Coherence for Monoidal Categories with Duals

Ebrahim Ebrahim

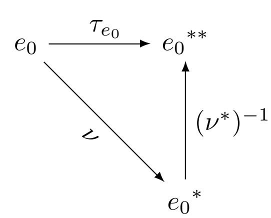
November 3, 2014

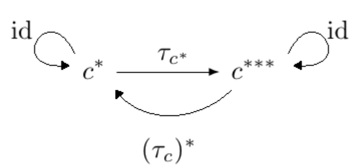
This document emerged from the notes I took while reading about dual structures on monoidal categories. Dual structures are often described in terms of *rigidity*, and it is often desired that rigid monoidal categories be at least *pivotal*. Under these circumstances we get relations like $(A \otimes B)^* \cong B^* \otimes A^*$, $A^{**} \cong A$, and $1^* \cong 1$. A general monoidal category with these relations will be called a monoidal category *with duals*. We aim to state and prove a coherence theorem for monoidal categories with duals. The objective is to make the proof clear, perhaps at the expense of efficiency.

The reader is assumed to be familiar with categories, functors, and natural transformations. Some familiarity with monoidal categories is assumed as well, although Maclane's coherence theorem for monoidal categories is subsumed within the treatment here. For a category $\mathcal{C}$ we will denote the collection of objects by $\mathcal{C}^0$ and the collection of arrows by $\mathcal{C}^1$.

1 Coherence Conditions

A *monoidal category with duals*, or MCD, is a decuple $(\mathcal{C}, \otimes, *, e_0, \alpha, \rho, \lambda, \gamma, \tau, \nu)$ where $(\mathcal{C}, \otimes, e_0, \alpha, \rho, \lambda)$ is a (relaxed) monoidal category, $*$ is a contravariant functor $\mathcal{C} \to \mathcal{C}$ (i.e. a functor $\mathcal{C}^{\text{op}} \to \mathcal{C}$), $\tau_A : A \to A^{**}$ and $\gamma_{A,B} : A^* \otimes B^* \to (B \otimes A)^*$ are natural isomorphims, $\nu : e_0 \to {e_0}^*$ is an isomorphism, and the following six diagrams commute for arbitrary objects a, b, c:





$$
\begin{array}{ccc}
a^* \otimes e_0 & \xrightarrow{1_{a^*} \otimes \nu} & a^* \otimes {e_0}^* \\
\downarrow{\scriptstyle \rho_{a^*}} & & \downarrow{\scriptstyle \gamma_{a,e_0}} \\
a^* & \xrightarrow[(\lambda_a)^*]{} & (e_0 \otimes a)^*
\end{array}
\qquad\qquad
\begin{array}{ccc}
e_0 \otimes a^* & \xrightarrow{\nu \otimes 1_{a*}} & {e_0}^* \otimes a^* \\
\downarrow{\scriptstyle \lambda_{a^*}} & & \downarrow{\scriptstyle \gamma_{e_0,a}} \\
a^* & \xrightarrow[(\rho_a)^*]{} & (a \otimes e_0)^*
\end{array}
$$

$$
\begin{array}{ccccc}
(c^* \otimes b^*) \otimes a^* & \xrightarrow{\gamma_{c,b} \otimes 1_{a^*}} & (b \otimes c)^* \otimes a^* & \xrightarrow{\gamma_{b\otimes c,a}} & (a \otimes (b \otimes c))^* \\
{\scriptstyle \alpha_{c^*,b^*,a^*}}\downarrow & & & & \downarrow{\scriptstyle (\alpha_{a,b,c})^*} \\
c^* \otimes (b^* \otimes a^*) & \xrightarrow[1_{c^*} \otimes \gamma_{b,a}]{} & c^* \otimes (a \otimes b)^* & \xrightarrow[\gamma_{c,a\otimes b}]{} & ((a \otimes b) \otimes c))^*
\end{array}
$$

$$
\begin{array}{ccc}
a \otimes b & \xrightarrow{\tau_{a\otimes b}} & (a \otimes b)^{**} \\
{\scriptstyle \tau_a \otimes \tau_b}\downarrow & & \downarrow{\scriptstyle (\gamma_{b,a})^*} \\
a^{**} \otimes b^{**} & \xrightarrow[\gamma_{a^*,b^*}]{} & (b^* \otimes a^*)^*
\end{array}
$$

These diagrams are referred to as the *coherence conditions*, since their commutativity is necessary and sufficient for *coherence*. Coherence is roughly the property that the diagrams in a category that really ought to commute do in fact commute. We will give a precise definition of coherence in MCDs and a proof that it is a consequence of the coherence conditions. It becomes clear why the coherence conditions take the form that they do in the proof of lemma 6 below.

The statement of the definition above views $\nu : e_0 \to {e_0}^*$ as an isomorphism in $\mathcal{C}$, but it will sometimes be convenient to view it as a natural isomorphism $\nu : e \to e^{\hat{*}}$, where $e : \circlearrowleft \to \mathcal{C}$ is the functor from the one-point category that picks out the identity object e_0, and $e^{\hat{*}}$ is the functor that picks out ${e_0}^*$

2 Motivation for Upcoming Definitions

Consider an algebraic structure consisting of a set with an associative multiplication, an identity, and a dual operation that agrees with the multiplication. An expression written in terms of elements of the algebraic structure can be manipulated by applying the various properties of the multiplication and the dual. One may ask of any given pair of expressions, “Can we use the axioms to show they are equal?” Often there are several different sequences of applications of the properties that take one given expression to another.

A MCD is a categorification of this algebraic structure, where the equations that describe its properties are weakened to natural isomorphisms. A new question arises: "If we can use the properties to show that two given algebraic expressions are isomorphic, then can *different* applications of the properties yield different isomorphisms?" Coherence is what obtains when the answer is *no*. A MCD is *coherent* if every sequence of "moves" from one expression to another that involve applying $\alpha, \rho, \lambda, \gamma, \tau, \nu$ and their inverses yields the same isomorphism.

3 Iterated Functors and Iterated Natural Transformations

We will build these sequences of "moves" explicitly. Iterated functors will play the role of "expressions" from above, and we will recursively construct the sorts of natural transformations that count as "moves" on expressions. Then we will discuss coherence from this perspective.

Given a MCD $\mathcal{C}$ we define a new category $\mathbf{Fct}(\mathcal{C})$ with objects being functors from products of $\mathcal{C}$ and $\mathcal{C}^{\text{op}}$ to $\mathcal{C}$ and with morphisms being all the natural transformations between them. We can write such products of copies of $\mathcal{C}$ as $\mathcal{C}^\alpha$, where α is a finite sequence in $\{-1, 1\}$ (each -1 denotes a factor of $\mathcal{C}^{\text{op}}$). $\alpha + \beta$ will denote concatenation of sequences and $(-\alpha)$ denotes swapping -1 and 1. The length of the sequence α is the *multiplicity* of a functor $F : \mathcal{C}^\alpha \to \mathcal{C}$. For example $\otimes$ and $*$ are themselves objects of $\mathbf{Fct}(\mathcal{C})$, of multiplicity 2 and 1 respectively. Multiplicity 0 is also allowed, and in this case the domain would be the one point category $\circlearrowleft$ (an example is e). Now $\mathbf{Fct}(\mathcal{C})$ can itself be given the structure of a MCD: For two functors $F : \mathcal{C}^\alpha \to \mathcal{C}, G : \mathcal{C}^\beta \to \mathcal{C}$ we define $F \hat{\otimes} G : \mathcal{C}^{\alpha+\beta} \to \mathcal{C}$ to be the composite:

$$\mathcal{C}^{\alpha+\beta} \cong \mathcal{C}^\alpha \times \mathcal{C}^\beta \xrightarrow{F \times G} \mathcal{C} \times \mathcal{C} \xrightarrow{\otimes} \mathcal{C}$$

which has multiplicity $n + m$. For natural transformations $\eta : F \to G$, $\theta : F' \to G'$ define $\eta \hat{\otimes} \theta : F \hat{\otimes} F' \to G \hat{\otimes} G'$ to be the natural transformation with components given by:

$$(\eta \hat{\otimes} \theta)_{(A,B)} = \eta_A \otimes \theta_B$$

for $A \in \text{dom}(F)$, $B \in \text{dom}(F')$. Similarly define $\hat{*}$ by:

$$F^{\hat{*}} \text{ is } \mathcal{C}^{-\alpha} = (\mathcal{C}^\alpha)^{\text{op}} \xrightarrow{F^{\text{op}}} \mathcal{C}^{\text{op}} \xrightarrow{*} \mathcal{C}$$

$$\eta^{\hat{*}} : G^{\hat{*}} \to F^{\hat{*}} \text{ has components } (\eta^{\hat{*}})_A = (\eta_A)^* \text{ for } A \in \text{dom}(G)$$

for any functor $F : \mathcal{C}^\alpha \to \mathcal{C}$ and natural transformation $\eta : F \to G$. It is easy to check that $\hat{\otimes}$ and $\hat{*}$ are functorial. Define $\hat{e} : \circlearrowleft \to \mathbf{Fct}(\mathcal{C})$ to be the functor that picks out e. For $F : \mathcal{C}^\alpha \to \mathcal{C}$, $G : \mathcal{C}^\beta \to \mathcal{C}$, $H : \mathcal{C}^\gamma \to \mathcal{C}$ objects in $\mathbf{Fct}(\mathcal{C})$ define $\hat{\alpha}_{F,G,H} : (F \hat{\otimes} G) \hat{\otimes} H \to F \hat{\otimes} (G \hat{\otimes} H)$ by:

$$(\hat{\alpha}_{F,G,H})_{A,B,C} = \alpha_{F(A),G(B),H(C)}$$

for any $A \in \mathcal{C}^\alpha, B \in \mathcal{C}^\beta, C \in \mathcal{C}^\gamma$. Similarly define:

$$(\hat{\rho}_F)_A = \rho_{F(A)}$$
$$(\hat{\lambda}_F)_A = \rho_{F(A)}$$
$$(\hat{\gamma}_{F,G})_{A,B} = \gamma_{F(A),G(B)}$$
$$(\hat{\tau}_F)_A = \tau_{F(A)}$$

It is easy to check that these are natural isomorphisms (for example $\hat{\alpha}$ is a natural transformation $(1_{\mathbf{Fct}(\mathcal{C})} \hat{\hat{\otimes}} 1_{\mathbf{Fct}(\mathcal{C})}) \hat{\hat{\otimes}} 1_{\mathbf{Fct}(\mathcal{C})} \longrightarrow 1_{\mathbf{Fct}(\mathcal{C})} \hat{\hat{\otimes}} (1_{\mathbf{Fct}(\mathcal{C})} \hat{\hat{\otimes}} 1_{\mathbf{Fct}(\mathcal{C})})$). Also define $\hat{\nu} : \hat{e} \to \hat{e}^{\hat{\hat{*}}}$ by $\hat{\nu}_{(\cdot)} = \nu : e \to e^{\hat{*}}$. The coherence conditions are easily checked, and we see that we have a MCD, $(\mathbf{Fct}(\mathcal{C}), \hat{\otimes}, \hat{*}, \hat{e}, \hat{\alpha}, \hat{\rho}, \hat{\lambda}, \hat{\gamma}, \hat{\tau}, \hat{\nu})$.

