

A sequent calculus for the Tamari order

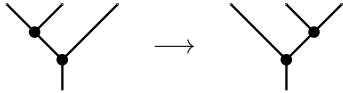
Noam Zeilberger
 University of Birmingham
 noam.zeilberger@gmail.com

Abstract—We introduce a sequent calculus with a simple restriction of Lambek’s product rules that precisely captures the classical Tamari order, i.e., the partial order on fully-bracketed words (equivalently, binary trees) induced by a semi-associative law (equivalently, tree rotation). We establish a focusing property for this sequent calculus (a strengthening of cut-elimination), which yields the following coherence theorem: every valid entailment in the Tamari order has exactly one focused derivation. One combinatorial application of this coherence theorem is a new proof of the Tutte–Chapoton formula for the number of intervals in the Tamari lattice Y_n . We also apply the sequent calculus and the coherence theorem to build a surprising bijection between intervals of the Tamari order and a certain fragment of lambda calculus, consisting of the β -normal planar lambda terms with no closed proper subterms.

I. INTRODUCTION

A. The Tamari order, and Tamari lattices

Suppose you are given a pair of binary trees A and B and the following problem: *transform A into B using only right rotations*. Recall that a right rotation is an operation acting locally on a pair of internal nodes of a binary tree, rearranging them like so:



Solving this problem amounts to showing that $A \leq B$ in the *Tamari order*. Originally introduced by Dov Tamari in the study of monoids with a partially-defined multiplication operation [10], [29], [30], the Tamari order is the partial ordering on words induced by asking that multiplication obeys a *semi-associative law*¹

$$(A * B) * C \leq A * (B * C)$$

and is monotonic in each argument:

$$\frac{A \leq A'}{A * B \leq A' * B} \quad \frac{B \leq B'}{A * B \leq A * B'}$$

For example, the word $(p * (q * r)) * s$ is below the word $p * (q * (r * s))$ in the Tamari order. The variables p, q, \dots are just placeholders and what really matters is the underlying shape of such “fully-bracketed words”, which is what justifies the above description of the Tamari order in terms of tree rotations. Since, binary trees are enumerated by the ubiquitous Catalan numbers (there are $C_n = \binom{2n}{n}/(n+1)$ distinct binary

¹Clearly, one has to make an arbitrary choice in orienting associativity from left-to-right or right-to-left, and Tamari’s original papers in fact took the opposite convention. The literature is inconsistent about this, but since the two possible orders defined are strictly dual it does not make much difference.

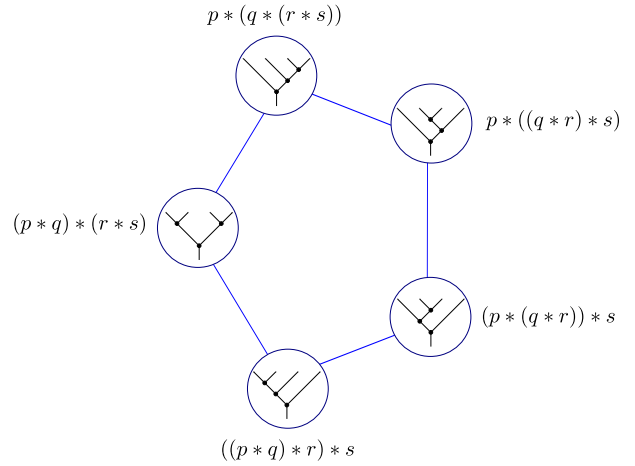


Fig. 1. The Tamari lattice Y_3 .

trees with n internal nodes) which also count many other isomorphic families of objects, the Tamari order has many other equivalent formulations as well, such as on strings of balanced parentheses [12], triangulations of a polygon [26], or Dyck paths [3] (see also [13, pp. 474–475]).

For any fixed natural number n , the C_n objects of that size form a lattice under the Tamari order, which is called the *Tamari lattice* Y_n . For example, Figure 1 shows the Hasse diagram of Y_3 , which has the shape of a pentagon, and readers familiar with category theory may recognize this as “Mac Lane’s pentagon” [20]. More generally, a fascinating property of the Tamari order is that each lattice Y_n generates via its Hasse diagram the underlying graph of an $(n - 1)$ -dimensional polytope called an “associahedron” [23], [28].

B. A Lambekian analysis of the Tamari order

In this paper we will consider a surprisingly elementary (but to the best of my knowledge previously unstudied) presentation of the Tamari order as a *sequent calculus* in the spirit of Lambek [16], [17]. This calculus consists of just one *left rule* and one *right rule*:

$$\frac{A, B, \Delta \longrightarrow C}{A * B, \Delta \longrightarrow C} *L \quad \frac{\Gamma \longrightarrow A \quad \Delta \longrightarrow B}{\Gamma, \Delta \longrightarrow A * B} *R$$

together with two *structural rules*:

$$\frac{}{A \longrightarrow A} id \quad \frac{\Theta \longrightarrow A \quad \Gamma, A, \Delta \longrightarrow B}{\Gamma, \Theta, \Delta \longrightarrow B} cut$$

Here, Γ , Δ , and Θ range over lists of formulas called *contexts*, and we write a comma to indicate concatenation of contexts (which is a strictly associative operation).

In fact, all of these rules come straight from Lambek [16], except for the $*L$ rule which is a restriction of his left rule for products. Lambek’s original rule looked like this:

$$\frac{\Gamma, A, B, \Delta \longrightarrow C}{\Gamma, A * B, \Delta \longrightarrow C} *L^{\text{amb}}$$

That is, Lambek’s left rule allowed the formula $A*B$ to appear anywhere in the context, whereas our more restrictive rule $*L$ requires the formula to appear at the leftmost end of the context. It turns out that this simple variation makes all the difference for capturing the Tamari order!

For example, here is a sequent derivation of the entailment $(p*(q*r)) * s \leq p*(q*(r*s))$ (we write L and R as short for $*L$ and $*R$, and don’t bother labelling instances of id):

$$\frac{\frac{\frac{\frac{q \longrightarrow q \quad r \longrightarrow r}{q, r \longrightarrow q * r} R \quad s \longrightarrow s}{p \longrightarrow p} R}{\frac{q, r, s \longrightarrow q * (r * s)}{q * r, s \longrightarrow q * (r * s)} L} R}{\frac{p, q * r, s \longrightarrow p * (q * (r * s))}{p * (q * r), s \longrightarrow p * (q * (r * s))} L} L}{(p * (q * r)) * s \longrightarrow p * (q * (r * s))} L$$

If we had full access to Lambek’s original rule then we could also derive the converse entailment (which is false for Tamari):

$$\frac{\frac{\frac{\frac{q \longrightarrow q \quad r \longrightarrow r}{q, r \longrightarrow q * r} R \quad s \longrightarrow s}{p, q, r \longrightarrow p * (q * r)} R}{\frac{p, q, r, s \longrightarrow (p * (q * r)) * s}{p, q, r * s \longrightarrow (p * (q * r)) * s} L^{\text{amb}}} R^{\text{amb}}}{\frac{p, q * (r * s) \longrightarrow (p * (q * r)) * s}{p * (q * (r * s)) \longrightarrow (p * (q * r)) * s} L} L$$

But with the more restrictive rule we can’t – the following soundness and completeness result will be established below.

Claim I.1. $A \longrightarrow B$ is derivable using the rules $*L$, $*R$, id , and cut if and only if $A \leq B$ holds in the Tamari order.

As Lambek emphasized, the real power of a sequent calculus comes when it is combined with Gentzen’s *cut-elimination* procedure [11]. We will prove the following somewhat stronger form of cut-elimination:

Claim I.2. If $\Gamma \longrightarrow A$ is derivable using the rules $*L$, $*R$, id , and cut , then it has a derivation using only $*L$ together with the following restricted forms of $*R$ and id :

$$\frac{\Gamma^{\text{irr}} \longrightarrow A \quad \Delta \longrightarrow B}{\Gamma^{\text{irr}}, \Delta \longrightarrow A * B} *R^{\text{foc}} \quad \frac{}{p \longrightarrow p} id^{\text{atm}}$$

where Γ^{irr} ranges over contexts that don’t have a product formula $C * D$ at their leftmost end.

This is actually a *focusing completeness* result in the sense of Andreoli [1], and we will refer to derivations constructed using

only the rules $*L$, $*R^{\text{foc}}$, and id^{atm} as *focused derivations*. (The above derivation of $(p*(q*r)) * s \leq p*(q*(r*s))$ is an example of a focused derivation.) A basic analysis of these three rules confirms that any sequent $\Gamma \longrightarrow A$ has at most one focused derivation. By combining Claims I.1 and I.2, we therefore have

Claim I.3. Every valid entailment in the Tamari order has exactly one focused derivation.

This *coherence theorem* is the main contribution of the paper, and we will see that it has several interesting applications.

C. The surprising combinatorics of Tamari intervals, planar maps, and planar lambda terms

The original impetus for this work came from wanting to better understand an apparent link between the Tamari order and lambda calculus, which was inferred indirectly via their mutual connection to the combinatorics of embedded graphs.

About a dozen years ago, Frédéric Chapoton [5] proved the following surprising formula for the number of *intervals* in the Tamari lattice Y_n :

$$\frac{2(4n+1)!}{(n+1)!(3n+2)!} \quad (1)$$

Here, by an “interval” of a partially ordered set we just mean a valid entailment $A \leq B$, which can also be identified with the corresponding set of elements $[A, B] = \{C \mid A \leq C \leq B\}$ (a poset with minimum and maximum elements). For example, the Tamari lattice Y_3 displayed in Figure 1 contains 13 intervals. Chapoton used generating function techniques to show that (1) gives the number of intervals in Y_n , and we will explain how the above coherence theorem can be used to give a new proof of this result. As Chapoton mentions, though, the formula itself did not come out of thin air, but rather was found by querying the *On-Line Encyclopedia of Integer Sequences* (OEIS) [27]. Formula (1) is included in OEIS entry A000260, and in fact it was derived over half a century ago by the graph theorist Bill Tutte [31] for a seemingly unrelated family of objects: it counts the number of (3-connected, rooted) triangulations of the sphere with $3(n+1)$ edges. The same formula is also known to count other natural families of embedded graphs, and in particular it counts the number of bridgeless rooted planar maps with n edges [32].² Sparked by Chapoton’s observation, Bernardi and Bonichon [3] found an explicit bijection between intervals of the Tamari order and 3-connected rooted planar triangulations, and quite recently, Fang [8] has proposed new bijections between these three different families of objects (i.e., between 3-connected rooted planar triangulations, bridgeless rooted planar maps, and Tamari intervals).

