r$, applying the block transposition $\tau[i,j,\ell]$ to $x_1 \cdots x_r$ is the same as swapping $x_{i+1} \cdots x_j$ with $x_{j+1} \cdots x_k$ by applying $\tau[i,j,k]$; then swapping $x_{i+1} \cdots x_j$, now comprising the $(i-j+k+1)^{\text{th}}$ through k^{th} positions, with $x_{k+1} \cdots x_\ell$:

$$\begin{aligned} &\tau[i-j+k,k,\ell]\tau[i,j,k] \cdot (x_1 \cdots x_r) \\ &= \tau[i-j+k,k,\ell] \cdot (x_1 \cdots x_i(x_{j+1} \cdots x_k)(x_{i+1} \cdots x_j)(x_{k+1} \cdots x_\ell)x_{\ell+1} \cdots x_r) \\ &= x_1 \cdots x_i(x_{j+1} \cdots x_k)(x_{k+1} \cdots x_\ell)(x_{i+1} \cdots x_j)x_{\ell+1} \cdots x_r \\ &= \tau[i,j,\ell] \cdot (x_1 \cdots x_r). \end{aligned}$$

Lemma B.12. *Given $0 \leq i < j < k < \ell \leq r$, we have $\tau[i,j,\ell] = \tau[i-j+k,k,\ell]\tau[i,j,k]$.*

Corollary B.13. *Given $0 \leq i < j < k < \ell \leq r$, we have $\tau[i,k,\ell] = \tau[i,j,j-k+\ell]\tau[j,k,\ell]$.*

Proof. Replacing j with $i-j+k$ in Lemma B.12 yields $\tau[i,i-j+k,\ell] = \tau[j,k,\ell]\tau[i,i-j+k,k]$. Then inverting both sides using Lemma B.9 yields $\tau[i,j-k+\ell,\ell] = \tau[i,j,k]\tau[j,j-k+\ell,\ell]$. Finally, replacing k with $j-k+\ell$ yields the result. □

Corollary B.14. *Block transpositions may be written as products of adjacent transpositions as follows. Given $0 \leq i < j < k \leq r$, we have*

$$\tau[i,j,k] = (\tau_{i-j+k} \cdots \tau_{i+2}\tau_{i+1}) \cdots (\tau_{k-2} \cdots \tau_j \tau_{j-1})(\tau_{k-1} \cdots \tau_{j+1}\tau_j).$$

This product on the right hand side is best read from right to left. The idea is that to swap $x_{i+1} \cdots x_j$ with $x_{j+1} \cdots x_k$ in $x_1 \cdots x_r$, as $\tau[i,j,k]$ does, we can use adjacent transpositions to swap one variable at a time:

- apply $\tau_{k-1} \cdots \tau_{j+1}\tau_j$ to swap x_j with all of $x_{j+1} \cdots x_k$, one variable at a time;
- apply $\tau_{k-2} \cdots \tau_j \tau_{j-1}$ to swap x_{j-1} with all of $x_{j+1} \cdots x_k$ (now in the j^{th} through $(k-1)^{\text{th}}$ positions), one variable at a time;
- …
- apply $\tau_{i-j+k} \cdots \tau_{i+2}\tau_{i+1}$ to swap x_{i+1} with all of $x_{j+1} \cdots x_k$ (now in the $(i+2)^{\text{th}}$ through $(i-j+k+1)^{\text{th}}$ positions), one variable at a time.

Proof of Corollary B.14. We induct on $k-i$. Our base case is $k-i=2$, for which we have $\tau[i,i+1,i+2] = \tau_{i+1}$. Then for $k-i>2$, if $j=i+1$, by Lemma B.12 with $k-1$ in place of k and k in place of ℓ, we have

$$\tau[i,i+1,k] = \tau[k-2,k-1,k]\tau[i,i+1,k-1] = \tau_{k-1}(\tau_{k-2} \cdots \tau_{i+2}\tau_{i+1})$$

by induction. Otherwise, $i+1<j$, so by Corollary B.13 with $i+1$ in place of j, j in place of k, and k in place of ℓ, we have

$$\begin{aligned}\tau[i,j,k] &= \tau[i,i+1,i-j+k+1]\tau[i+1,j,k] \\ &= \tau_{i-j+k} \cdots \tau_{i+2}\tau_{i+1}(\tau_{i-j+k+1} \cdots \tau_{i+3}\tau_{i+2}) \cdots (\tau_{k-2} \cdots \tau_j\tau_{j-1})(\tau_{k-1} \cdots \tau_{j+1}\tau_j)\end{aligned}$$

by induction. □

It turns out these are all the relations we need to present Σ_r by generators B_r.