A certain collection of objects in $\mathbf{Fct}\,(\mathcal{C})$ will play the role of expressions to be manipulated, and a certain collection of morphisms in $\mathbf{Fct}\,(\mathcal{C})$ will provide the allowed manipulations. We start with some definitions to efficiently describe it:

Definition 1: Let $\mathcal{C}$ be a category, and let $S \subset \mathbf{Fct}\,(\mathcal{C})^1$ be a collection of natural transformations. We define $S_\blacksquare \subset \mathcal{C}^1$ to be the collection of all components of the natural transformations in S. The elements of $S_\blacksquare$ are called *instances* of S.

Definition 2: Let $\mathcal{C}$ be a MCD, and let $S \subset \mathcal{C}^1$ be a collection of morphisms in $\mathcal{C}$.
A *reduced expansion of S of depth* 0 is an element of $\llbracket S \rrbracket_0 := \{1_A \mid A \in \mathcal{C}^0\} \cup S$.
A *reduced expansion of S of depth* $n+1$ is an element of $\llbracket S \rrbracket_{n+1} :=$

$$\{\beta \otimes 1_A \mid A \in \mathcal{C}^0 \text{ and } \beta \in \llbracket S \rrbracket_n\} \ \cup \ \{1_A \otimes \beta \mid A \in \mathcal{C}^0 \text{ and } \beta \in \llbracket S \rrbracket_n\}$$

An *expansion of S of depth* 0 is an element of $\llbracket S \rrbracket^{*}_0$, which we define to be the closure of $\llbracket S \rrbracket_0$ under $*$.
An *expansion of S of depth* $n+1$ is an element of $\llbracket S \rrbracket^{*}_{n+1}$, which we define to be the closure of

$$\{\beta \otimes 1_A \mid A \in \mathcal{C}^0 \text{ and } \beta \in \llbracket S \rrbracket^{*}_n\} \ \cup \ \{1_A \otimes \beta \mid A \in \mathcal{C}^0 \text{ and } \beta \in \llbracket S \rrbracket^{*}_n\}$$

under $*$.
A *reduced expansion of S* is an element of $\llbracket S \rrbracket := \bigcup_n \llbracket S \rrbracket_n$.
An *expansion of S* is an element of $\llbracket S \rrbracket^{*} := \bigcup_n \llbracket S \rrbracket^{*}_n$.

Definition 3: Let $\mathcal{C}$ be a category, and let S be a collection of morphisms in $\mathcal{C}$. We define $\llbracket S \rrbracket^{\circ} \subset \mathcal{C}^1$ to be the collection of morphsisms in the wide subcategory of $\mathcal{C}$ generated by S. It consists of all identities in $\mathcal{C}$, and all composites of elements of S. The elements of $\llbracket S \rrbracket^{\circ}$ are called *iterates* of S.

Finally we may define the "expressions" and the "moves" allowed on those expressions. The idea is that we start with a set of natural transformations S that encodes elementary moves that can be made on the outermost part of an expression, and then we extend to elementary moves made "inside" an expression by taking expansions of S.

Definition 4: Let $\mathcal{C}$ be a MCD. An *iterated functor* is an element of the smallest subcollection of $\mathbf{Fct}\,(\mathcal{C})^0$ that contains $\{e, 1_\mathcal{C}\}$ and is closed under $\hat{\otimes}$ and $\hat{*}$. An iterated natural transformation is an element of:

$$\llbracket \llbracket \{\hat{\alpha}, \hat{\rho}, \hat{\lambda}, \hat{\gamma}, \hat{\tau}, \hat{\nu}, \hat{\alpha}^{-1}, \hat{\rho}^{-1}, \hat{\lambda}^{-1}, \hat{\gamma}^{-1}, \hat{\tau}^{-1}, \hat{\nu}^{-1}\}_\blacksquare \rrbracket^{*} \rrbracket^{\circ}$$

We are now ready to give one definition of coherence:

Coherence (iterated functor definition): A MCD $\mathcal{C}$ is *coherent* if any two iterated natural transformations between the same two iterated functors are equal. In other words, coherence occurs when any diagram with iterated functors for vertices and expansions of instances of $\hat{\alpha}, \hat{\rho}, \hat{\lambda}, \hat{\gamma}, \hat{\tau}, \hat{\nu}$ (and their inverses) for edges commutes.

One issue that we need to worry about is "vertex collapse." Consider a diagram with iterated functors for vertices, and expansions of instances as above for edges. One vertex might be $(1_\mathcal{C} \hat{\otimes} 1_\mathcal{C}^{\hat{*}}) \hat{\otimes} 1_\mathcal{C}$, while another might be $(e \hat{\otimes} 1_\mathcal{C}) \hat{\otimes} (1_\mathcal{C}^{\hat{*}} \hat{\otimes} 1_\mathcal{C})$. When thinking about coherence we usually imagine these to be distinct vertices, but in the context of a specific MCD they could wind up being equal! When this happens it seems possible that it could spoil the commutativity of the diagram (I like to think of the equality of vertices as providing an unnatural "portal" through which one can teleport while considering paths in the diagram). It could be that iterated functors were constructed carefully enough for this not to be an issue, but I have yet to find a convincing argument for this. The next construction has the same spirit as the iterated functors definition, but it ensures in a very direct way that we do not get this vertex collapse. For this reason we pursue a proof of coherence in terms of the language of the following section.

4 The Free Monoidal Category with Duals

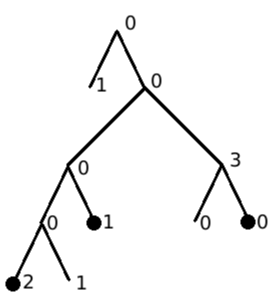
Define a (strict) morphism of MCDs $(\mathcal{C}, \otimes, *, e, \alpha, \rho, \lambda, \gamma, \tau, \nu)$ and $(\mathcal{C}', \otimes', *', e', \alpha', \rho', \lambda', \gamma', \tau', \nu')$ to be a functor $F : \mathcal{C} \to \mathcal{C}'$ such that $F(A \otimes B) = F(A) \otimes' F(B)$, $F(A^*) = F(A)^{*'}$, $F \circ e = e'$, and:

$$F(f \otimes g) = F(f) \otimes' F(g) \qquad F(f^*) = F(f)^{*'} \qquad F(\alpha_{A,B,C}) = \alpha'_{F(A),F(B),F(C)}$$
$$F(\gamma_{A,B}) = \gamma'_{F(A),F(B)} \qquad F(\rho_A) = \rho'_{F(A)} \qquad F(\lambda_A) = \lambda'_{F(A)}$$
$$F(\tau_A) = \tau'_{F(A)} \qquad F(\nu_{(\cdot)}) = \nu'_{(\cdot)} \tag{1}$$

With respect to this notion of morphism between MCDs we will explicitly construct the free MCD on one object.

4.1 Explicit Construction

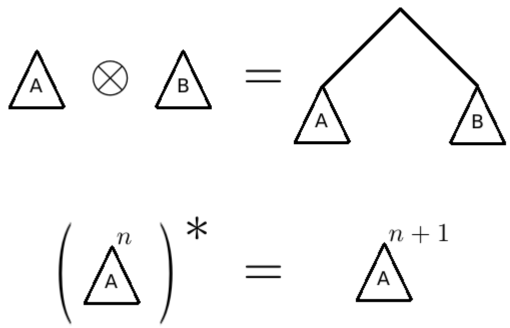
Let us define $\mathbf{ɯ}[x]$, the free MCD based on $\{x\}$. The objects will be called *formal expressions*. They are finite proper binary trees (trees with all nodes having 0 or 2 children), with each node bearing a weight in $\mathbb{N}$ and with each leaf bearing a label in $\{e, x\}$. Here is a graphic representation:



The leaves with a black circle are labelled by x, while the others are labelled by e.
This formal expression can be thought of as:

" $e^*(((x^{**}e^*)x^*)(ex)^{***})$ "

Define $\otimes$ and $*$ on formal expressions by:



Define the ground object e_0 to be the single vertex labeled by e with weight 0. Similarly define x to be the single vertex labeled by x with weight 0.

Arrows between formal expressions will be certain ordered pairs of formal expressions, with the components being the domain and codomain (so by construction there can be at most one arrow between any two formal expressions). Let's start with a much larger category where all pairs are included as arrows (i.e. the complete directed graph with formal expressions as vertices). Call this category $\overline{\mathbf{ш}}[x]$. Define $\otimes$, $*$, and $\circ$ on its arrows by:

$$(A,B)\otimes(C,D)=(A\otimes C,B\otimes D)\quad\text{and}\quad(A,B)^*=(B^*,A^*)\quad\text{and}\quad(B,C)\circ(A,B)=(A,C)$$

For formal expressions A, B, C define arrows:

$$\alpha_{A,B,C}=((A\otimes B)\otimes C,A\otimes(B\otimes C))\qquad\rho_A=(A\otimes e_0,A)\qquad\lambda_A=(e_0\otimes A,A)$$
$$\gamma_{A,B}=((A^*\otimes B^*),(B\otimes A)^*)\qquad\tau_A=(A,A^{**})\qquad\nu=(e_0,{e_0}^*)$$

For $(\overline{\mathbf{ш}}[x],\otimes,*,e_0,\alpha,\rho,\lambda,\gamma,\tau,\nu)$ to be a MCD, we need closure of arrows under $\otimes$ and $*$, naturality of $\alpha,\rho,\lambda,\gamma,\tau,\nu$, functoriality of $\otimes,*,e$, and the coherence conditions. All of these hold trivially, including coherence conditions (all diagrams in this category commute by construction).