²A *rooted planar map* is a connected graph embedded in the 2-sphere or the plane, with one half-edge chosen as the root. A (rooted planar) *triangulation* (dually, trivalent map) is a (rooted planar) map in which every face (dually, vertex) has degree three. A map is said to be *bridgeless* (respectively, *3-connected*) if it has no edge (respectively, pair of vertices) whose removal disconnects the underlying graph. (Cf. [18].)

Meanwhile, in [35], Alain Giorgetti and I gave a bijection between rooted planar maps and a simple fragment of linear lambda calculus consisting of the β -normal planar terms. (Here, “planarity” of a lambda term essentially means that the order in which variables are used is fixed following a stack discipline; we will discuss the precise definition of planarity and its potential variations later on.) As with Chapoton’s result, this connection between maps and lambda calculus was found using hints from the OEIS, since the sequence enumerating rooted planar maps was already known – and once again this sequence was first computed by Tutte, who derived another simple closed formula for the number of rooted planar maps with n edges ($\frac{2(2n)!3^n}{n!(n+2)!}$). It is not difficult to check that the bijection described in [35] restricts to a bijection between bridgeless rooted planar maps and β -normal planar terms *with no closed proper subterms*. This restriction on a (not necessarily β -normal or planar) term was called “indecomposability” in [34], where it was used to give a characterization of bridgeless rooted trivalent maps as indecomposable linear lambda terms (and, in turn, to reformulate the Four Color Theorem as a statement about indecomposable planar terms). In any case, this property of the bijection in [35] means that formula (1) also enumerates indecomposable β -normal planar terms by size, and a natural question is whether there is a direct bijection between such terms and intervals of the Tamari order.

As we will explain, in a fairly natural way, every closed (not necessarily β -normal) indecomposable planar term induces both an *application tree* (describing its underlying applicative structure) and a *binding tree* (describing its underlying binding structure, that is, the matching between abstractions and variables). Well, it so happens that the binding tree is always below the application tree in the Tamari order! Trying to prove this fact by induction leads directly to consideration of sequents $\Gamma \rightarrow A$, because the binding structure of an *open* indecomposable planar term (with an arbitrary number of free variables) is naturally described as a *list* of trees. We can then easily build by induction a mapping

$$M \mapsto \Gamma_M \xrightarrow{\mathcal{D}_M} A_M$$

from indecomposable planar terms M to derivations \mathcal{D}_M showing that the binding forest Γ_M is below the application tree A_M in the the Tamari order. Composing with the “forgetful” transformation from derivations to their conclusions, we therefore obtain a mapping

$$M \mapsto (\Gamma_M, A_M)$$

from indecomposable planar terms to (generalized) intervals. One can show that this mapping is surjective, but not injective. This is where the coherence theorem comes in: by inspection, the derivation \mathcal{D}_M is focused if and only if the term M is β -normal, and therefore, the mapping $M \mapsto (\Gamma_M, A_M)$ is bijective from indecomposable β -normal planar terms to Tamari intervals.

D. Outline

The remainder of the paper is organized as follows. In Section II we establish all of the proof-theoretic properties claimed above, including soundness and completeness, focusing, and coherence of the sequent calculus for the Tamari order. In Section III we concisely discuss how the coherence theorem can be applied to give a new proof of formula (1) for the number of intervals in Y_n , simplifying Chapoton’s original proof. Finally, in Section IV we recall some basic lambda calculus notions, then turn to the combinatorics of linear lambda terms, and explain how to construct the aforementioned bijection between Tamari intervals and indecomposable β -normal planar terms.

II. A SEQUENT CALCULUS FOR THE TAMARI ORDER

A. Definitions and terminology

For reference, we recall here the definition of the sequent calculus introduced in I-B, and clarify some notational conventions. The four rules of the sequent calculus are:

$$\frac{A, B, \Delta \rightarrow C}{A * B, \Delta \rightarrow C} *L \quad \frac{\Gamma \rightarrow A \quad \Delta \rightarrow B}{\Gamma, \Delta \rightarrow A * B} *R$$

$$\frac{}{A \rightarrow A} id \quad \frac{\Theta \rightarrow A \quad \Gamma, A, \Delta \rightarrow B}{\Gamma, \Theta, \Delta \rightarrow B} cut$$

Uppercase Latin letters (A, B, \dots) range over **formulas**, which can be either **compound** ($A*B$) or **atomic** (ranged over by lowercase Latin letters p, q, \dots). Uppercase Greek letters Γ, Δ, \dots range over **contexts**, which are (possibly empty) lists of formulas, with concatenation of contexts indicated by a comma. (Let us emphasize that as in Lambek’s system [16] but in contrast to Gentzen’s original sequent calculus [11], there are no rules of “weakening”, “contraction”, or “exchange”, so the order and the number of occurrences of a formula within a context matters.) A **sequent** is a pair of a context Γ and a formula A .

Abstractly, a **derivation** is a tree (formally, a *rooted planar tree with boundary*, cf. [14]) whose internal nodes are labelled by the names of rules and whose edges are labelled by sequents satisfying the constraints indicated by the given rule. The **conclusion** of a derivation is the sequent labelling its outgoing root edge, while its **premises** are the sequents labelling any incoming leaf edges. A derivation with no premises is said to be **closed**.

We write “ $\Gamma \rightarrow A$ ” as a notation for sequents, but also sometimes as a shorthand to indicate that the given sequent is **derivable** using the above rules, in other words that there exists a closed derivation whose conclusion is that sequent (it will always be clear which of these two senses we mean). Sometimes we will need to give an explicit name to a derivation with a given conclusion, in which case we place it over the sequent arrow.

As in I-B, when constructing derivations we sometimes write L and R as shorthand for $*L$ and $*R$, and usually don’t bother labelling the instances of *id* and *cut* since they are clear from context.

Finally, define the **frontier** $\text{fr}(A)$ of a formula A to be the ordered list of atoms occurring in A (i.e., by $\text{fr}(A * B) = \text{fr}(A), \text{fr}(B)$ and $\text{fr}(p) = p$), and the frontier of a context $\Gamma = A_1, \dots, A_n$ as the concatenation of frontiers $\text{fr}(\Gamma) = \text{fr}(A_1), \dots, \text{fr}(A_n)$. The following properties are immediate by examination of the four sequent calculus rules.

Proposition II.1. *Suppose that $\Gamma \longrightarrow A$. Then*

- 1) (*Refinement:*) $\text{fr}(\Gamma) = \text{fr}(A)$.
- 2) (*Relabelling:*) $\sigma\Gamma \longrightarrow \sigma A$, where σ is any relabelling function on atoms.

B. Completeness

We begin by establishing completeness relative to the Tamari order, which is the easier direction.

Theorem II.2 (Completeness). *If $A \leq B$ then $A \longrightarrow B$.*

Proof. We must show that the relation $A \longrightarrow B$ is reflexive and transitive, and that the multiplication operation satisfies a semi-associative law and is monotonic in each argument. All of these properties are straightforward:

- 1) Reflexivity: immediate by *id*.
- 2) Transitivity: immediate by *cut*.
- 3) Semi-associativity:

$$\frac{\frac{\frac{A \longrightarrow A \quad \overline{B \longrightarrow B} \quad \overline{C \longrightarrow C}}{B, C \longrightarrow B * C} R}{A, B, C \longrightarrow A * (B * C)} R}{\frac{A * B, C \longrightarrow A * (B * C)}{(A * B) * C \longrightarrow A * (B * C)} L} L$$

- 4) Monotonicity:

$$\frac{\frac{A \longrightarrow A' \quad \overline{B \longrightarrow B}}{A, B \longrightarrow A' * B} R}{A * B \longrightarrow A' * B} L \quad \frac{\frac{\overline{A \longrightarrow A} \quad \overline{B \longrightarrow B'}}{A, B \longrightarrow A * B'} R}{A * B \longrightarrow A * B'} L$$

□

C. Soundness

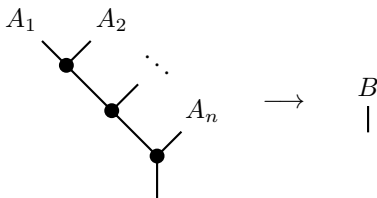
To prove soundness relative to the Tamari order, first we have to explain the interpretation of general sequents. The basic idea is that we can interpret a *non-empty context* as a *left-associated product*. Thus, a general sequent of the form

$$A_1, A_2, \dots, A_n \longrightarrow B$$

(where $n \geq 1$) is interpreted as an entailment of the form

$$(\dots(A_1 * A_2) * \dots) * A_n \leq B$$

in the Tamari order. Visualizing everything in terms of binary trees, the sequent can be interpreted like so:



That is, the context provides information about the *left-branching spine* of the tree which is below in the Tamari order.

Let $\phi[-]$ be the operation taking any non-empty context Γ to a formula $\phi[\Gamma]$ by the above interpretation. The operation is defined by the following equations:

$$\begin{aligned} \phi[A] &= A \\ \phi[\Gamma, A] &= \phi[\Gamma] * A \end{aligned}$$

Critical to soundness of the sequent calculus is the following “colax” property of $\phi[-]$:

Proposition II.3. $\phi[\Gamma, \Delta] \leq \phi[\Gamma] * \phi[\Delta]$ for all non-empty contexts Γ and Δ .

Proof. By induction on Δ . The case of a singleton context $\Delta = A$ is immediate. Otherwise, if $\Delta = (\Delta', A)$, we have

$$\begin{aligned} \phi[\Gamma, \Delta', A] &= \phi[\Gamma, \Delta'] * A \\ &\leq (\phi[\Gamma] * \phi[\Delta']) * A \\ &\leq \phi[\Gamma] * (\phi[\Delta'] * A) \\ &= \phi[\Gamma] * \phi[\Delta', A] \end{aligned}$$

where the first inequality is by the inductive hypothesis and monotonicity, while the second inequality is by the semi-associative law. □

The operation $\phi[-]$ can also be equivalently described in terms of a *right action* $A \otimes \Delta$ of an arbitrary context on a formula, where this action is defined by the following equations:

$$\begin{aligned} A \otimes \cdot &= A \\ A \otimes (\Delta, B) &= (A \otimes \Delta) * B \end{aligned}$$

We will make use of a few simple properties of $- \otimes \Delta$:

Proposition II.4. $\phi[\Gamma, \Delta] = \phi[\Gamma] \otimes \Delta$ for all non-empty contexts Γ and arbitrary contexts Δ .