Proposition B.15. *The symmetric group Σ_r is presented as a group by generators*

$$B_r = \{\tau[i,j,k] \mid 0 \leq i < j < k \leq r\},$$

denoting block transpositions, and relations R' given by

(1) $\tau[i, i-j+k, k]\tau[i,j,k] \; R' \; \mathrm{id}$ *for* $0 \leq i < j < k \leq r$*,*

(2) $\tau[i,j,k]\tau[i',j',k'] \; R' \; \tau[i',j',k']\tau[i,j,k]$ *for* $0 \leq i < j < k \leq i' < j' < k' \leq r$*,*

(3) $\tau[i,j,\ell]\tau[j,k,\ell] \; R' \; \tau([j,k,\ell]-(j-i))\tau[i,j,\ell]$ *for* $0 \leq i' < j' \leq i < j < k \leq k' \leq r$*, where* $\tau([j,k,\ell]-(j-i)) \coloneqq \tau[j-(j-i), k-(j-i), \ell-(j-i)]$*,*

(4) $\tau[i,j,\ell] \; R' \; \tau[i-j+k,k,\ell]\tau[i,j,k]$ *for* $0 \leq i < j < k < \ell \leq r$*,*

where id *is the identity, so* $\Sigma_r \cong F_{\mathrm{grp}}B_r/\langle\langle R'\rangle\rangle$*.*

By Lemma B.3, the symmetric group Σ_r is presented as a monoid by generators B_r and relations R' as well: $\Sigma_r \cong F_{\mathrm{mon}}B_r/\langle R'\rangle$ as monoids and thus also as groups.

Proof. Our previous work exhibits a group homomorphism $\varphi\colon F_{\mathrm{grp}}B_r/\langle\langle R'\rangle\rangle \to \Sigma_r$ sending each block transposition to itself as a permutation, and by Proposition B.7 there is an isomorphism $\kappa\colon \Sigma_r \to F_{\mathrm{grp}}T_r/\langle\langle R\rangle\rangle$ sending the permutation corresponding to each τ_i to τ_i. We now construct a group homomorphism $\psi\colon F_{\mathrm{grp}}T_r/\langle\langle R\rangle\rangle \to F_{\mathrm{grp}}B_r/\langle\langle R'\rangle\rangle$ as follows. Define a function $T_r \to F_{\mathrm{grp}}B_r/\langle\langle R'\rangle\rangle$ sending $\tau_i \mapsto \tau[i-1,i,i+1]$, inducing a group homomorphism $\psi'\colon F_{\mathrm{grp}}T_r \to F_{\mathrm{grp}}B_r/\langle\langle R'\rangle\rangle$. To show that ψ' descends to ψ, it suffices to show that ψ' agrees on R-related elements of $F_{\mathrm{grp}}T_r$, as follows.

- For $1 \leq j \leq r-1$, we have $\tau_j^2 \, R \, \mathrm{id}$, and $\psi'(\tau_j^2) = (\tau[j-1,j,j+1])^2 \; R' \; \mathrm{id}$ by (1), with $i = j-1$ and $k = j+1$.

- For $1 \leq i < i+2 \leq j \leq r-1$, we have $\tau_i\tau_j \; R \; \tau_j\tau_i$, and by (2),

$$\begin{aligned} \psi'(\tau_i\tau_j) &= \tau[i-1,i,i+1]\tau[j-1,j,j+1] \\ &R' \; \tau[j-1,j,j+1]\tau[i-1,i,i+1] = \psi'(\tau_j\tau_i). \end{aligned}$$

- For $1 \leq i \leq r-2$, we have $\tau_i\tau_{i+1}\tau_i \; R \; \tau_{i+1}\tau_i\tau_{i+1}$, and by (3) and (4),

$$\begin{aligned} \psi'(\tau_i\tau_{i+1}\tau_i) &= \tau[i-1,i,i+1](\tau[i,i+1,i+2]\tau[i-1,i,i+1]) \\ &\langle\langle R' \rangle\rangle \; \tau[i-1,i,i+1]\tau[i-1,i,i+2] && (4) \\ &\langle\langle R' \rangle\rangle \; \tau[i-1,i,i+2]\tau[i,i+1,i+2] && (3) \\ &\langle\langle R' \rangle\rangle \; (\tau[i,i+1,i+2]\tau[i-1,i,i+1])\tau[i,i+1,i+2] && (4) \\ &= \psi'(\tau_{i+1}\tau_i\tau_{i+1}). \end{aligned}$$

It remains to show that $\kappa \circ \varphi$ and ψ are mutually inverse. We have that $\kappa \circ \varphi \circ \psi$ sends

$$\tau_i \mapsto \tau[i-1,i,i+1] \mapsto [1,\ldots,i-1,i+1,i,i+2,\ldots,r] \mapsto \tau_i$$

and is therefore the identity. Conversely, Corollary B.14 shows how $\kappa \circ \varphi$ sends $\tau[i,j,k]$ to a product of adjacent transpositions, which ψ then sends to the corresponding product of block transpositions that is $\langle\langle R' \rangle\rangle$-related to $\tau[i,j,k]$, as the proof of Corollary B.14 relies only on the R'-relations (1) and (4) above. Hence $\psi \circ \kappa \circ \varphi$ is the identity as well. $\square$

Here we compile more coherence schemata to which we may apply the techniques throughout this thesis.