Recall that $\{\alpha,\rho,\lambda,\gamma,\tau,\nu\}_{\blacksquare}$ is the collection of all components of the natural transformations α, ρ, λ, γ, τ, and ν. Here let us use the word *expansion* to refer to elements of $\{\!\!\{\{\alpha,\rho,\lambda,\gamma,\tau,\nu\}_{\blacksquare}\}\!\!\}$. The collection of arrows of $\mathbf{ш}[x]$ shall be the closure of the collection of expansions under composition and inverse:

$$\mathbf{ш}[x]^1:=\{(A,B)\mid\text{there is a path of expansions from }A\text{ to }B\text{, not necessarily directed}\}$$

In other words it is:

$$\mathbf{ш}[x]^1:=\lbrace\!\!\lbrace\{\!\!\{\{\alpha,\rho,\lambda,\gamma,\tau,\nu,\alpha^{-1},\rho^{-1},\lambda^{-1},\gamma^{-1},\tau^{-1},\nu^{-1}\}_{\blacksquare}\}\!\!\}\rbrace\!\!\rbrace$$

For $(\mathbf{ш}[x],\otimes,*,e_0,\alpha,\rho,\lambda,\gamma,\tau,\nu)$ to be a MCD, we still need to verify closure of arrows under $\otimes$ and $*$. Closure of arrows under $*$ follows from functoriality of $*$ and the fact that expansions are closed under $*$. Here is an argument for closure of arrows under $\otimes$:

First notice that if $(A,B),(C,D)$ are expansions then

$$(A,B)\otimes(C,D)=(AC,BD)=(AD,BD)\circ(AC,AD)=((A,B)\otimes 1_D)\circ(1_A\otimes(C,D))$$

is a composite of expansions, and so is an arrow in $\mathbf{ш}[x]$. If $(A,B),(C,D)$ are arrows, then there is a path of expansions from A to B and from C to D. By appending sufficiently many identities to the shorter path we may take them to be of the same length. Then by $\otimes$-ing the individual expansions in each path, we obtain a path of arrows from $A \otimes C$ to $B \otimes D$. And so $\mathbf{ш}[x]$ is indeed a MCD.

Observe that the set $\mathbf{ш}[x]^0$ of formal expressions can be generated by starting with $\{x, e\}$ and iteratively closing under $*$ and $\otimes$:

Proposition 5: Start with the set E_0 of formal expressions of the form $x^{*\cdots*}$ or ${e_0}^{*\cdots*}$. Let E_{n+1} be the set of formal expressions of the form $(a \otimes b)^{*\cdots*}$ for $a, b \in E_n$ (i.e. E_{n+1} is the result of applying " $\overset{m}{\Delta\Delta}$ " to expressions of E_n). Then $\mathbf{ш}[x]^0 = \bigcup_n E_n$.

4.2 Is it really free?

Showing that the construction we gave above actually produces a free MCD is the bulk of the work when it comes to proving coherence. The following lemma makes the task much easier, and it also explains the six coherence conditions on the first page.

Lemma 6: Suppose $\mathcal{C}$ is a MCD. Let $S = \{\alpha, \rho, \lambda, \gamma, \tau, \nu, \alpha^{-1}, \rho^{-1}, \lambda^{-1}, \gamma^{-1}, \tau^{-1}, \nu^{-1}\}$. Then:

$$\{\!\{ \ast\!\!\ast\, S_{\blacksquare} \ast\!\!\ast \}\!\} = \{\!\{ [\![S_{\blacksquare}]\!] \}\!\}$$

Proof: We must prove "$\subseteq$." It suffices to show that the second collection is closed under $*$ (to see this look back at the recursive definition of expansions). By functoriality of $*$ we need only show that the $*$ of a reduced expansion of an instance of S is a composite of reduced expansions of instances of S. For reduced expansions of depth 1: Identity maps are a trivial case since $(1_A)^* = 1_{A^*}$, and instances of S are handled by the six coherence conditions. Each of the six coherence conditions gives us a way to write the $*$ of an instance of S as a composite of reduced expansions of instances of S!

We get the result at higher depths by induction: If $\beta : A \to B$ is a reduced expansion of an instance of S such that β^* is a composite of reduced expansions of instances of S, then can we say the same of $1_C \otimes \beta$ or $\beta \otimes 1_C$? The two cases are similar, and naturality of γ answers in the affirmative:

$$\begin{array}{ccc} C^* \otimes B^* & \xrightarrow{1_{C^*} \otimes \beta^*} & C^* \otimes A^* \\ \downarrow{\scriptstyle \gamma_{C,B}} & & \downarrow{\scriptstyle \gamma_{C,A}} \\ (B \otimes C)^* & \xrightarrow{(\beta \otimes 1_C)^*} & (A \otimes C)^* \end{array}$$

□

The proof of freeness will involve showing that a formal expression can be "simplified" uniquely. The definitions that follow will help guide this simplification.

Definition 7: A formal expression is *reduced* if the only nodes with nonzero weight are leaves labelled x (i.e. if its stars are "pushed all the way in").

Definition 8: A formal expression is *trimmed* if it is reduced and has no leaves labelled e unless it is just one vertex (i.e. all extra copies of the unit have been removed).

Definition 9: A formal expression is *standard* if it is trimmed and the right-hand child of any non-leaf node is a leaf (i.e. all parentheses are in the front).

Lemma 10: Reduced, trimmed, and standard expressions are generated as follows:
• Reduced formal expressions are the smallest collection of formal expressions that (1) contains $\{e_0, x, x^*\}$ and (2) contains $a \otimes b$ whenever it contains a and b.
• Trimmed formal expressions are the smallest collection of formal expressions that (1) contains $\{e_0, x, x^*\}$ and (2) contains $a \otimes b$ whenever it contains a and b with $a, b \neq e_0$.
• Standard formal expressions are the smallest collection of formal expressions that (1) contains $\{e_0, x, x^*\}$ and (2) contains $a \otimes x$ and $a \otimes x^*$ whenever it contains a with $a \neq e_0$.

Proof: Easy. □

Theorem 11: For any MCD $\mathcal{C}$ and any object c in $\mathcal{C}$, there is a unique morphism of MCDs, $F : \mathbf{m}[x] \to \mathcal{C}$, sending x to c.

Proof: The behavior of F on the set of objects $\mathbf{m}[x]^0$ is fully determined by the properties:

$$F(x) = c \qquad F(e_0) = \overline{e_0} \qquad F(A^*) = F(A)^* \qquad F(A \otimes B) = F(A) \otimes F(B)$$

where $\overline{e_0}$ is the ground object in $\mathcal{C}$. So we *must* define the mapping F on objects as above; our task is to see if it can be uniquely extended to arrows as a morphism of MCDs.

Let us adopt the convention of using an overbar on objects and arrows that are in $\mathcal{C}$. For example α is the associator for $\mathbf{m}[x]$ and $\overline{\alpha}$ is the associator for $\mathcal{C}$.

For any arrow (A, B) in $\mathbf{m}[x]$ we must define $F((A, B))$ to be some morphism $F(A) \to F(B)$. Let $S = \{\alpha, \rho, \lambda, \gamma, \tau, \nu, \alpha^{-1}, \rho^{-1}, \lambda^{-1}, \gamma^{-1}, \tau^{-1}, \nu^{-1}\}$ and let $\overline{S}$ be similar but using the natural isomorphisms that came with $\mathcal{C}$. The value of F on identity arrows and on $S_\blacksquare$ is uniquely determined by functoriality and the properties listed in (1) above. Similarly by functoriality and those properties we see that F is uniquely determined on $[\![S_\blacksquare]\!]$ and must take values in $[\![\overline{S}_\blacksquare]\!]$. So now we have a working and unique definition of F on objects and on reduced expansions. Can we uniquely extend it further to iterates?

If we could extend it to iterates at all then the functoriality of F grants us *uniqueness* of F on $\oint[\![S_\blacksquare]\!]\oint$, which is equal to all of $\mathbf{m}[x]^1$ by lemma 6. And it tells us that F must take values in $\oint[\![\overline{S}_\blacksquare]\!]\oint$. However *existence* still requires proof; it is not clear that we get a well-defined mapping if we attempt to send any composite of reduced expansions to the composite of their images. To complete the proof we need to show:

Claim: Define F on $\mathbf{m}[x]^0$ (objects) and on $[\![S_\blacksquare]\!]$ (reduced expansions) as above. Let overbars denote the application of this partial definition of F to formal expressions and reduced expansions. If (A, B) is an arrow in $\mathbf{m}[x]$ and $f_n \circ \cdots \circ f_1$, $g_m \circ \cdots \circ g_1$ are paths of reduced expansions from A to B, then

$$\overline{f_n} \circ \cdots \circ \overline{f_1} = \overline{g_m} \circ \cdots \circ \overline{g_1}$$

If we proved this claim then we could extend the definition of F uniquely to all of $\mathbf{m}[x]^1$. It would remain to check that $F(f \otimes g) = F(f) \otimes F(g)$ and $F(f^*) = F(f)^*$ on iterates f, g, but this would follow immediately because we have it for reduced expansions f, g and because $\otimes$ and $*$ are functorial. So if we prove this claim, F will be defined and will be a morphism of MCDs.

Proof of claim:

Terminology for proof: Let K be a set of natural tranformations in $\mathbf{Fct}\,(\mathbf{m}[x])^1$. I will refer to elements of $[\![K_\blacksquare]\!]$ as "K-moves in $\mathbf{m}[x]$," and their images under F will be called "K-moves in $\mathcal{C}$." If K' is another set of natural transformations then a "K-move followed by a K'-move in $\mathcal{C}$" is a composite $\overline{f_2} \circ \overline{f_1}$ where $f_2 \circ f_1$ is a K-move in $\mathbf{m}[x]$ followed by a K'-move in $\mathbf{m}[x]$. More generally a "K-path in $\mathcal{C}$ from A to B" is a composite $\overline{f_n} \circ \cdots \circ \overline{f_1}$ where $f_n \circ \cdots \circ f_1 : A \to B$ is a composite of K-moves in $\mathbf{m}[x]$. The claim is essentially asserting that there is only one S-path in $\mathcal{C}$ from A to B for any arrow $(A, B) \in \mathbf{m}[x]^1$.