Proposition II.5 (Monotonicity). *If $A \leq A'$ then $A \otimes \Delta \leq A' \otimes \Delta$.*

Proof. Both properties are immediate by induction on Δ , where in the case of Prop. II.5 we apply monotonicity of the operations $- * B$. □

We are now ready to prove soundness.

Theorem II.6 (Soundness). *If $\Gamma \longrightarrow A$ then $\phi[\Gamma] \leq A$.*

Proof. By induction on the (closed) derivation of $\Gamma \longrightarrow A$. There are four cases, corresponding to the four rules of the sequent calculus:

- (Case **L*): The derivation ends in

$$\frac{A, B, \Delta \longrightarrow C}{A * B, \Delta \longrightarrow C} *L$$

By induction we have $\phi[A, B, \Delta] \leq C$, but by Prop. II.4 we have $\phi[A * B, \Delta] = \phi[A * B] \otimes \Delta = (A * B) \otimes \Delta = \phi[A, B] \otimes \Delta = \phi[A, B, \Delta]$.

- (Case $*R$): The derivation ends in

$$\frac{\Gamma \longrightarrow A \quad \Delta \longrightarrow B}{\Gamma, \Delta \longrightarrow A * B} *R$$

By induction we have $\phi[\Gamma] \leq A$ and $\phi[\Delta] \leq B$, hence

$$\phi[\Gamma, \Delta] \leq \phi[\Gamma] * \phi[\Delta] \leq A * B$$

where we apply Prop. II.3 for the first inequality, and monotonicity in both arguments for the second.

- (Case id): Immediate by reflexivity.
- (Case cut): The derivation ends in

$$\frac{\Theta \longrightarrow A \quad \Gamma, A, \Delta \longrightarrow B}{\Gamma, \Theta, \Delta \longrightarrow B} cut$$

We can reason like so:

$$\begin{aligned} \phi[\Gamma, \Theta, \Delta] &= \phi[\Gamma, \Theta] * \Delta && \text{(II.4)} \\ &\leq (\phi[\Gamma] * \phi[\Theta]) * \Delta && \text{(II.3 + monotonicity)} \\ &\leq (\phi[\Gamma] * A) * \Delta && \text{(i.h. + monotonicity)} \\ &= \phi[\Gamma, A, \Delta] && \text{(II.4)} \\ &\leq B && \text{(i.h.)} \end{aligned} \quad \square$$

D. Focusing completeness

Cut-elimination theorems are a staple of proof theory, and often provide a rich source of information about a given logic. In this section we will prove a *focusing completeness* theorem, which is an even stronger form of cut-elimination originally formulated by Andreoli in the setting of linear logic [1].

Definition II.7. A context Γ is said to be **reducible** if its leftmost formula is compound, and **irreducible** otherwise. A sequent $\Gamma \longrightarrow A$ is said to be:

- **left-inverting** if Γ is reducible;
- **right-focusing** if Γ is irreducible and A is compound;
- **atomic** if Γ is irreducible and A is atomic.

Proposition II.8. Any sequent is either left-inverting, right-focusing, or atomic.

Definition II.9. A closed derivation \mathcal{D} is said to be **focused** if left-inverting sequents only appear as the conclusions of $*L$, right-focusing sequents only as the conclusions of $*R$, and atomic sequents only as the conclusions of id .

We write “ Γ^{irr} ” to indicate that a context Γ is irreducible.

Proposition II.10. A closed derivation is focused if and only if it is constructed using only $*L$ and the following restricted forms of $*R$ and id (and no instances of cut):

$$\frac{\Gamma^{\text{irr}} \longrightarrow A \quad \Delta \longrightarrow B}{\Gamma^{\text{irr}}, \Delta \longrightarrow A * B} *R^{\text{foc}} \quad \frac{}{p \longrightarrow p} id^{\text{atm}}$$

Example 1. One way to derive

$$((p * q) * r) * s \longrightarrow p * ((q * r) * s)$$

is by cutting together the two derivations

$$\frac{\mathcal{S}\mathcal{A}_{p,q,r} \quad \frac{(p * q) * r \longrightarrow p * (q * r) \quad s \longrightarrow s}{(p * q) * r, s \longrightarrow (p * (q * r)) * s} R}{((p * q) * r) * s \longrightarrow (p * (q * r)) * s} L$$

and

$$\mathcal{S}\mathcal{A}_{p,q*r,s} \quad \frac{}{(p * (q * r)) * s \longrightarrow p * ((q * r) * s)}$$

where $\mathcal{S}\mathcal{A}_{A,B,C}$ is the derivation of the semi-associative law $(A * B) * C \longrightarrow A * (B * C)$ from the proof of Theorem II.2. Clearly this is *not* a focused derivation (besides the cut rule, it also uses instances of $*R$ and id with a left-inverting conclusion). However, it is possible to give a focused derivation of the same sequent:

$$\frac{\frac{\frac{q \longrightarrow q \quad r \longrightarrow r}{q, r \longrightarrow q * r} R \quad \frac{}{s \longrightarrow s} R}{p \longrightarrow p} R \quad \frac{p, q, r, s \longrightarrow p * ((q * r) * s)}{p * q, r, s \longrightarrow p * ((q * r) * s)} L}{(p * q) * r, s \longrightarrow p * ((q * r) * s)} L \quad \frac{}{((p * q) * r) * s \longrightarrow p * ((q * r) * s)} L$$

In the below, we write “ $\Gamma \Longrightarrow A$ ” as a shorthand notation to indicate that $\Gamma \longrightarrow A$ has a (closed) focused derivation, and “ $\mathcal{D} : A \Longrightarrow B$ ” to indicate that \mathcal{D} is a particular focused derivation of $A \longrightarrow B$.

Theorem II.11 (Foc. comp’ness). If $\Gamma \longrightarrow A$ then $\Gamma \Longrightarrow A$.

To prove the focusing completeness theorem, it suffices to show that the cut rule as well as the unrestricted forms of id and $*R$ are all *admissible* for focused derivations, in the proof-theoretic sense that given focused derivations of their premises, we can obtain a focused derivation of their conclusion. We begin by proving a *focused deduction* lemma (cf. [25]), which entails the admissibility of id , then show cut and $*R$ in turn.

Lemma II.12 (Deduction). If $\Gamma^{\text{irr}} \Longrightarrow A$ implies $\Gamma^{\text{irr}}, \Delta \Longrightarrow B$ for all Γ^{irr} , then $A, \Delta \Longrightarrow B$. In particular, $A \Longrightarrow A$.

Proof. By induction on the formula A :

- (Case $A = p$): Immediate by assumption, taking $\Gamma = p$ and $p \longrightarrow p$ derived by the id^{atm} rule.
- (Case $A = A_1 * A_2$): By composing with the $*L$ rule,

$$\frac{A_1, A_1, \Delta \longrightarrow B}{A_1 * A_2, \Delta \longrightarrow B} *L$$

we reduce the problem to showing $A_1, A_2, \Delta \Longrightarrow B$, and by the i.h. on A_1 it suffices to show that $\Gamma_1^{\text{irr}} \Longrightarrow A_1$ implies $\Gamma_1^{\text{irr}}, A_2, \Delta \Longrightarrow B$ for all contexts Γ_1^{irr} . Let $\mathcal{D}_1 : \Gamma_1^{\text{irr}} \Longrightarrow A_1$. We can derive $\Gamma_1^{\text{irr}}, A_2 \longrightarrow A_1 * A_2$ by

$$\mathcal{D} = \frac{\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\Gamma_1^{\text{irr}} \longrightarrow A_1 \quad A_2 \longrightarrow A_2} *R}{\Gamma_1^{\text{irr}}, A_2 \longrightarrow A_1 * A_2} *R$$

where we apply the i.h. on A_2 to obtain \mathcal{D}_2 . Finally, applying the assumption to \mathcal{D} (with $\Gamma^{\text{irr}} = \Gamma_1^{\text{irr}}, A_2$) we obtain the desired derivation of $A_1, A_2, \Delta \rightarrow B$. \square

Lemma II.13 (Cut). *If $\Theta \Rightarrow A$ and $\Gamma, A, \Delta \Rightarrow B$ then $\Gamma, \Theta, \Delta \Rightarrow B$.*

Proof. Let $\mathcal{D} : \Theta \Rightarrow A$ and $\mathcal{E} : \Gamma, A, \Delta \Rightarrow B$. We proceed by a lexicographic induction, first on the cut formula A and then on the pair of derivations $(\mathcal{D}, \mathcal{E})$ (i.e., at each inductive step of the proof, either A gets smaller, or it stays the same as one of \mathcal{D} or \mathcal{E} gets smaller while the other stays the same).

In the case that $A = p$ we can apply the ‘‘frontier refinement’’ property (Prop. II.1) to deduce that $\Theta = p$, so the cut is trivial and we just reuse the derivation $\mathcal{E} : \Gamma, p, \Delta \Rightarrow B$. Otherwise we have $A = A_1 * A_2$ for some A_1, A_2 , and we proceed by case-analyzing the root rule of \mathcal{E} :

- (Case id^{atm}): Impossible since A is non-atomic.
- (Case $*R^{\text{foc}}$): This case splits in two possibilities:

- 1) $\exists \Delta_1, \Delta_2$ such that $\Delta = \Delta_1, \Delta_2$ and

$$\mathcal{E} = \frac{\frac{\mathcal{E}_1}{\Gamma^{\text{irr}}, A, \Delta_1 \rightarrow B_1} \quad \frac{\mathcal{E}_2}{\Delta_2 \rightarrow B_2}}{\Gamma^{\text{irr}}, A, \Delta_1, \Delta_2 \rightarrow B_1 * B_2} *R$$

- 2) $\exists \Gamma_1^{\text{irr}}, \Gamma_2$ such that $\Gamma^{\text{irr}} = \Gamma_1^{\text{irr}}, \Gamma_2$ and

$$\mathcal{E} = \frac{\frac{\mathcal{E}_1}{\Gamma_1^{\text{irr}} \rightarrow B_1} \quad \frac{\mathcal{E}_2}{\Gamma_2, A, \Delta \rightarrow B_2}}{\Gamma_1^{\text{irr}}, \Gamma_2, A, \Delta \rightarrow B_1 * B_2} *R$$

In the first case, we cut \mathcal{D} with \mathcal{E}_1 to obtain $\Gamma^{\text{irr}}, \Theta, \Delta_1 \Rightarrow B_1$, then recombine that with \mathcal{E}_2 using $*R^{\text{foc}}$ to obtain $\Gamma^{\text{irr}}, \Theta, \Delta_1, \Delta_2 \Rightarrow B_1 * B_2$. The second case is similar.