BIBLIOGRAPHY

[1] Awodey, S. *Category Theory.* Oxford University Press, 2006.

[2] Blumberg, A. J., and Hill, M. A. Operadic multiplications in equivariant spectra, norms, and transfers. *Advances in Mathematics 285* (2015), 658–708.

[3] Elmendorf, A. D., and Mandell, M. A. Rings, modules, and algebras in infinite loop space theory. *Advances in Mathematics 205*, 1 (2006), 163–228.

[4] Enderton, H. B. *A Mathematical Introduction to Logic.* Elsevier, 2001.

[5] Eriksson, H., Karlander, J., Svensson, L., and Wästlund, J. Sorting a bridge hand. *Discrete Mathematics 241* (10 2001), 289–300.

[6] Guillou, B. J., May, J. P., Merling, M., and Osorno, A. M. Symmetric monoidal G-categories and their strictification. *The Quarterly Journal of Mathematics 71*, 1 (12 2019), 207–246.

[7] Hill, M. A., Hopkins, M. J., and Ravenel, D. C. On the nonexistence of elements of kervaire invariant one. *Annals of Mathematics* (2016), 1–262.

[8] Huet, G. Confluent reductions: Abstract properties and applications to term rewriting systems: Abstract properties and applications to term rewriting systems. *J. ACM 27*, 4 (1980), 797–821.

[9] Johnson, N., and Yau, D. *Bimonoidal Categories, E_n-Monoidal Categories, and Algebraic K-Theory*, vol. 283–285 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2024. 3 volumes.

[10] Kelly, G. M. On maclane's conditions for coherence of natural associativities, commutativities, etc. *Journal of Algebra 1*, 4 (1964), 397–402.

[11] Korchmaros, A. *Combinatorial Properties of Block Transpositions in Symmetric Groups.* PhD thesis, Università degli Studi della Basilicata, March 2015.

[12] Laplaza, M. L. Coherence for associativity not an isomorphism. *Journal of Pure and Applied Algebra 2*, 2 (1972), 107–120.

[13] Laplaza, M. L. Coherence for distributivity. In *Coherence in Categories* (Berlin, Heidelberg, 1972), G. M. Kelly, M. Laplaza, G. Lewis, and S. Mac Lane, Eds., Springer Berlin Heidelberg, pp. 29–65.

[14] Leinster, T. *Higher Operads, Higher Categories*. London Mathematical Society Lecture Note Series. Cambridge University Press, 2004.

[15] Mac Lane, S. Natural associativity and commutativity. *Rice University Studies 49*, 4 (1963), 28–46.

[16] Mac Lane, S. *Categories for the Working Mathematician*. Springer-Verlag, New York, 1971. Graduate Texts in Mathematics, Vol. 5.

[17] May, J. P. *The Geometry of Iterated Loop Spaces*, vol. 271 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1972.

[18] May, J. P. E_∞ spaces, group completions, and permutative categories. In *New Developments in Topology (Proc. Sympos. Algebraic Topology, Oxford, 1977)* (1977), Cambridge University Press, pp. 61–93.

[19] May, J. P. Multiplicative infinite loop space theory. *Journal of Pure and Applied Algebra 26*, 1 (1982), 1–69.

[20] May, J. P. The construction of E_∞ ring spaces from bipermutative categories. *New topological contexts for Galois theory and algebraic geometry (BIRS 2008) 16* (2009), 283–330.

[21] May, J. P. What precisely are E_∞ ring spaces and E_∞ ring spectra? *Geometry & Topology Monographs 16* (2009), 215–282.

[22] Mimram, S. Rewriting techniques for relative coherence, 2024.

[23] Newman, M. H. A. On theories with a combinatorial definition of 'equivalence'. *Annals of mathematics 43*, 2 (1942), 223–243.

[24] Rubin, J. Normed symmetric monoidal categories. *Journal of Homotopy and Related Structures 20* (Feb. 2025), 195–250.

[25] Segal, G. Categories and cohomology theories. *Topology 13*, 3 (1974), 293–312.

[26] Yau, D. Multifunctorial equivariant algebraic k-theory, 2024.