Subclaim 1: Suppose A is a formal expression. Then there are unique R_A, η_A such that R_A is a reduced formal expression and $\eta_A : \overline{R_A} \to \overline{A}$ is a $\{\gamma, \tau, \nu\}$-path in $\mathcal{C}$ from R_A to A.
Proof:

Every formal expression is of one of the seven forms e_0, x, ${e_0}^*$, x^*, $B \otimes C$, $(B \otimes C)^*$, or B^{**}, as is evident from proposition 5. Reducing a formal expression is a matter of "pushing" the stars all the way inside and "cancelling" extra stars. To make this proof inductive we need a way to measure "how pushed in" the stars are in an expression, and how many stars there are. And since we need to work with arbitrarily long expressions we also need to measure the length of an expression in an appropriate way.

Define the *rank* $\rho(\cdot)$ and *luminosity* $\ell(\cdot)$ of formal expressions by the following recursive formulae:

$$\rho(e_0) = 1 \quad \rho(x) = 1 \quad \rho(A \otimes B) = \rho(A) + \rho(B) + 1 \quad \rho(A^*) = \rho(A)$$
$$\ell(e_0) = 0 \quad \ell(x) = 0 \quad \ell(A \otimes B) = \ell(A) + \ell(B) \quad \ell(A^*) = \ell(A) + \rho(A)$$

A calculation shows that non-identity $\{\gamma, \tau, \nu\}$-moves strictly increase luminosity:

$$\ell((A \otimes B)^*) - \ell(B^* \otimes A^*) = 1 > 0$$
$$\ell(A^{**}) - \ell(A) = 2\rho(A) > 0$$
$$\ell({e_0}^*) - \ell(e_0) = 1 > 0$$
$$\ell(A \otimes C) - \ell(B \otimes C) = \ell(A) - \ell(B) \qquad \ell(C \otimes A) - \ell(C \otimes A) = \ell(A) - \ell(B)$$

where the last line allows us to use induction on depth to extend the result to all reduced expansions.

The proof of the subclaim will be by induction on the sum of rank and luminosity. If this quantity is 1 for a formal expression A, then it must be either e or x (see proposition 5). In this case $R_A = A$ and $\eta_A = 1_{\overline{A}}$ is an identity. We have uniqueness because any non-identity $\{\gamma, \tau, \nu\}$-move would increase the sum-of-rank-and-luminosity past 1.

Now suppose the subclaim holds for any formal expression with sum-of-rank-and-luminosity lower than that of A. A is of one of the forms e_0, x, x^*, ${e_0}^*$, B^{**}, $(B \otimes C)^*$, or $B \otimes C$. We already dealt with the first two cases. In the third case A is already reduced and no non-identity reduced expansion of an instance of $\{\gamma, \tau, \nu\}$ has codomain x^* so $\eta_A = 1_{F(A)}$ works and is unique. In the fourth case set $R_A = e$ and set $\eta_A = \overline{\nu_{(\cdot)}}$. These are unique because only $\nu_{(\cdot)}$ has codomain e^*.

If A is B^{**}: Among non-identity $\{\gamma, \tau, \nu\}$-moves only the instance $\tau_B : B \to B^{**}$ has codomain B^{**}. B has strictly lower luminosity than A so there are unique R_B and $\eta' : R_B \to B$ satisfying the subclaim. Set $R_A = R_B$ and $\eta_A = \overline{\tau}_{\overline{B}} \circ \eta'$. This satisfies the subclaim, and it is unique because the last move in the composite *must* be $\overline{\tau}_{\overline{B}}$, and because η' is unique.

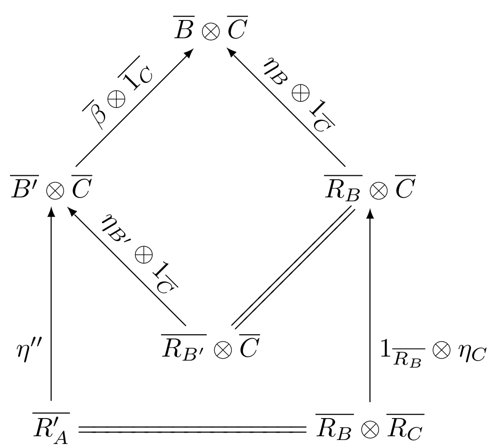
If A is $(B \otimes C)^*$: Among non-identity $\{\gamma, \tau, \nu\}$-moves only the instance $\gamma_{C,B} : C^* \otimes B^* \to (B \otimes C)^*$ has codomain $(B \otimes C)^*$. Since $C^* \otimes B^*$ has strictly lower luminosity than A we may use the same argument as above.

For the final case suppose $A = B \otimes C$. Since B and C have lower rank, we have unique R_B, R_C, $\eta_B : \overline{R_B} \to \overline{B}$, $\eta_C : \overline{R_C} \to \overline{C}$ satisfying the subclaim. Then $R_A := R_B \otimes R_C$ is a reduced formal expression (lemma 10). Let $\eta_A = (\eta_B \otimes 1_{\overline{C}}) \circ (1_{\overline{R_B}} \otimes \eta_C)$. Observe that in general if $g = \overline{f_n} \circ \cdots \circ \overline{f_1}$ is a K-path in $\mathcal{C}$ then

$$\begin{aligned} 1_{\overline{E}} \otimes g = 1_{\overline{E}} \otimes (\overline{f_n} \circ \cdots \circ \overline{f_1}) &= (1_{\overline{E}} \otimes \overline{f_n}) \circ \cdots \circ (1_{\overline{E}} \otimes \overline{f_1}) \\ &= \overline{1_E \otimes f_n} \circ \cdots \circ \overline{1_E \otimes f_1} \end{aligned}$$

is too. So η_A as we defined it is a $\{\gamma, \tau, \nu\}$-path in $\mathcal{C}$, and we have proven the existence part of the subclaim. Uniqueness is the tricky bit.

Suppose we had some reduced formal expression R'_A and suppose η'_A is a $\{\gamma, \tau, \nu\}$-path in $\mathcal{C}$ from R'_A to A. Write η'_A as $\overline{f} \circ \eta''$ where f is a non-identity $\{\gamma, \tau, \nu\}$-move in $\mathbf{w}[x]$ and η'' is some $\{\gamma, \tau, \nu\}$-path in $\mathcal{C}$ from R'_A to $\mathrm{dom}(f)$. Since the codomain of f is $B \otimes C$ we know f must be either $\beta \otimes 1_C$ or $1_B \otimes \beta$ for some $\{\gamma, \tau, \nu\}$-move β in $\mathbf{w}[x]$. Suppose we are in the former case, and $\beta : B' \to B$. Then:



commutes. $\eta_{B'}$ and $R_{B'}$ come from applying the induction hypothesis to B', which has lower rank than that of A. Commutativity of the diamond and the equality in the middle come from applying the induction hypothesis to B, which has lower rank than A. And commutativity of the bottom portion of the diagram and the bottom equality come from applying the induction hypothesis to $B' \otimes C$, which has lower luminosity than A because B' has lower luminosity than B (β is not an identity because we postulated that f is not an identity). The path on left edge of the diagram is η'_A, and the path on the right is η_A. So $\eta_A = \eta'_A$ and $R_A = R'_A$.

The case $A = 1_B \otimes \beta$ is similar, if we observe that another way to describe $\eta_A = (\eta_B \otimes 1_{\overline{C}}) \circ (1_{\overline{R_B}} \otimes \eta_C)$ is $(1_{\overline{B}} \otimes \eta_C) \circ (\eta_B \otimes 1_{\overline{R_C}})$.

Subclaim 2: Suppose A is a reduced formal expression. Then there are unique T_A, θ_A such that T_A is a trimmed formal expression and $\theta_A : \overline{T_A} \to \overline{A}$ is a $\{\rho^{-1}, \lambda^{-1}\}$-path in $\mathcal{C}$ from T_A to A.
Proof:

Define the *fluff* $\mathcal{F}(\cdot)$ and *length* $|\cdot|$ of a reduced formal expression by

$$\mathcal{F}(x) = \mathcal{F}(x^*) = 0 \qquad \mathcal{F}(e_0) = 1 \qquad \mathcal{F}(A \otimes B) = \mathcal{F}(A) + \mathcal{F}(B)$$
$$|e_0| = 0 \qquad |x| = |x^*| = 1 \qquad |A \otimes B| = |A| + |B|$$

Observe that any non-identity $\{\rho^{-1}, \lambda^{-1}\}$-move strictly increases fluff. We will use induction on the sum of length and fluff of a reduced formal expression to iteratively trim away the "fluff' until no unnecessary copies of e_0 remain. At its lowest this sum would be 1, and it is only 1 for $A \in \{e_0,\ x,\ x^*\}$. These three formal expressions are already trimmed, and so $\theta_A = 1_{\overline{A}}$ works. Uniqueness follows from the fact that a nonidentity $\{\rho^{-1}, \lambda^{-1}\}$-move strictly increases fluff.

Now suppose that A has $|A| + \mathcal{F}(A) > 1$ and suppose that the subclaim holds for any reduced formal expression with sum-of-length-and-fluff less than that of A. A is necessarily $B \otimes C$ for some reduced formal expressions B, C.

If B, C are both e_0 then $\theta_A = \overline{\rho_{e_0}^{-1}} = (\overline{\rho}_{\overline{e_0}})^{-1}$ would work. So would $\theta_A = \overline{\lambda_{e_0}^{-1}}$, but these two are equal from the coherence conditions for monoidal categories. And there is nothing else θ_A could be since no $\{\rho^{-1}, \lambda^{-1}\}$-moves in $\mathbf{ш}[x]$ other than $\rho_{e_0}^{-1}$ and $\lambda_{e_0}^{-1}$ have $e_0 \otimes e_0$ for a codomain.

Now suppose at least one of B, C is not e_0. Suppose $B \neq e_0$ (the case $C \neq e_0$ is symmetric to this one). B could have less fluff than A (if C were $e_0 \otimes \cdots \otimes e_0$) or B could have lower length than A (if C were otherwise). In either case we may apply the induction hypothesis to obtain a trimmed T_B and a $\{\rho^{-1}, \lambda^{-1}\}$-path $\theta_B : \overline{T_B} \to \overline{B}$.