- (Case $*L$): This case splits into two possibilities:

- 1) $\exists C_1, C_2, \Gamma'$ such that $\Gamma = C_1 * C_2, \Gamma'$ and

$$\mathcal{E} = \frac{\frac{\mathcal{E}'}{C_1, C_2, \Gamma', A, \Delta \rightarrow B}}{C_1 * C_2, \Gamma', A, \Delta \rightarrow B} *L$$

We cut \mathcal{D} into \mathcal{E}' and reapply the $*L$ rule.

- 2) $\Gamma = \cdot$ and

$$\mathcal{E} = \frac{\frac{\mathcal{E}'}{A_1, A_2, \Delta \rightarrow B}}{A_1 * A_2, \Delta \rightarrow B} *L$$

We further analyze the root rule of \mathcal{D} :

- (Case $*L$): $\exists C_1, C_2, \Theta'$ s.t. $\Theta = C_1 * C_2, \Theta'$ and

$$\mathcal{D} = \frac{\frac{\mathcal{D}'}{C_1, C_2, \Theta' \rightarrow A_1 * A_2}}{C_1 * C_2, \Theta' \rightarrow A_1 * A_2} *L$$

We cut \mathcal{D}' into \mathcal{E} and reapply the $*L$ rule.

- (Case id^{atm}): Impossible.

- (Case $*R^{\text{foc}}$): $\exists \Theta_1^{\text{irr}}, \Theta_2$ s.t. $\Theta^{\text{irr}} = \Theta_1^{\text{irr}}, \Theta_2$ and

$$\mathcal{D} = \frac{\frac{\mathcal{D}_1}{\Theta_1^{\text{irr}} \rightarrow A_1} \quad \frac{\mathcal{D}_2}{\Theta_2 \rightarrow A_2}}{\Theta_1^{\text{irr}}, \Theta_2 \rightarrow A_1 * A_2} *R$$

We cut both \mathcal{D}_1 and \mathcal{D}_2 into \mathcal{E}' (the cuts are at smaller formulas so the order doesn't matter). \square

Lemma II.14 ($*R$ admiss.). *If $\Gamma \Rightarrow A$ and $\Delta \Rightarrow B$ then $\Gamma, \Delta \rightarrow A * B$.*

Proof. We can derive $A, B \Rightarrow A * B$ using two instances of the deduction lemma together with the $*R$ rule. Then we obtain $\Gamma, \Delta \Rightarrow A * B$ using two instances of *cut*. \square

Proof of Theorem II.11. An arbitrary closed derivation can be turned into a focused one by starting at the top of the derivation tree and using the above lemmas to interpret any instance of the *cut* rule and of the unrestricted forms of *id* and $*R$. \square

Finally, we mention two simple applications of the focusing completeness theorem.

Proposition II.15 (Frontier invariance). *Let σ be any relabelling function on atoms. Then $\Gamma \rightarrow A$ if and only if $\text{fr}(\Gamma) = \text{fr}(A)$ and $\sigma\Gamma \rightarrow \sigma A$.*

Proof. The forward direction is Prop. II.1. For the backward direction we use induction on focused derivations, which is justified by Theorem II.11. The only interesting case is $*R^{\text{foc}}$, where we can assume $\text{fr}(\Gamma, \Delta) = \text{fr}(A * B)$ and $\sigma\Gamma \Rightarrow \sigma A$ ($\sigma\Gamma$ irreducible) and $\sigma\Delta \Rightarrow \sigma B$. By Prop. II.1 we have $\text{fr}(\sigma\Gamma) = \text{fr}(\sigma A)$ and $\text{fr}(\sigma\Delta) = \text{fr}(\sigma B)$, but then elementary properties of lists imply that $\text{fr}(\Gamma) = \text{fr}(A)$ and $\text{fr}(\Delta) = \text{fr}(B)$, from which $\Gamma \Rightarrow A$ (Γ irreducible) and $\Delta \Rightarrow B$ follow by the induction hypothesis, hence $\Gamma, \Delta \Rightarrow A * B$. \square

If we let $\sigma = _ \mapsto p$ be any constant relabelling function, then speaking in terms of the Tamari order, Proposition II.15 says that to check that two ‘‘fully-bracketed words’’ (a.k.a., formulas) are related, it suffices to check that their frontiers are equal and that the unlabelled binary trees describing their underlying multiplicative structure are related. Although this fact is intuitively obvious, trying to prove it directly by induction on general derivations fails, because in the case of the *cut* rule we cannot assume anything about the frontier of the cut formula A .

Definition II.16. We say that an irreducible context Γ^{irr} is a **maximal decomposition** of A if $\Gamma^{\text{irr}} \rightarrow A$, and for any other $\Theta^{\text{irr}}, \Theta^{\text{irr}} \rightarrow A$ implies $\Theta^{\text{irr}} \rightarrow \phi[\Gamma^{\text{irr}}]$.

Proposition II.17. *If Γ^{irr} is a maximal decomposition of A , then $A, \Delta \rightarrow B$ if and only if $\Gamma^{\text{irr}}, \Delta \rightarrow B$.*

Proof. The forward direction is by cutting with $\Gamma^{\text{irr}} \rightarrow A$, the backwards direction is by the deduction lemma (II.12) and the universal property of Γ^{irr} . \square

Proposition II.18. *Let $\psi[A]$ be the irreducible context defined inductively by:*

$$\psi[p] = p \quad \psi[A * B] = \psi[A], B$$

Then $\psi[A]$ is a maximal decomposition of A .

Proof. We construct $\psi[A] \longrightarrow A$ by induction on A , and prove the universal property of $\psi[A]$ by induction on focused derivations of $\Delta \longrightarrow A$. \square

Proposition II.19. *$\phi[\psi[A]] = A$ and $\psi[\phi[\Theta^{\text{irr}}]] = \Theta^{\text{irr}}$.*

The maximal decomposition $\psi[A]$ of A is essentially the same thing as what Chapoton [5] calls a “décomposition maximale” of a binary tree. The logical characterization expressed in Defn. II.16 is quite general, though, and is familiar from studies of focusing in other settings (cf. [33]).

E. The coherence theorem

We now come to our main result:

Theorem II.20 (Coherence). *Every derivable sequent has exactly one focused derivation.*

Coherence is a direct consequence of focusing completeness and the following lemma:

Lemma II.21. *For any context Γ and formula A , there is at most one focused derivation of $\Gamma \longrightarrow A$.*

Proof. We proceed by a well-founded induction on sequents, which can be reduced to multiset induction as follows. Define the **size** $|A|$ of a formula A by

$$|A * B| = 1 + |A| + |B| \quad |p| = 0$$

(That is, $|A|$ counts the number of multiplication operations occurring in A .) Then any sequent $A_1, \dots, A_n \longrightarrow B$ induces a multiset of size $(\uplus_{i=1}^n |A_i|) \uplus |B|$, and at each step of our induction this multiset will decrease in the multiset ordering.

There are three cases:

- (A left-inverting sequent $A * B, \Delta \longrightarrow C$): Any focused derivation must end in $*L$, so we apply the i.h. to $A, B, \Delta \longrightarrow C$.
- (A right-focusing sequent $\Gamma^{\text{irr}} \longrightarrow A * B$): Any focused derivation must end in $*R$, and decide some splitting of the context into contiguous pieces Γ_1^{irr} and Δ_2 . However, Γ_1^{irr} and Δ_2 are uniquely determined by frontier refinement ($\text{fr}(\Gamma_1^{\text{irr}}) = \text{fr}(A)$ and $\text{fr}(\Delta_2) = \text{fr}(B)$) and the equation $\Gamma^{\text{irr}} = \Gamma_1^{\text{irr}}, \Delta_2$. So, we apply the i.h. to $\Gamma_1^{\text{irr}} \longrightarrow A$ and $\Delta_2 \longrightarrow B$.
- (An atomic sequent $\Gamma^{\text{irr}} \longrightarrow p$): The sequent has exactly one focused derivation if and only if $\Gamma^{\text{irr}} = p$. \square

Proof of Theorem II.20. By Theorem II.11 and Lemma II.21. \square

F. Notes

The coherence theorem says in a sense that focused derivations provide a canonical representation for intervals of the Tamari order. Although the representations are quite different, in this respect it seems roughly comparable to the “unicity of maximal chains” that was established by Tamari and Friedman as part of their original proof of the lattice property of Y_n [10], [30]. A natural question is whether the sequent calculus can be used to better understand and further simplify the proofs (cf. [12] [22, §4]) of this lattice property.

An easy observation is that one obtains the dual Tamari order (cf. Footnote 1) via a dual restriction of Lambek’s original rule, in other words by requiring the product formula to appear at the *rightmost* end of the context. These two forms of product might also be considered in combination with left and right units, or in combination with Lambek’s left and right division operations. Interestingly, Lambek also introduced a fully non-associative version of his original “syntactic calculus” [17].

The name “coherence theorem” for Theorem II.20 is inspired by the terminology from category theory and Mac Lane’s coherence theorem for monoidal categories [19]. Laplaza [21] extended Mac Lane’s coherence theorem to the situation (very close to Tamari’s) where there is no monoidal unit and the *associator* $\alpha_{A,B,C} : (A \otimes B) \otimes C \rightarrow A \otimes (B \otimes C)$ is only a natural transformation rather than an isomorphism. (In the presence of units, this gives rise to the notion of a *skew monoidal category* [15].) The precise relationship with our coherence theorem remains to be clarified.

III. COUNTING INTERVALS IN TAMARI LATTICES

In this section we explain how the coherence theorem can be used to give a new proof of Chapoton’s result (mentioned in the Introduction) that the number of intervals in Y_n is given by Tutte’s formula (1) for planar triangulations. We will assume some basic familiarity with generating functions (say, as provided by a combinatorics textbook like [9]).

The problem of “counting intervals” is to compute the cardinality of the set

$$\mathcal{I}_n = \{ (A, B) \in Y_n \times Y_n \mid A \leq B \}$$

as a function of n . By the soundness and completeness theorems as well as the frontier invariance property (Prop. II.15), each Y_n is isomorphic as a partial order to the set of formulas A of size n with any fixed frontier of length $n + 1$ (remember that a binary tree with n internal nodes has $n + 1$ leaves), ordered by sequent derivability. By the coherence theorem, the problem of counting intervals can therefore be reduced to the problem of *counting focused derivations*.

This problem lends itself readily to being solved using generating functions. Consider the generating functions $L(z, x)$ and $R(z, x)$ defined as formal power series $L(z, x) = \sum_{k,n \in \mathbb{N}} \ell_{k,n} x^k z^n$ and $R(z, x) = \sum_{k,n \in \mathbb{N}} r_{k,n} x^k z^n$, where $\ell_{k,n}$ (respectively, $r_{k,n}$) is the number of focused derivations of sequents whose left-hand side is a context (respectively, irreducible context) of length k and whose right-hand side is a formula of size n . (Without loss of generality in this analysis,

we assume that all formulas A of size n have a fixed frontier $\text{fr}(A) = p^{n+1}$.) We write $L_1(z)$ to denote the coefficient of x^1 in $L(z, x)$.

Proposition III.1. $L_1(z)$ is the ordinary generating function counting Tamari intervals by size.

Proof. The coefficients $\ell_{1,n}$ give the number of focused derivations of sequents of the form $A \Rightarrow B$, where $|B| = n$ (and hence $|A| = n$), so $\ell_{1,n} = |\mathcal{I}_n|$ by Theorem II.20. \square

Proposition III.2. L and R satisfy the equations:

$$L(z, x) = \frac{L(z, x) - xL_1(z)}{x} + R(z, x) \quad (2)$$

$$R(z, x) = zR(z, x)L(z, x) + x \quad (3)$$

Proof. The equations are derived directly from the inductive structure of focused derivations:

- The first summand in (2) corresponds to the contribution from the $*L$ rule, which transforms any $A, B, \Gamma \Rightarrow C$ into $A*B, \Gamma \Rightarrow C$. The context in the premise must have length ≥ 2 which is why we subtract the $xL_1(x)$ factor, and the context in the conclusion is one formula shorter which is why we divide by x . The second summand is the contribution from irreducible contexts.
- The first summand in (3) corresponds to the contribution from the $*R^{\text{foc}}$ rule, which transforms $\Gamma^{\text{irr}} \Rightarrow A$ and $\Delta \Rightarrow B$ into $\Gamma^{\text{irr}}, \Delta \Rightarrow A * B$: the length of the context in the conclusion is the sum of the lengths of Γ^{irr} and Δ , while the size of $A * B$ is one plus the sum of the sizes of A and B , which is why we multiply R and L together and then by an extra factor of z . The second summand is the contribution from $id^{\text{atm}} : p \Rightarrow p$. \square

Proposition III.3. $L_1(z) = R(z, 1)$.

Proof. This follows immediately from (2), but we can interpret it constructively as well. The coefficient of z^n in $R(z, 1)$ is the formal sum $\sum_k r_{k,n}$, giving the number of focused derivations of sequents whose right-hand side is a formula of size n and whose left-hand side is an irreducible context of arbitrary length. But by Props. II.17–II.19, the operations $\phi[-]$ and $\psi[-]$ realize a 1-to-1 correspondence between derivable sequents of the form $\Gamma^{\text{irr}} \rightarrow B$ and ones of the form $A \rightarrow B$. \square

After substituting $L_1(z) = R(z, 1)$ into (2) and applying a bit of algebra, we obtain another formula for L in terms of a “discrete difference operator” acting on R :

$$L(z, x) = x \frac{R(z, x) - R(z, 1)}{x - 1} \quad (4)$$

The recursive (or “functional”) equations (3) and (4) can be easily unrolled (preferably using computer algebra software) to compute the first few dozen coefficients of R and L :

$$R(z, x) = x + x^2z + (x^2 + 2x^3)z^2 + (3x^2 + 5x^3 + 5x^4)z^3 + (13x^2 + 20x^3 + 21x^4 + 14x^5)z^4 + (68x^2 + 100x^3 + 105x^4 + 84x^5 + 42x^6)z^5 + \dots$$

$$L_1(z) = R(z, 1) = 1 + z + 3z^2 + 13z^3 + 68z^4 + 399z^5 + 2530z^6 + 16965z^7 + \dots$$

Theorem III.4 (Chapoton [5]). $|\mathcal{I}_n| = \frac{2(4n+1)!}{(n+1)!(3n+2)!}$.

Proof. At this point, we can directly appeal to results of Cori and Schaeffer, because equations (3) and (4) are a special case of the functional equations given in [7] for the generating functions of *description trees of type* (a, b) , where $a = b = 1$. Cori and Schaeffer explained how to solve these equations abstractly for $R(z, 1)$ using Brown and Tutte’s “quadratic method”, and then how to derive the explicit formula above in the specific case that $a = b = 1$ via Lagrange inversion (essentially as the formula was originally derived by Tutte for planar triangulations). \square

Let’s take a moment to discuss Chapoton’s original proof of Theorem III.4, which it should be emphasized is actually not all that different from the one given here. Chapoton likewise defines a two-variable generating function $\Phi(z, x)$ enumerating intervals in the Tamari lattices \mathcal{Y}_n , where the parameter z keeps track of n , and the parameter x keeps track of the *number of segments along the left border of the tree at the lower end of the interval*.³ By a combinatorial analysis, Chapoton derives the following functional equation for Φ :

$$\Phi(z, x) = x^2z(1 + \Phi(z, x)/x) \left(1 + \frac{\Phi(z, x) - \Phi(z, 1)}{x - 1} \right) \quad (5)$$

He manipulates this equation and eventually solves for $\Phi(z, 1)$ as the root of a certain polynomial, from which he derives Tutte’s formula (1), again by appeal to another result in the paper by Cori and Schaeffer [7].

If we give a bit of thought to these definitions, it is easy to see that the number of segments along the left border of a tree (= formula) A is equal to the length of its maximal decomposition $\psi[A]$ – meaning that the generating function $\Phi(z, x)$ apparently contains exactly the same information as $R(z, x)$! There is a small technicality, however, due to the fact that Chapoton only considers the \mathcal{Y}_n for $n \geq 1$. In fact, the two generating functions are related by a small offset (corresponding to the coefficient of z^0 in $R(z, x)$):

$$\Phi(z, x) = R(z, x) - x \quad (6)$$

Indeed, it can be readily verified that equation (5) follows from (3) and (4), applying the substitution $R(z, x) = x + \Phi(z, x)$.

IV. THE INTERPRETATION OF INDECOMPOSABLE PLANAR LAMBDA TERMS BY TAMARI INTERVALS

A. Preliminaries

We recall a few basic definitions from lambda calculus [2]. A **term** (ranged over by uppercase Latin letters M, N, \dots) is either a **variable** (ranged over by lowercase Latin letters x, y, \dots) or an **application** $M(N)$ (where M and N are terms) or an **abstraction** $\lambda x.M$ (where x is a variable and M is a term). By syntactic convention, abstractions take scope to the right as far as possible, so that for example “ $\lambda x.x(\lambda y.y)$ ”

³Or rather at the upper end, since Chapoton uses the dual convention for orienting the Tamari order (cf. Footnote 1).

should be read as an abstraction term, whereas “ $(\lambda x.x)(\lambda y.y)$ ” is an application term.

We define the **subterms** of a term as follows:

- M is a subterm of itself
- any subterm of M or N is a subterm of $M(N)$
- any subterm of M is a subterm of $\lambda x.M$

The subterms of M are said to “occur” in M . Among variables occurring within a term, we distinguish **free** variables from **bound** variables: an abstraction term $\lambda x.M$ is said to bind any free occurrences of x in M , and otherwise all variables which are not bound by an abstraction are said to be free. A term with no free variables is said to be **closed**. Terms are usually considered up to α -conversion, or renaming of bound variables (e.g., the terms $\lambda x.x$ and $\lambda y.y$ are α -equivalent). We will assume the **Barendregt convention**, which says that all bound variables have distinct names and that no variable is both free and bound (this condition is always possible to achieve via α -conversion).

The basic computation rule of lambda calculus is the rule of β -reduction,

$$(\lambda x.M)(N) \rightarrow M[N/x] \quad (\beta)$$

where $M[N/x]$ denotes the (“capture-avoiding”) substitution of N for any free occurrences of x in M . The β -reduction rule can be performed on any subterm, in other words there are also the following “congruence” rules:

$$\frac{M \rightarrow M'}{M(N) \rightarrow M'(N)} \quad \frac{N \rightarrow N'}{M(N) \rightarrow M(N')} \quad \frac{M \rightarrow M'}{\lambda x.M \rightarrow \lambda x.M'}$$

This defines a rewriting system which is confluent (the Church-Rosser theorem), although there exist infinite reduction sequences $M_1 \rightarrow M_2 \rightarrow \dots$ (since pure lambda calculus is a universal model of computation). Two terms are said to be β -equivalent $M \stackrel{\beta}{=} N$ if they both β -reduce to the same term $M \rightarrow P \leftarrow N$ in any number of steps. A term which cannot be further reduced is said to be β -normal.

An abstraction $\lambda x.M$ is said to be **linear** if the variable x has exactly one free occurrence in M . By extension, a term N is said to be linear if every abstraction subterm of N is linear, and all free variables of N occur exactly once as subterms. For example, the terms $\lambda x.\lambda y.y(x)$ and $\lambda x.x(\lambda y.y)$ are linear, but the terms $\lambda x.x(x)$ and $\lambda x.\lambda y.y$ are not.