There are two subcases we can proceed from: $C = e_0$ and $C \neq e_0$. Consider the former case, $A = B \otimes e_0$. Define $\theta_A = \overline{\rho_B^{-1}} \circ \theta_B : \overline{T_B} \to \overline{A}$, and define $T_A = T_B$. This clearly works, but it remains to show uniqueness. Suppose we had some T'_A, θ'_A satisfying the subclaim. Write θ'_A as $\overline{f} \circ \theta''$ for a non-identity $\{\rho^{-1}, \lambda^{-1}\}$-move f and a $\{\rho^{-1}, \lambda^{-1}\}$-path θ'' in $\mathcal{C}$. If f were ρ_B^{-1} then we get uniqueness by applying the induction hypothesis to B (i.e. ($\theta_B = \theta''$ and $T'_A = T_B = T_A$)). If f is not ρ_B^{-1} then it must be $\beta \otimes 1_{e_0}$ for some non-identity $\{\rho^{-1}, \lambda^{-1}\}$-move $\beta : B' \to B$ in $\mathbf{ш}[x]$. Applying the induction hypothesis to $B' \otimes e_0$ shows the following equality of $\overline{T_{B'}} \to \overline{B}$ arrows:

$$(\theta_{B'} \otimes 1_{\overline{e_0}}) \circ \overline{\rho_{T_{B'}}}^{-1} \quad = \quad \theta''$$

And it shows $T_{B'} = T'_A$. Applying the induction hypothesis to B shows the following equality of $\overline{T_{B'}} \to \overline{B}$ arrows:

$$\overline{\beta} \circ \theta_{B'} \quad = \quad \theta_B$$

And it shows $T_B = T_{B'}$. So we have uniqueness:

$$\begin{aligned}\theta'_A &= \overline{f} \circ \theta'' = (\overline{\beta} \otimes 1_{\overline{e_0}}) \circ (\theta_{B'} \otimes 1_{\overline{e_0}}) \circ \overline{\rho_{T_{B'}}^{-1}} = ((\overline{\beta} \circ \theta_{B'}) \otimes 1_{\overline{e_0}}) \circ \overline{\rho_{T_B}^{-1}} \\ &= (\theta_B \otimes 1_{\overline{e_0}}) \circ \overline{\rho_{T_B}^{-1}} = \overline{\rho_B^{-1}} \circ \theta_B = \theta_A\end{aligned}$$

(Naturality of ρ is used in the second line).

Finally there is the case in which neither B nor C are e_0. We may apply the induction hypothesis to obtain unique T_B, T_C and θ_B, θ_C satisfying the subclaim. Since $\{\rho^{-1}, \lambda^{-1}\}$-moves never modify the length of a formal expression, we can be sure that T_B and T_C are not e_0 (since e_0 is the only formal expression of length 0). So $T_A := T_B \otimes T_C$ is indeed trimmed. Let θ_A be

$$(\theta_B \otimes 1_{\overline{C}}) \circ (1_{\overline{T_{B'}}} \otimes \theta_C)$$

which is a $\{\rho^{-1},\lambda^{-1}\}$-path in $\mathcal{C}$ from T_A to A. For uniqueness let us suppose T'_A, θ'_A also satisfied the subclaim. Write θ'_A as $\overline{f}\circ\theta''$ for a non-identity $\{\rho^{-1},\lambda^{-1}\}$-move f and a $\{\rho^{-1},\lambda^{-1}\}$-path θ'' in $\mathcal{C}$. f must either be of the form $\beta\otimes 1_C$ or of the form $1_B\otimes\beta$ (i.e. it must work "inside" one of the factors). Suppose the former, with $\beta : B'\to B$. Apply the induction hypothesis to B to see that $T_B = T_{B'}$ and that we have the following equality of arrows $\overline{T_{B'}}\to\overline{B}$:

$$\overline{\beta}\circ\theta_{B'} \quad = \quad \theta_B$$

Apply the induction hypothesis to $B'\otimes C$ to see that $T_{A'} = T_{B'}\circ T_C$ and that we have the following equality of arrows $\overline{T'_A}\to\overline{B'}\otimes\overline{C}$:

$$(\theta_{B'}\otimes 1_{\overline{C}})\circ(1_{\overline{T_{B'}}}\otimes\theta_C) \quad = \quad \theta''$$

And so we have uniqueness:

$$\begin{aligned}\theta'_A &= \overline{f}\circ\theta'' = (\overline{\beta}\otimes 1_{\overline{C}})\circ(\theta_{B'}\otimes 1_{\overline{C}})\circ(1_{\overline{T_{B'}}}\otimes\theta_C) = ((\overline{\beta}\circ\theta_{B'})\otimes 1_{\overline{C}})\circ(1_{\overline{T_{B'}}}\otimes\theta_C)\\ &= (\theta_B\otimes 1_{\overline{C}})\circ(1_{\overline{T_{B'}}}\otimes\theta_C) = \theta_A\end{aligned}$$

The other case, $f = 1_B\otimes\beta$, can be handled similarly if we write θ_A as $(1_{\overline{B}}\otimes\theta_C)\circ(\theta_B\otimes 1_{\overline{T_C}})$.

Subclaim 3: Suppose A is a trimmed formal expression. Then there are unique S_A, ψ_A such that S_A is a standard formal expression and $\psi_A : \overline{S_A}\to\overline{A}$ is an $\{\alpha\}$-path in $\mathcal{C}$ from S_A to A.
Proof:

We will use the definition of *length* $|\cdot|$ given in the preceding subclaim. Define the *parenthesis-rank* "$p(\cdot)$" of a trimmed formal expression by:

$$p(e_0) = p(x) = p(x^*) = 0$$
$$p(A\otimes B) = p(A) + p(B) + |B| - 1$$

First we argue that a trimmed expression A has $p(A) = 0$ iff it is standard: We will use the recursive description of trimmed expressions given in lemma 10. First notice that e_0, x, x^* have a parenthesis-rank of 0 and are standard. Next suppose that $A, B\neq e_0$ each have the property that they are standard iff their parenthesis-rank vanishes. Suppose $p(A\otimes B) = 0$, i.e. $p(A)+p(B)+|B|-1 = 0$. Since $B\neq e_0$, we have $|B|\geq 1$. So $p(A)+p(B) = 1-|B|\leq 0$. So $p(A) = p(B) = 0$ and $|B| = 1$. It follows, since A, B are standard and $|B| = 1$, that $A\otimes B$ is standard. Inversely, suppose $p(A\otimes B) > 0$. Then either A is nonstandard, B is nonstandard, or $|B| > 1$. In any case, $A\otimes B$ is nonstandard.

For the next step we show that non-identity $\{\alpha\}$-moves in $\mathbf{u}[x]$ strictly increase parenthesis-rank. It works out nicely for instances of α:

$$\begin{aligned}p(B\otimes(C\otimes D)) - p((B\otimes C)\otimes D)&\\ &= p(B) + p(C\otimes D) + |C\otimes D| - p(B\otimes C) - p(D) - |D|\\ &= p(B) + p(C) + p(D) + |D| + |C\otimes D| - p(B) - p(C) - |C| - p(D) - |D|\\ &= |D| > 0 \qquad \text{(because we cannot have } |D| = 0 \text{ for a trimmed expression)}\end{aligned}$$

The rest is by induction on the depth of reduced expansions of instances of α. If the $\{\alpha\}$-move $\beta : B'\to B$ strictly increases rank then so does $\beta\otimes 1_C$:

$$p(B\otimes C) - p(B'\otimes C) = p(B) - p(B')$$

If the $\{\alpha\}$-move $\beta : C' \to C$ strictly increases rank then so does $1_B \otimes \beta$:

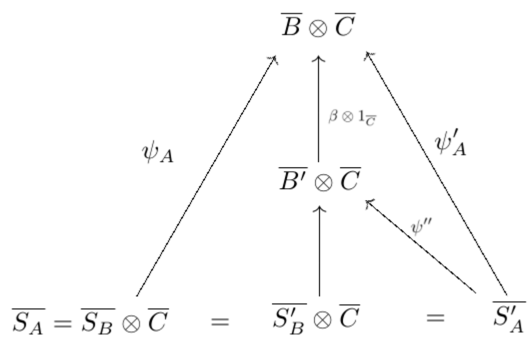
$$p(B \otimes C) - p(B \otimes C') = p(C) - p(C') + |C| - |C'| = p(C) - p(C')$$

where we have used the fact that $|C'| = |C|$ if C and C' are trimmed expressions related by an $\{\alpha\}$-move (this is also easy to see by induction on the depth of reduced expansions of instances of α).

Now we may prove the subclaim by induction on parenthesis-rank. If $p(A) = 0$ then A is already standard, so $\psi_A = 1_{\overline{A}}$ will do. Uniqueness follows from the fact that any non-identity $\{\alpha\}$-move would strictly increase parenthesis-rank.

Suppose A has $p(A) > 0$ and suppose the subclaim holds for any trimmed expression with lower parenthesis-rank. Then A is of the form $B \otimes C$ for $B, C \neq e_0$.

Consider the case $p(C) = 0$. Apply the induction hypothesis to B to obtain S_B, ψ_B. Let $S_A = S_B \otimes C$, a standard formal expression. Let $\psi_A = \psi_B \otimes 1_{\overline{C}}$, an $\{\alpha\}$-path in $\mathcal{C}$ from S_A to A. For uniqueness: Suppose S'_A, ψ'_A also satisfy the subclaim. Write ψ'_A as $\overline{f} \circ \psi''$ for some non-identity $\{\alpha\}$-move f and for an $\{\alpha\}$-path ψ'' in $\mathcal{C}$. Since $|C| = 0$, f must be of the form $\beta \otimes 1_C$ for some non-identity $\{\alpha\}$-move $\beta : B' \to B$. The diagram



commutes. The left side commutes and $S_B = S'_B$ by an application of the induction hypothesis to B. The bottom-right commutes and $S'_B \otimes C = S'_A$ by an application of the induction hypothesis to $B' \otimes C$.

Now consider the case $p(C) \neq 0$. C is then of the form $D \otimes E$ for $D, E \neq e_0$, so $A = B \otimes (D \otimes E)$. Apply the induction hypothesis to the lower rank object $(B \otimes D) \otimes E$ to obtain $S_{(B\otimes D)\otimes E}$, $\psi_{(B\otimes D)\otimes E}$. Let $S_A = S_{(B\otimes D)\otimes E}$ and let $\psi_A = \overline{\alpha}_{B,D,E} \circ \psi_{(B\otimes D)\otimes E}$. It remains to prove uniqueness. Suppose S'_A, ψ'_A also satisfy the subclaim, and write ψ'_A as $\overline{f} \circ \psi''$ for some non-identity $\{\alpha\}$-move f and for an $\{\alpha\}$-path ψ'' in $\mathcal{C}$. Since the codomain of f is $B \otimes (D \otimes E)$ we know that f has one of the three forms:

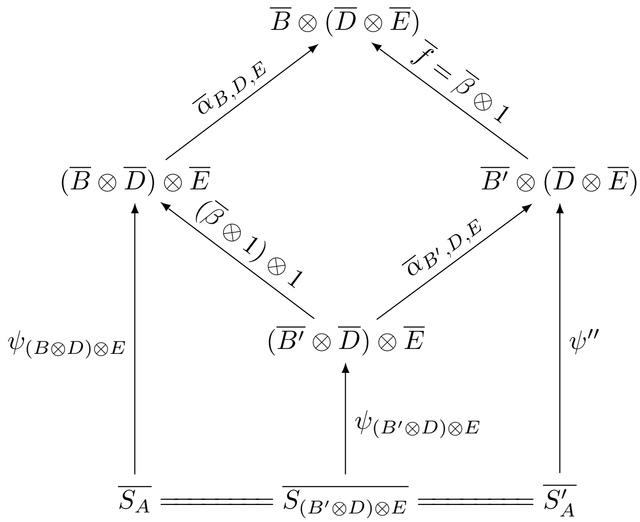
$$\alpha_{B,D,E}, \ \beta \otimes 1, \ 1 \otimes \beta$$

where β denotes an $\{\alpha\}$-move. We can subdivide the final case into three subcases to see that f is one of:

$$(1)\ \alpha_{B,D,E}, \quad (2)\ \beta \otimes 1, \quad (3)\ 1 \otimes \alpha_{D,F,G}, \quad (4)\ 1 \otimes (\beta \otimes 1), \quad (5)\ 1 \otimes (1 \otimes \beta)$$

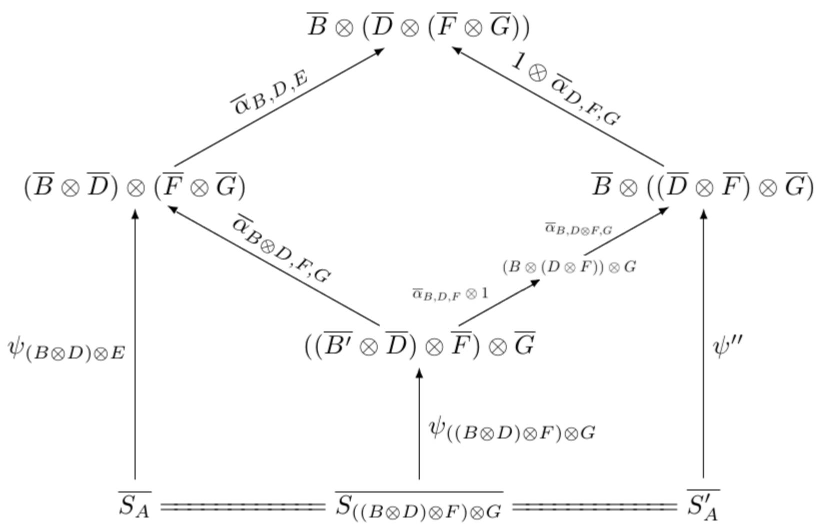
where in case (3) we have written E as $F \otimes G$ with $F, G \neq e_0$. In case (1) we may simply apply uniqueness of $\psi_{(B\otimes D)\otimes E}$ to see that $S_A = S'_A$ and $\psi_A = \psi'_A$. Cases (2), (4), and (5) are handled similarly using the naturality of $\overline{\alpha}$, so we will only show case (2) and then case (3).

Suppose $f = \beta\otimes 1_{D\otimes E}$, where $\beta : B' \to B$. In this case we get uniqueness from the commutativity of the following diagram:



The diamond commutes by the naturality of $\overline{\alpha}$. Commutativity of the left side and $S_A = S_{(B'\otimes D)\otimes E}$ are obtained by applying the induction hypothesis to $(B\otimes D)\otimes E$. Commutativity of the right side and $S'_A = S_{(B'\otimes D)\otimes E}$ are obtained by applying the induction hypothesis to $B'\otimes(D\otimes E)$. The left and right edges of the diagram are ψ_A and ψ'_A, so we get $\psi_A = \psi'_A$ and $S_A = S'_A$.

Only case (3) remains. Suppose $f = 1_B\otimes\alpha_{D,F,G}$, and $A = B\otimes(D\otimes(F\otimes G))$. Uniqueness is obtained using a similar method to the other cases, but this time the diamond is replaced by the pentagon from the coherence condition that $\mathcal{C}$ satisfies!



Subclaim 4: A $\{\gamma,\tau,\nu\}$-move followed by an $\{\alpha,\rho^{-1},\lambda^{-1}\}$-move in $\mathcal{C}$ can be written as an $\{\alpha,\rho^{-1},\lambda^{-1}\}$-

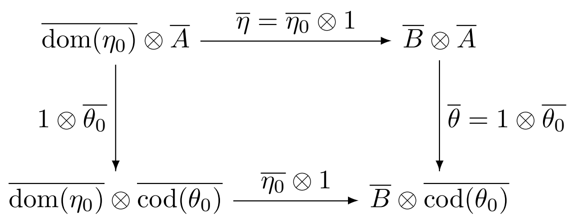
move followed by a $\{\gamma, \tau, \nu\}$-move in $\mathcal{C}$.
Proof:

Let η be a $\{\gamma, \tau, \nu\}$-move in $\mathbf{m}[x]$ and let θ be an $\{\alpha, \rho^{-1}, \lambda^{-1}\}$-move in $\mathbf{m}[x]$ such that $\mathrm{dom}(\theta) = \mathrm{cod}(\eta)$. Let n be the depth of η and let m be the depth of θ. The goal is to show that $\overline{\theta} \circ \overline{\eta}$ can be written as some $\overline{\eta'} \circ \overline{\theta'}$ with η' a $\{\gamma, \tau, \nu\}$-move and θ' an $\{\alpha, \rho^{-1}, \lambda^{-1}\}$-move. We will use induction on (n, m). There will be a "diagonal" induction step deducing the $(n+1, m+1)$ case from the (n, m) case, and then we will have to prove the "edge" cases $(n, 0)$ and $(0, m)$.

If either θ or η are identities then we are done, so suppose neither is an identity.

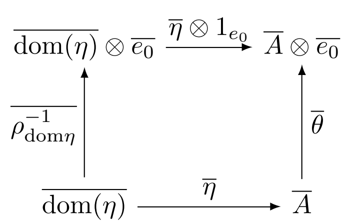
Diagonal induction part: Suppose $n, m > 0$. If $\eta = \eta_0 \otimes 1_A$ and $\theta = \theta_0 \otimes 1_B$ then $\mathrm{dom}(\theta_0) \otimes B = \mathrm{dom}(\theta) = \mathrm{cod}(\eta) = \mathrm{cod}(\eta_0) \otimes A$, so $\mathrm{dom}(\theta_0) = \mathrm{cod}(\eta_0)$ and $A = B$. (This sort of argument works due to the construction of the free MCD. From now on I will be less explicit about it). Then applying the induction hypothesis to η_0, θ_0, which have strictly lower depth, handles the case.

Consider the case where $\eta = \eta_0 \otimes 1_A$ and $\theta = 1_B \otimes \theta_0$. We get $\mathrm{cod}(\eta_0) = B$ and $\mathrm{dom}(\theta_0) = A$, and an easy commutative diagram finishes off the argument:



The remaining two cases are symmetric to the two we already handled.

"Corner" base case (0,0): Suppose both η and θ have zero depth. Then η is an instance of γ, τ, or ν. So $\mathrm{cod}(\eta)$ has one of the forms: ${e_0}^*, (A \otimes B)^*$, or A^{**}. It follows that θ cannot be an instance of α, for then $\mathrm{dom}(\theta)$ would be of the form $(A \otimes B) \otimes C$. So θ must be an instance of ρ^{-1} or λ^{-1}. The two cases are similar; we will treat the case $\theta = \rho_A^{-1} : A \to A \otimes e_0$. In this case $\mathrm{cod}(\eta) = \mathrm{dom}(\theta) = A$, and we are done by the naturality of $\overline{\rho}^{-1}$:



Edge base case (n,0): Suppose $n > 0$ and $m = 0$. If θ is an instance of λ^{-1} or ρ^{-1} then the method applied in the corner base case above applies, since η is either of the form $\eta_0 \otimes 1$ or $1 \otimes \eta_0$ (i.e. use the naturality of $\overline{\rho}^{-1}$ or $\overline{\lambda}^{-1}$). So suppose that θ is an instance of α, say $\theta = \alpha_{A,B,C}$. Then $\mathrm{cod}(\eta) = \mathrm{dom}(\theta) = (A \otimes B) \otimes C$. If $\eta = 1_{A \otimes B} \otimes \eta_0$ then $\mathrm{cod}(\eta_0) = C$ and we need only cite the naturality of $\overline{\alpha}$:

$$\begin{array}{ccc}
(\overline{A}\otimes\overline{B})\otimes\overline{\mathrm{dom}(\eta_0)} & \xrightarrow{\overline{\eta}} & (\overline{A}\otimes\overline{B})\otimes\overline{C} \\
\Big\downarrow \overline{\alpha_{A,B,\mathrm{dom}\eta_0}} & & \Big\downarrow \overline{\theta} \\
\overline{A}\otimes(\overline{B}\otimes\overline{\mathrm{dom}(\eta_0)}) & \xrightarrow{1_{\overline{A}}\otimes(1_{\overline{B}}\otimes\overline{\eta_0})} & \overline{A}\otimes(\overline{B}\otimes\overline{C})
\end{array}$$