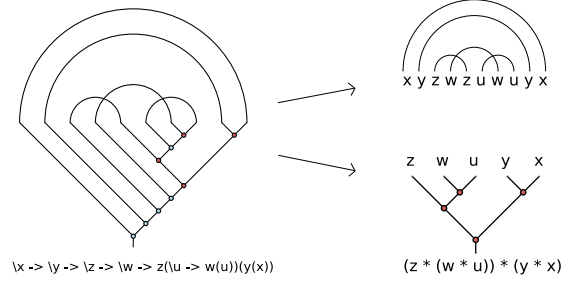
A term is said to be **indecomposable** [34] if it has no closed proper subterms. For example, the term $\lambda x.\lambda y.y(x)$ is indecomposable, but the term $\lambda x.x(\lambda y.y)$ is not.

Proposition IV.1. *A closed indecomposable term is necessarily an abstraction term $\lambda x.M$, where M is indecomposable.*

B. Application trees and binding diagrams

Any linear lambda term is naturally associated with a pair of basic combinatorial objects: a binary tree describing its underlying structure of applications, and a *rooted chord diagram* describing the matching between lambda abstractions

and variables. The quickest way of explaining this is with a picture:



On the left we have a diagram that faithfully represents the linear term $\lambda x.\lambda y.\lambda z.\lambda w.z(\lambda u.w(u))(y(x))$ up to α -equivalence, with nodes representing applications colored in red and nodes representing abstractions colored in blue. (This kind of diagrammatic representation is folklore in lambda calculus; for a more thorough discussion, see [34].) On the right we’ve selectively gotten rid of some of the structure of the term to produce two simpler objects. The *application tree* (depicted in the lower right) collapses abstraction nodes, keeping only the underlying binary tree of applications (with variables labelling the leaves).

Definition IV.2. The **application tree** of a term M is a (leaf-labelled) binary tree $\alpha[M]$, defined by induction as follows:

$$\begin{aligned} \alpha[x] &= x \\ \alpha[M(N)] &= \alpha[M] * \alpha[N] \\ \alpha[\lambda x.M] &= \alpha[M] \end{aligned}$$

On the other hand, the *binding diagram* (depicted in the upper right of the picture) collapses application nodes, recording only the *order* of successive lambda abstractions and variable occurrences (or “uses”). Reading the diagram from left to right, rising arcs correspond to abstractions and falling arcs to uses. The fact that this is the binding diagram of a *closed linear* term means that every rising arc is met by exactly one falling arc, and we get the classical notion of rooted chord diagram (also sometimes referred to as an “arc diagram” [18] or “matching diagram” [4]), which has many equivalent purely combinatorial representations such as by *double-occurrence words* ($x y z w z u w u y x$) or by *fixed point-free involutions* $((1\ 10)(2\ 9)(3\ 5)(4\ 7)(6\ 8))$. More generally, the binding diagram of a linear term with free variables corresponds to a rooted chord diagram *where chords can have unattached ends* (i.e., to an “open matching” in the sense of [4]). Such diagrams can be represented faithfully by “at-most-double-occurrence” words (that is, sequences where every symbol occurs either once or twice), or equivalently by involutions with fixed points.

Definition IV.3. The **binding diagram** of a linear term M is a sequence $\gamma[M]$, defined by induction as follows:

$$\begin{aligned} \gamma[x] &= x \\ \gamma[M(N)] &= \gamma[M], \gamma[N] \\ \gamma[\lambda x.M] &= x, \gamma[M] \end{aligned}$$

Proposition IV.4. *If M is linear, then $\gamma[M]$ is an at-most-double-occurrence word (assuming the Barendregt convention). If moreover M is closed, then $\gamma[M]$ is a double-occurrence word.*

By the isomorphism between rooted chord diagrams and double-occurrence words, we will therefore view the binding diagram of a linear term interchangeably either as an (at-most-)double-occurrence word or as a rooted chord diagram (potentially with unattached chords). We refer to chords with an unattached end (i.e., single-occurrence letters) as *free chords*, and to chords with both ends attached (i.e., double-occurrence letters) as *full chords*.

Let the **size** $|M|$ of a linear term M be defined here as the number of internal nodes in its application tree $\alpha[M]$.

Proposition IV.5. *If M is a closed linear term of size n , then its binding diagram $\gamma[M]$ is a rooted chord diagram with $n + 1$ full chords (i.e., the corresponding double-occurrence word has $2n + 2$ letters). More generally, if M is a linear term of size n with k free variables, then its binding diagram has $n + 1 - k$ full chords and k free chords.*

A rooted chord diagram (with no free chords) is said to be *indecomposable* if it is not the juxtaposition of two rooted chord diagrams – the corresponding condition on a double-occurrence word is that it is not the concatenation of two double-occurrence words (cf. [6], [24]).

Proposition IV.6. *Let M be a closed linear term. If M is indecomposable (i.e., has no closed proper subterms), then $\gamma[M]$ is an indecomposable double-occurrence word.*

(Note the converse is false: for example, $\lambda x.(\lambda y.y)x$ is not indecomposable, but has an indecomposable binding diagram.)

C. Planarity and the lambda rotation (ρ) rule

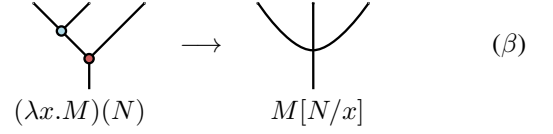
A rooted chord diagram is said to be *planar* if it has no pair of crossing arcs. This translates to at-most-double-occurrence words as follows: γ is planar if for any x, δ, x occurring as a subword, δ is a double-occurrence word.

Definition IV.7. We say that a linear term M is **planar** just in case its binding diagram $\gamma[M]$ is planar.

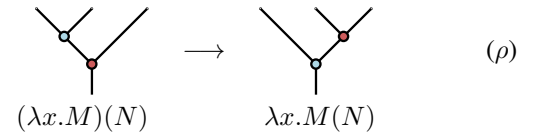
An important comment about this definition is that the class of terms which are considered planar clearly depends upon the “layout convention” we use in drawing diagrams, which is implicit in the definition of the binding diagram $\gamma[M]$. The precise notion of planarity we consider here was studied (in an equivalent formulation) in [35], where it was called “LR-planarity” and contrasted with “RL-planarity” (see [35, §3.1], and also [34, §4]): the latter can be defined by replacing the abstraction case of Defn. IV.3 by $\gamma[\lambda x.M] = \gamma[M], x$. For example, the term $\lambda x.\lambda y.y(\lambda z.z(x))$ is LR-planar, while the term $\lambda x.\lambda y.x(\lambda z.y(z))$ is RL-planar. A simple observation about planarity is that for any given underlying “skeleton” of abstractions and applications sans variable names (i.e., an expression like “ $\lambda_.\lambda_._(\lambda_._(_))$ ”), there is a unique (up to

α -equivalence) way of filling in the variable names to produce either an LR-planar term or an RL-planar term.

Even though the difference is seemingly trivial, the asymmetry of the lambda calculus means that these two notions of planarity have very different properties. Notably, the set of LR-planar terms is *not* closed under β -reduction, which diagrammatically (under these conventions) corresponds to the following operation:



On the other hand, LR-planar terms *are* closed under a “colored” version of the right rotation operation, which rotates a lambda abstraction out from the left of an application:



The ρ -reduction rule is certainly not a typical lambda calculus rule – but the amusing coincidence is that the set of β -normal terms and the set of ρ -normal terms coincide. This is true even though the induced notions of equivalence are very different: let us say that two terms are ρ -equivalent ($M =^\rho N$) just in case they both reduce to a common term via some sequence of ρ -reductions (applied to any subterms). The following observation is key: it says that the pair of the application tree and binding diagram of a linear lambda term form a *complete invariant* for that term up to ρ -equivalence.

Theorem IV.8. *For all linear terms M and N , $M =^\rho N$ if and only if $\alpha[M] = \alpha[N]$ and $\gamma[M] = \gamma[N]$.*

Proof. For the forward direction, we check that ρ -reduction preserves application trees and binding diagrams, which is evident by inspection of the rule. For the backward direction, we first verify that every term has a ρ -normal form (obvious, since any lambda abstraction can be rotated only finitely many times), then that any two ρ -normal forms with equal application trees and binding diagrams must be equal. Here we can use the fact that ρ -normal forms (which, again, are identical with β -normal forms) have a simple inductive structure. In particular, we can read off the *head normal form* $\lambda x_1 \dots \lambda x_n.((xM_1) \dots M_p)$ of a term by examining the leading arcs of its binding diagram (in this case, n rising arcs x_1, \dots, x_n followed by a falling arc x) together with the left-branching spine of its application tree (in this case, with the leftmost leaf labelled x , and p subtrees along the right). Linearity determines how the rest of the binding diagram is partitioned among the subterms M_1, \dots, M_p , and we proceed by induction. \square

D. From indecomposable planar terms to Tamari intervals

In this section we at last present the bijection between (ρ/β)-normal planar indecomposable terms and Tamari intervals,

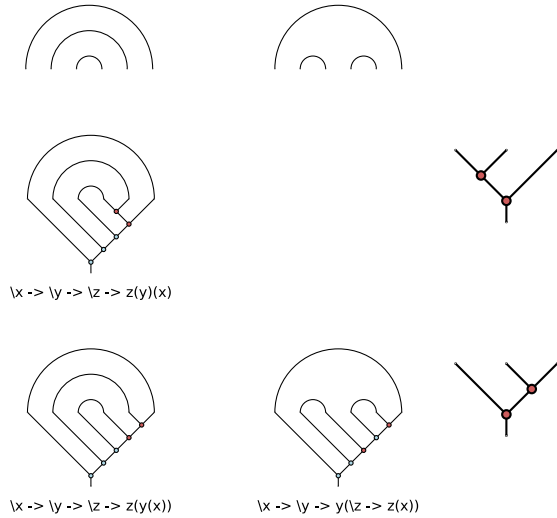


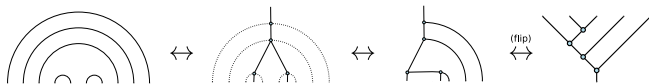
Fig. 2. The three indecomposable normal planar terms of size two, organized into rows according to their underlying application trees, and into columns by their underlying binding diagrams.

beginning first by explaining how to map any planar indecomposable term to an interval. It is well-known that rooted planar chord diagrams are enumerated by the Catalan numbers: there are C_n rooted planar chord diagrams with n chords, and hence they can be placed in bijection with binary trees with n internal nodes. Since we were hoping to extract a pair of binary trees and we already have the application tree, that seems like a good start. Remember, though, that a term with an application tree of size n has $n + 1$ chords in its binding diagram (Prop. IV.5), so there is an apparent indexing mismatch that has to be accounted for.