If $\eta = \eta_0 \otimes 1_C$ on the other hand, then $\mathrm{cod}(\eta_0) = A \otimes B$, and we need to consider the depth of η_0. If the depth of η_0 were zero, then $\mathrm{cod}(\eta_0)$ would have the form ${e_0}^*$, D^{**}, or $(D \otimes E)^*$, which could not be. So η_0 has positive depth and we can write η as either $(\eta_1 \otimes 1_B) \otimes 1_C$ or $(1_A \otimes \eta_1) \otimes 1_C$ for some η_1 of depth $n-2$. In either case we are done by the naturality of $\overline{\alpha}$:

$$\begin{array}{ccc}
(\overline{\mathrm{dom}(\eta_1)}\otimes\overline{B})\otimes\overline{C} & \xrightarrow{(\overline{\eta_1}\otimes 1)\otimes 1} & (\overline{A}\otimes\overline{B})\otimes\overline{C} \\
\Big\downarrow \overline{\alpha} & & \Big\downarrow \overline{\alpha} \\
\overline{\mathrm{dom}(\eta_1)}\otimes(\overline{B}\otimes\overline{C}) & \xrightarrow{\overline{\eta_1}\otimes 1} & \overline{A}\otimes(\overline{B}\otimes\overline{C})
\end{array}$$

$$\begin{array}{ccc}
(\overline{A}\otimes\overline{\mathrm{dom}(\eta_1)})\otimes\overline{C} & \xrightarrow{1\otimes(\overline{\eta_1})\otimes 1} & (\overline{A}\otimes\overline{B})\otimes\overline{C} \\
\Big\downarrow \overline{\alpha} & & \Big\downarrow \overline{\alpha} \\
\overline{A}\otimes(\overline{\mathrm{dom}(\eta_1)}\otimes\overline{C}) & \xrightarrow{1\otimes(\overline{\eta_1}\otimes 1)} & \overline{A}\otimes(\overline{B}\otimes\overline{C})
\end{array}$$

Edge base case (0,m): Suppose $m > 0$ and $n = 0$. Then $\mathrm{cod}(\eta)$ has one of the forms ${e_0}^*$, A^{**}, $(A \otimes B)^*$. But $\mathrm{dom}(\theta)$ cannot have this form unless θ were an instance of ρ^{-1} or λ^{-1}, in which case the method applied in the corner base case applies here (i.e. use the naturality of $\overline{\rho}^{-1}$ and $\overline{\lambda}^{-1}$).

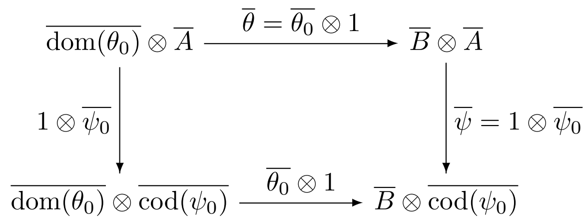
Subclaim 5: A $\{\rho^{-1}, \lambda^{-1}\}$-move followed by an $\{\alpha\}$-move in $\mathcal{C}$ can be written as an $\{\alpha\}$-move followed by a $\{\rho^{-1}, \lambda^{-1}\}$-move in $\mathcal{C}$.
Proof:

Let θ be a $\{\rho^{-1}, \lambda^{-1}\}$-move in $\mathbf{ш}[x]$ and let ψ be an $\{\alpha\}$-move in $\mathbf{ш}[x]$ such that $\mathrm{dom}(\psi) = \mathrm{cod}(\theta)$. Let n be the depth of ψ and let m be the depth of θ. The goal is to show that $\overline{\psi} \circ \overline{\theta}$ can be written as some $\overline{\theta'} \circ \overline{\psi'}$ with θ' a $\{\rho^{-1}, \lambda^{-1}\}$-move and ψ' an $\{\alpha\}$-move. We will use induction on (n, m). There will be a "diagonal" induction step deducing the $(n+1, m+1)$ case from the (n, m) case, and then we will have to prove the "edge" cases $(n, 0)$ and $(0, m)$.

If either θ or ψ are identities then we are done, so suppose neither is an identity.

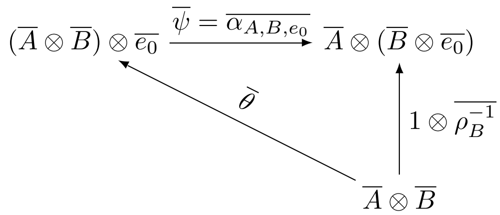
Diagonal induction part: Suppose $n, m > 0$. In the case that $\theta = \theta_0 \otimes 1_A$ and $\psi = \psi_0 \otimes 1_B$ (with θ_0 a $\{\rho^{-1}, \lambda^{-1}\}$-move in $\mathbf{ш}[x]$ and ψ_0 and $\{\alpha\}$-move in $\mathbf{ш}[x]$) we can deduce that $A = B$ and $\mathrm{dom}(\psi_0) = \mathrm{cod}(\theta_0)$. Applying the induction hypothesis to θ_0 and ψ_0 (i.e. applying the induction hypothesis to the move $\overline{\psi}_0 \circ \overline{\theta}_0$) then handles the case.

Consider now the case where $\theta = \theta_0 \otimes 1_A$ and $\psi = 1_B \otimes \psi_0$. We have $\mathrm{dom}(\psi_0) = A$ and $\mathrm{cod}(\theta_0) = B$, and the following easy commutative diagram finishes off the argument:



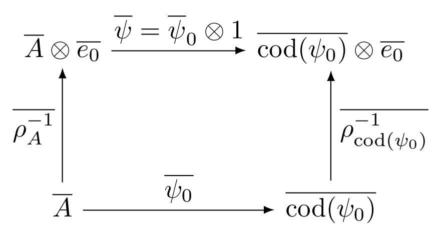
The remaining two cases are symmetric to the two we already handled.

"Corner" base case (0,0): Suppose both θ and ψ have zero depth. Then $\psi = \alpha_{A,B,C}$ for some formal expressions A, B, C, so $\mathrm{cod}(\theta) = \mathrm{dom}(\psi) = (A \otimes B) \otimes C$. θ is either an instace of ρ^{-1} or an instance of λ^{-1}, but it cannot be the latter because e_0 cannot have the form $A \otimes B$. It follows that we must have $C = e_0$ and $\theta = \rho^{-1}_{A \otimes B}$. The commutativity of



is then all we need. Showing this from the coherence conditions for a monoidal category is a (good) exercise in [1, p161].

Edge base case (n,0): Suppose $n > 0$ and $m = 0$. We know θ is an instance of ρ^{-1} or λ^{-1}. The two cases are similar so let show the former. Suppose that $\theta = \rho_A^{-1}$. Note that ψ cannot have the form $1_B \otimes \psi_0$, for this would imply that $A \otimes e_0 = \mathrm{cod}(\theta) = \mathrm{dom}(\psi) = B \otimes \mathrm{dom}(\psi_0)$ which gives $\mathrm{dom}(\psi_0) = e_0$, forcing ψ_0, and thereby ψ, to be an identity. It follows that ψ has the form $\psi_0 \otimes 1_B$. Then $A \otimes e_0 = \mathrm{cod}(\theta) = \mathrm{dom}(\psi) = \mathrm{dom}(\psi_0) \otimes B$, so $\mathrm{dom}(\psi_0) = A$ and $B = e_0$. Naturality of $\overline{\rho}^{-1}$ does the trick:



Edge base case (0,m): Suppose $m > 0$ and $n = 0$, and assume the subclaim for smaller m. We know $\psi = \alpha_{A,B,C}$ for some formal expressions A, B, C. θ can take one of two forms; suppose first that $\theta = 1 \otimes \theta_0$. Then $\theta_0 : D \to C$ for some D and we have

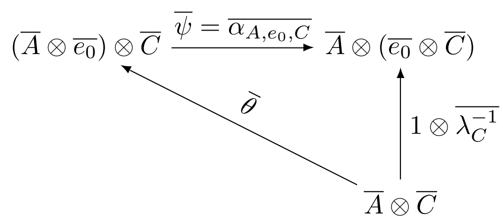
$$\begin{aligned}\overline{\psi} \circ \overline{\theta} &= \overline{\alpha_{A,B,C}} \circ (1_{\overline{A} \otimes \overline{B}} \otimes \overline{\theta}_0) \\ &= (1_{\overline{A}} \otimes (1_{\overline{B}} \otimes \overline{\theta}_0)) \circ \overline{\alpha_{A,B,D}}\end{aligned}$$

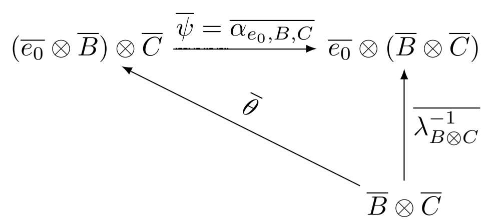
from the naturality of $\bar{\alpha}$. Now consider the remaining case where θ takes the form $\theta = \theta_0 \otimes 1$. Then $\theta_0 : \mathrm{dom}(\theta_0) \to A \otimes B$, so we must consider the possibilities for θ_0. If it has positive depth (i.e. $m > 1$), then θ is one of the two forms

$$1 \otimes (1 \otimes \theta_1)$$

$$1 \otimes (\theta_1 \otimes 1)$$

We can then prove the subclaim by using the naturality of $\overline{\alpha}$ in a manner similar to the previous case we handled. If, on the other hand, the depth of θ_0 is zero (i.e. $m = 1$), then θ_0 is either ρ_A^{-1} or λ_B^{-1}. We can handle these cases with the following diagrams, respectively:





The first diagram commutes because it is one of the original coherence conditions! The second diagram is a mirror image of the exercise that appeared in the $(0, 0)$ case.

Subclaim 6: Suppose A is a formal expression. Then there are unique S_A, ϕ_A such that S_A is a standard formal expression and $\phi_A : F(S_A) \to F(A)$ is a $\{\gamma, \tau, \nu, \alpha, \rho^{-1}, \lambda^{-1}\}$-path in $\mathcal{C}$ from S_A to A.
Proof:

Existence is a straight-forward application of subclaims 1-3; reduce, trim, and standardize to obtain a composite

$$\overline{S_{T_{R_A}}} \xrightarrow{\psi_{T_{R_A}}} \overline{T_{R_A}} \xrightarrow{\theta_{R_A}} \overline{R_A} \xrightarrow{\eta_A} \overline{A}$$

which we define to be ϕ_A. Rename the arrows above to

$$\overline{S_A} \xrightarrow{\psi_A} \overline{T_A} \xrightarrow{\theta_A} \overline{R_A} \xrightarrow{\eta_A} \overline{A}$$

for convenience. Each arrow in the composite above is itself a composite of $\{\gamma, \tau, \nu, \alpha, \rho^{-1}, \lambda^{-1}\}$-moves in $\mathbf{uu}[x]$. The composite of the images of all those moves under F gives an isomorphism $\overline{S_A} \to \overline{A}$. To show uniqueness, suppose that $\overline{f_n} \circ \cdots \circ \overline{f_1}$ is another $\{\gamma, \tau, \nu, \alpha, \rho^{-1}, \lambda^{-1}\}$-path in $\mathcal{C}$ from S'_A to A, with $f_1, \cdots, f_n$ being the $\{\gamma, \tau, \nu, \alpha, \rho^{-1}, \lambda^{-1}\}$-moves in $\mathbf{uu}[x]$ that make up the path, and with S'_A being a standard formal expression. By subclaim 4 we may write $\overline{f_n} \circ \cdots \circ \overline{f_1}$ as

$$\overline{g_n} \circ \cdots \circ \overline{g_{k+1}} \circ \overline{g_k} \circ \cdots \circ \overline{g_1}$$

where $\overline{g_k} \circ \cdots \circ \overline{g_1}$ is a $\{\alpha, \rho^{-1}, \lambda^{-1}\}$-path in $\mathcal{C}$ and $\overline{g_n} \circ \cdots \circ \overline{g_{k+1}}$ is a $\{\gamma, \tau, \nu\}$-path in $\mathcal{C}$. Let $R'_A = \mathrm{dom}(g_{k+1}) = \mathrm{cod}(g_k)$. R'_A must be reduced because $\{\alpha, \rho^{-1}, \lambda^{-1}\}$-moves preserve rank and luminosity and S'_A is reduced. Then by subclaim 1 we have $R'_A = R_A$ and

$$\overline{g_n} \circ \cdots \circ \overline{g_{k+1}} = \eta_A$$

By subclaim 5 we may write $\overline{g_k} \circ \cdots \circ \overline{g_1}$ as

$$\overline{h_k} \circ \cdots \circ \overline{h_{l+1}} \circ \overline{h_l} \circ \cdots \circ \overline{h_1} \ ,$$

where $\overline{h_k} \circ \cdots \circ \overline{h_{l+1}}$ is a $\{\rho^{-1}, \lambda^{-1}\}$-path in $\mathcal{C}$ and $\overline{h_l} \circ \cdots \circ \overline{h_1}$ is an $\{\alpha\}$-path in $\mathcal{C}$. Let $T'_A = \mathrm{cod}(h_l)$. T'_A must be trimmed because S'_A is trimmed and $\{\alpha\}$-moves preserve fluff and length. It follows from subclaim 2 that $T'_A = T_A$ and

$$\overline{h_k} \circ \cdots \circ \overline{h_{l+1}} = \theta_A.$$

Subclaim 3 then ensures that $S'_A = S_A$ and

$$\overline{h_l} \circ \cdots \circ \overline{h_1} = \psi_A.$$

This shows uniqueness:

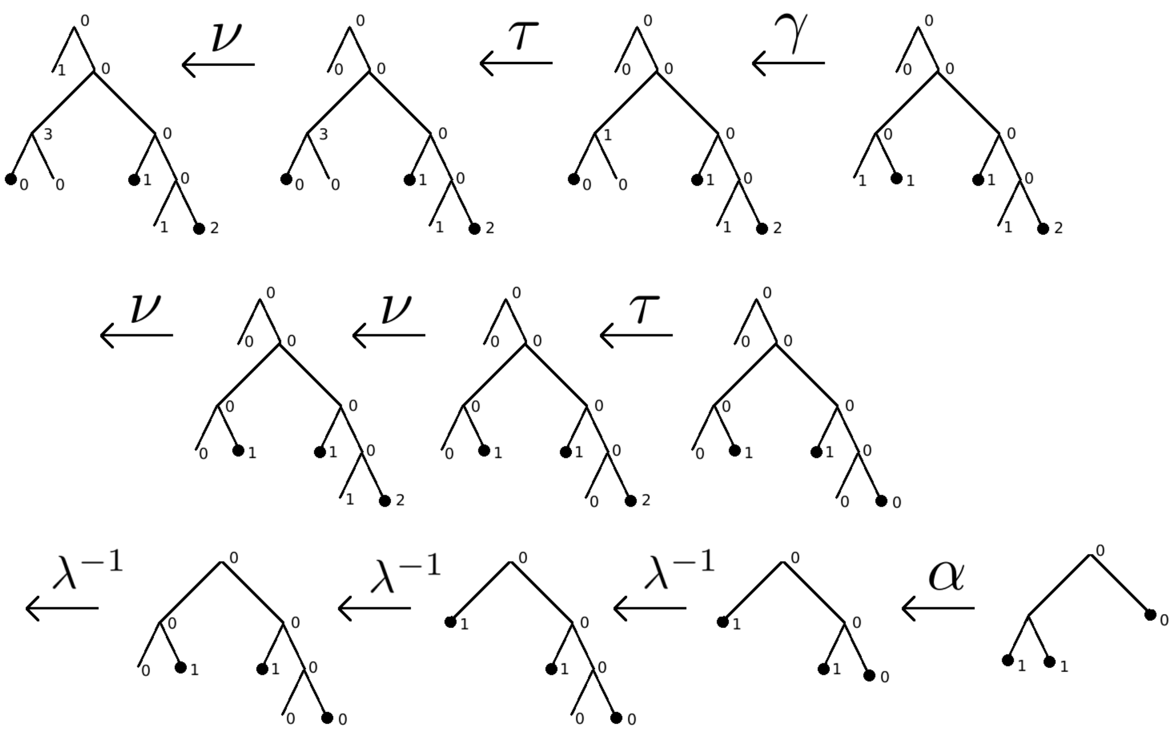
$$\overline{f_n} \circ \cdots \circ \overline{f_1} = \eta_A \circ \theta_A \circ \psi_A = \phi_A$$

Back to proof of claim: Let $K = \{\gamma, \tau, \nu, \alpha, \rho^{-1}, \lambda^{-1}\}$ and let " K^{-1} " denote $\{\gamma^{-1}, \tau^{-1}, \nu^{-1}, \alpha^{-1}, \rho, \lambda\}$. Suppose (A, B) is an arrow in $\mathbf{m}[x]$ and $f_n \circ \cdots \circ f_1$ is a path of reduced expansions of instances of S from A to B. We will show that $\overline{f_n} \circ \cdots \circ \overline{f_1}$ is something that depends only on the endpoints A and B of the path. Suppose $f_i : A_{i-1} \to A_i$ for $1 \leq i \leq n$, with $A_0 = A$ and $A_n = B$. f_1 is either a K-move in $\mathbf{m}[x]$ or a K^{-1}-move in $\mathbf{m}[x]$ (since $S = K \cup K^{-1}$). In the former case we have (using the notation of subclaim 6) $S_{A_1} = S_{A_0}$ and $\phi_{A_1} = \overline{f_1} \circ \phi_{A_0}$ from the uniqueness in subclaim 6. In the latter case we have $S_{A_1} = S_{A_0}$ and $(\overline{f_1})^{-1} \circ \phi_{A_1} = \phi_{A_0}$ for the same reason, so again we have $\phi_{A_1} = \overline{f_1} \circ \phi_{A_0}$. Iterating this result gets us $\phi_{A_n} = \overline{f_n} \circ \cdots \circ \overline{f_1} \circ \phi_{A_0}$. In other words:

$$\overline{f_n} \circ \cdots \circ \overline{f_1} = \phi_B \circ \phi_A^{-1}$$

□

Here is an illustration of reducing, trimming, and standardizing a formal expression:



5 Coherence

Let $(\mathcal{C}, \otimes, *, e_0, \alpha, \rho, \lambda, \gamma, \tau, \nu)$ be an arbitrary MCD. Theorem 11 shows that $\mathbf{w}[x]$ is free on $\{x\}$. It follows that there is a unique morphism of MCDs

$$F : \mathbf{w}[x] \to \mathbf{Fct}\,(\mathcal{C})$$

sending x to the identity functor $1_{\mathcal{C}}$. The iterated functors and iterated natural transformations of section 3 are the results of applying F to formal expressions and the arrows between them. Recall that in section 3 we used hats to denote the natural isomorphisms associated to $\mathbf{Fct}\,(\mathcal{C})$.

Coherence (formal expression definition): $\mathcal{C}$ is *coherent* if any diagram with formal expressions for vertices and suitable expansions of instances of $\hat{\alpha}, \hat{\rho}, \hat{\lambda}, \hat{\gamma}, \hat{\tau}, \hat{\nu}$ (and their inverses) for edges commutes. An expansion is suitable as an edge from formal expression A to formal expression B if it is an arrow $F(A) \to F(B)$.

Notice that the issue of vertex collapse is avoided by letting the vertices be formal expressions. The sort of diagram considered in this definition of coherence has vertices in $\mathbf{w}[x]$ and edges in

$$\llbracket \{\hat{\alpha}, \hat{\rho}, \hat{\lambda}, \hat{\gamma}, \hat{\tau}, \hat{\nu}, \hat{\alpha}^{-1}, \hat{\rho}^{-1}, \hat{\lambda}^{-1}, \hat{\gamma}^{-1}, \hat{\tau}^{-1}, \hat{\nu}^{-1}\}_{\blacksquare} \rrbracket \subset \mathbf{Fct}\,(\mathcal{C})^1 \ .$$

Diagrams of this form are images under F of diagrams in $\mathbf{w}[x]$. But all diagrams in $\mathbf{w}[x]$ commute by construction, so coherence always holds. That is, the coherence conditions described in section 1 were indeed sufficient to ensure coherence.

References

[1] S. Maclane (1971), *Categories for the Working Mathematician*, Springer-Verlag , New York .

[2] J. W. Barrett and B. W. Westbury, *Spherical categories*, preprint, hep-th/9310164, University of Nottingham, 1993.

things to add to this:

6 perhaps a section on strictification?

see Joyal's "cliques"

7 perhaps a section showing how to get γ, τ, ν for a pivotal cat?

see personal notes from muger, and 8 exercises