Proposition IV.9. *Any indecomposable planar double-occurrence word is necessarily of the form $\beta = x, \gamma, x$, where γ is a planar double-occurrence word.*

In other words, every indecomposable rooted planar chord diagram has a unique outermost chord. This simple property is readily apparent from inspection of the binding diagrams of indecomposable planar terms (see Figures 2–5), and it enables us to transform any indecomposable rooted planar chord diagram with $n + 1$ (full) chords into an n -node binary tree conveying equivalent information.

Geometrically, the construction could be visualized like so:



Going from left to right, we begin by removing the outermost chord and applying the canonical isomorphism between rooted planar chord diagrams and rooted planar trees (which interprets chords as nodes, and the covering relation between chords as the parent-child relation between nodes). Then we apply the (non-canonical) “left-child, right-sibling” correspondence between rooted planar trees and binary trees, and flip the resulting binary tree vertically to get our preferred orientation.

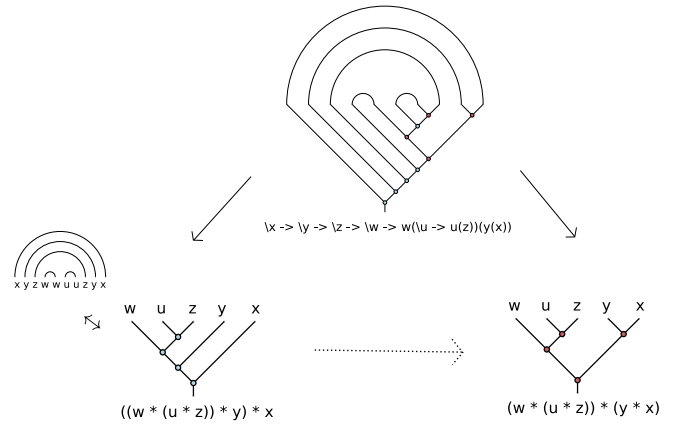


Fig. 3. Extracting a Tamari interval from an indecomposable planar term

The miracle is that whenever we apply this process to the binding diagram $\gamma[M]$ of a closed indecomposable planar term M , we get a tree which is below the application tree $\alpha[M]$ in the Tamari order! Figure 3 shows a demonstration of this on a term of size four, and the reader is invited to test it out for herself on any of the diagrams in Figures 2, 4 or 5.

So how do we prove that this always works? It turns out that the sequent calculus of Section II is especially well-suited (indeed, this was its original *raison d’être*). If one tries to give a formal inductive definition of the trees constructed above from the binding diagrams of closed indecomposable planar terms, this quickly leads to the consideration of *binding forests* for open terms – these correspond to sequent contexts, but they also make sense geometrically, since the binding diagram of an open planar term contains unattached arcs which partition the diagram into disjoint subdiagrams.

Definition IV.10. Let M be any indecomposable planar term with k free variables, for $k \geq 1$. The **binding forest** of M is a list of k (leaf-labelled) binary trees $\gamma^F[M]$, defined by induction as follows:

$$\begin{aligned} \gamma^F[x] &= x \\ \gamma^F[M(N)] &= \Gamma, \Delta \quad \text{where } \gamma^F[M] = \Gamma \text{ and } \gamma^F[N] = \Delta \\ \gamma^F[\lambda x.M] &= A * B, \Gamma \quad \text{where } \gamma^F[M] = A, B, \Gamma \\ &\quad \text{assuming } x \text{ at rightmost end of } \text{fr}(A) \end{aligned}$$

By extension, the **binding tree** of a closed indecomposable planar term (cf. Prop. IV.1) is defined by $\gamma^T[\lambda x.M] = \gamma^F[M]$.

Note the side-condition on the variable x in the definition of $\gamma^F[\lambda x.M]$, which enforces planarity.

Proposition IV.11. *If M is an indecomposable planar term with $k \geq 1$ free variables, then $\gamma^F[M]$ is well-defined, and the leftmost free variable of M occurs at the rightmost end of the frontier of the first formula in $\gamma^F[M]$.*

With this definition of binding forests as contexts, verifying the “miracle” becomes a routine exercise in sequent calculus.

Proposition IV.12. $\gamma^F[M] \longrightarrow \alpha[M]$ for all indecomposable planar terms M .

Proof. Immediate by induction on M , interpreting abstractions by the $*L$ rule, applications by the $*R$ rule, and variables by the id rule (in its atomic instance id^{atm}). \square

Corollary IV.13. For all closed indecomposable planar terms M , the binding tree $\gamma^T[M]$ is below the application tree $\alpha[M]$ in the Tamari order.

Finally, why does this give a bijection from closed β -normal planar indecomposable terms to Tamari intervals? Well, by examination of the proof of Prop. IV.12, it is clear that we can also go in the opposite direction and turn any sequent calculus derivation involving the rules $*L$, $*R$, and id^{atm} back into an indecomposable planar term: just interpret $*L$ as an abstraction (binding the atom at the rightmost end of the frontier of the first formula of the context in the premise), $*R$ as an application, and id^{atm} as a variable occurrence. (The frontier invariance property can be used to rename atoms as necessary, if we want to obtain a term satisfying the Barendregt convention.) Since such derivations include all the focused ones, this establishes that the mapping from closed indecomposable planar terms to Tamari intervals is surjective, by the focusing completeness theorem. Moreover, within this fragment of the sequent calculus, the only possible source of a focusing violation is if we have a derivation whose root is $*L$ as the left premise of a derivation whose root is $*R$ – but such violations can be resolved by permuting $*L$ under $*R$:

$$\frac{\frac{A, B, \Gamma_1 \longrightarrow C}{A * B, \Gamma_1 \longrightarrow C} L \quad \Gamma_2 \longrightarrow D}{A * B, \Gamma_1, \Gamma_2 \longrightarrow C * D} R \quad \longrightarrow \quad \frac{\frac{A, B, \Gamma_1 \longrightarrow C \quad \Gamma_2 \longrightarrow D}{A, B, \Gamma_1, \Gamma_2 \longrightarrow C * D} R}{A * B, \Gamma_1, \Gamma_2 \longrightarrow C * D} L$$

This transformation on derivations corresponds precisely to performing a ρ -reduction on the corresponding lambda terms. Therefore focused derivations correspond precisely to ρ -normal indecomposable planar terms, but the latter also happen to be β -normal, and by the coherence theorem we have:

Theorem IV.14. Closed indecomposable β -normal planar terms with n applications are in bijection with intervals in Y_n . There are exactly $\frac{2(4n+1)!}{(n+1)!(3n+2)!}$ of them.

ACKNOWLEDGMENT

I thank Tarmo Uustalu for telling me about skew monoidal categories. The diagrams in this paper were created using the help of Brent Yorgey’s Diagrams library for Haskell.

REFERENCES

- [1] J.-M. Andreoli, “Logic programming with focusing proofs in linear logic,” *J. Logic and Computation*, vol. 2, no. 3, pp. 297–347, 1992.
- [2] H. P. Barendregt, *The Lambda Calculus: Its Syntax and Semantics*, ser. Studies in Logic and the Found. of Math. Elsevier, 1984, vol. 103.
- [3] O. Bernardi and N. Bonichon, “Intervals in Catalan lattices and realizers of triangulations,” *J. Combin. Theory Ser. A*, vol. 116, no. 1, pp. 55–75, 2009.
- [4] S. Burrill, J. Courtiel, E. Fusy, S. Melczer, , and M. Mishna, “Tableau sequences, open diagrams, and Baxter families,” *European J. Combin.*, vol. 58, pp. 144–165, 2016.
- [5] F. Chapoton, “Sur le nombre d’intervalles dans les treillis de Tamari,” *Sém. Lothar. Combin.*, no. B55f, 2006, 18 pp. (electronic).

- [6] R. Cori, “Indecomposable permutations, hypermaps and labeled Dyck paths,” *J. Combin. Theory Ser. A*, vol. 116, no. 8, pp. 1326–1343, 2009.
- [7] R. Cori and G. Schaeffer, “Description trees and Tutte formulas,” *Theoretical Computer Science*, vol. 292, no. 1, pp. 165–183, 2003, selected papers in honor of Jean Berstel.
- [8] W. Fang, “Planar triangulations, bridgeless planar maps and Tamari intervals,” arXiv:1611.07922.
- [9] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*. Camb. Univ. Press, 2009.
- [10] H. Friedman and D. Tamari, “Problèmes d’associativité: une structure de treillis finis induite par une loi demi-associative,” *J. Combinatorial Theory*, vol. 2, pp. 215–242, 1967.
- [11] G. Gentzen, “Investigations into logical deductions,” in *The Collected Papers of Gerhard Gentzen*, M. E. Szabo, Ed. Amsterdam: North-Holland, 1969, pp. 68–131.
- [12] S. Huang and D. Tamari, “Problems of associativity: A simple proof for the lattice property of systems ordered by a semi-associative law,” *J. Combin. Theory Ser. A*, vol. 13, no. 1, pp. 7–13, 1972.
- [13] D. E. Knuth, *The Art of Computer Programming, Volume 4A: Combinatorial Algorithms (Part 1)*. Addison-Wesley, 2006.
- [14] J. Kock, “Polynomial functors and trees,” *Internat. Math. Res. Notices*, pp. 609–673, 2011.
- [15] S. Lack and R. Street, “Triangulations, orientals, and skew monoidal categories,” *Advances in Mathematics*, vol. 258, pp. 351–396, 2014.
- [16] J. Lambek, “The mathematics of sentence structure,” *The American Mathematical Monthly*, vol. 65, no. 3, pp. 154–170, 1958.
- [17] —, “On the calculus of syntactic types,” in *Structure of Language and its Mathematical Aspects*, ser. Proceedings of Symposia in Applied Mathematics, XII, R. Jakobson, Ed. AMS, 1961, pp. 166–178.
- [18] S. K. Lando and A. K. Zvonkin, *Graphs on Surfaces and Their Applications*, ser. Encyclopaedia of Mathematical Sciences. Springer-Verlag, 2004, no. 141.
- [19] S. M. Lane, “Natural associativity and commutativity,” *Rice University Studies*, vol. 49, no. 4, pp. 28–46, 1963.
- [20] —, *Categories for the Working Mathematician*. Springer, 1998.
- [21] M. L. Laplaza, “Coherence for associativity not an isomorphism,” *J. Pure and Appl. Alg.*, vol. 2, no. 2, pp. 107–120, 1972.
- [22] P.-A. Melliès, “Axiomatic rewriting theory VI: Residual theory revisited,” in *Rewriting Techniques and Applications*, ser. Lecture Notes in Computer Science, S. Tison, Ed. Springer, 2002, vol. 2378, pp. 24–50.
- [23] F. Müller-Hoissen and H.-O. Walther, Eds., *Associahedra, Tamari Lattices and Related Structures: Tamari Memorial Festschrift*, ser. Progress in Mathematics, vol. 299. Birkhauser, 2012.
- [24] P. Ossona de Mendez and P. Rosenstiehl, “Encoding pointed maps by double occurrence words,” in *Volume of essays in honour of Prof. A. C. Panayotopoulos*, U. of Piraeus, Ed. Eptalofos, 2006, pp. 701–712.
- [25] J. Reed and F. Pfenning, “Focus-preserving embeddings of substructural logics in intuitionistic logic,” unpublished manuscript, January 2010.
- [26] D. D. Sleator, R. E. Tarjan, and W. P. Thurston, “Rotation distance, triangulations, and hyperbolic geometry,” *J. Amer. Math. Soc.*, vol. 1, no. 3, pp. 647–681, 1988.
- [27] N. J. A. Sloane, “The On-Line Encyclopedia of Integer Sequences,” published electronically at <https://oeis.org>.
- [28] J. D. Stasheff, “Homotopy associativity of H-spaces, I,” *Trans. Amer. Math. Soc.*, vol. 108, pp. 293–312, 1963.
- [29] D. Tamari, “Monoïdes préordonnés et chaînes de Malcev,” Thèse, Université de Paris, 1951.
- [30] —, “Sur quelques problèmes d’associativité,” *Ann. sci. de Univ. de Clermont-Ferrand 2, Sér. Math.*, vol. 24, pp. 91–107, 1964.
- [31] W. T. Tutte, “A census of planar triangulations,” *Canad. J. Math.*, vol. 14, pp. 21–38, 1962.
- [32] T. R. S. Walsh and A. B. Lehman, “Counting rooted maps by genus III: Nonseparable maps,” *J. Combin. Theory Ser. B*, vol. 18, pp. 222–259, 1975.
- [33] N. Zeilberger, “On the unity of duality,” *Ann. Pure and Applied Logic*, vol. 153, no. 1–3, pp. 66–96, 2008.
- [34] —, “Linear lambda terms as invariants of rooted trivalent maps,” *J. Functional Programming*, vol. 26, no. e21, 2016.
- [35] N. Zeilberger and A. Giorgetti, “A correspondence between rooted planar maps and normal planar lambda terms,” *Logical Methods in Computer Science*, vol. 11, no. 3:22, 2015.

APPENDIX

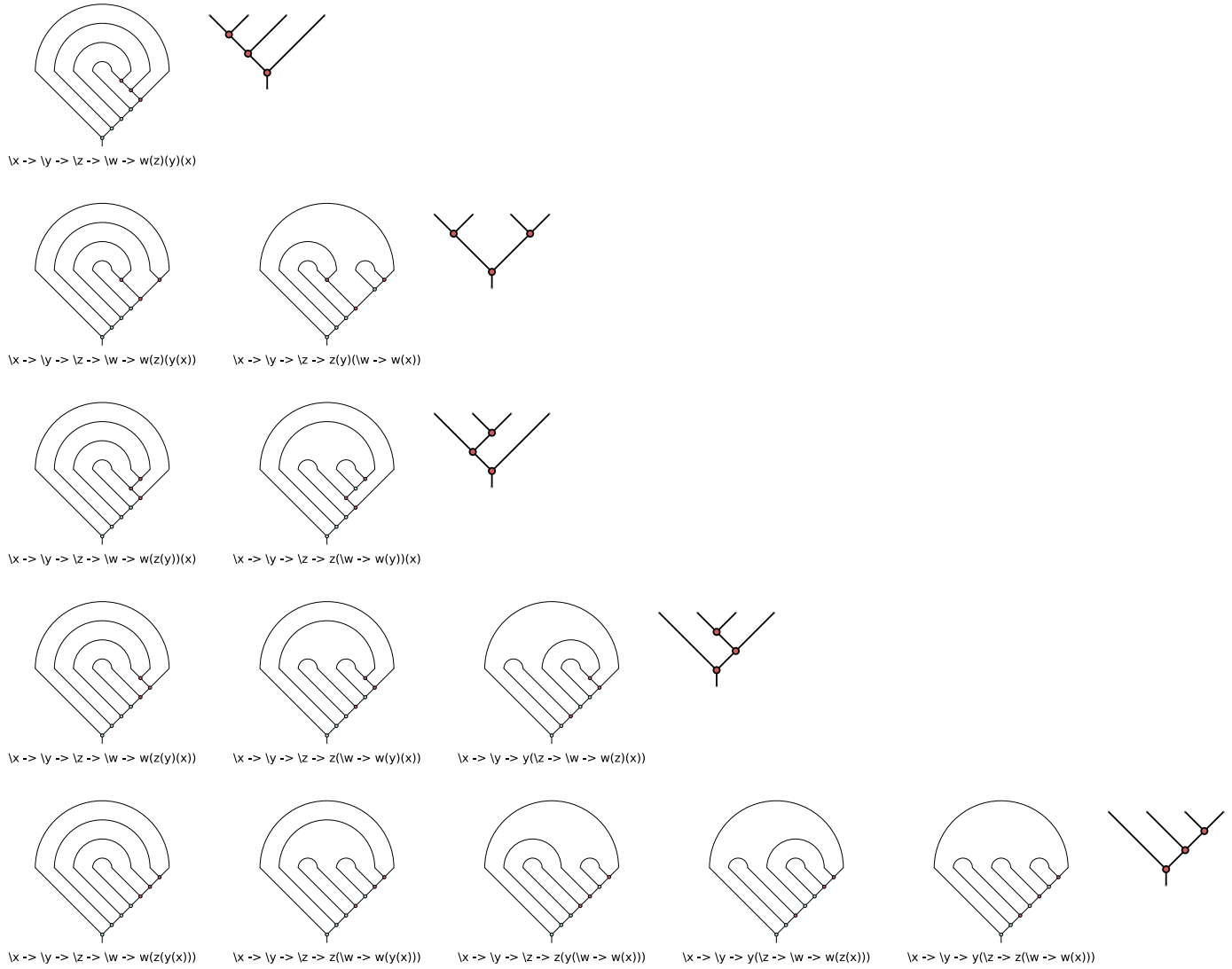


Fig. 4. All 13 indecomposable normal planar terms of size three, organized into rows according to their underlying application tree.

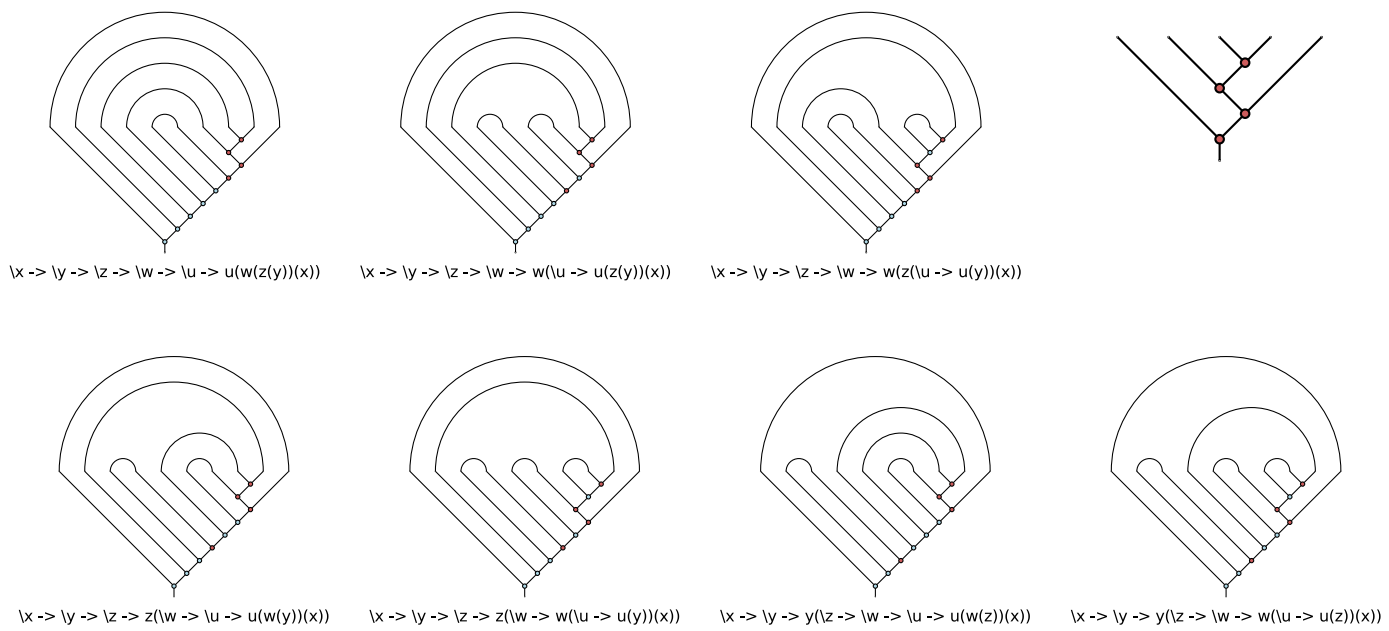


Fig. 5. A list of indecomposable normal planar terms of size four, all sharing the same application tree.