A Sequent Calculus for a Semi-Associative Law

Noam Zeilberger
University of Birmingham, Birmingham, UK
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 $Y_n$. Elsewhere, we have also used the sequent calculus and the coherence theorem to build a surprising bijection between intervals of the Tamari order and a natural fragment of lambda calculus, consisting of the $\beta$-normal planar lambda terms with no closed proper subterms.

1998 ACM Subject Classification F.4.1 Mathematical Logic, G.2.1 Combinatorics

Keywords and phrases proof theory, combinatorics, substructural logic, coherence theorem

Digital Object Identifier 10.4230/LIPIcs.FSCD.2017.33

1 Introduction

1.1 The Tamari order, Tamari lattices and associahedra

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 [25, 26, 6], the Tamari order is the partial ordering on words induced by asking that multiplication obeys a semi-associative law\footnote{Clearly, 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.}

\[(A * B) * C \leq A * (B * C)\]

and is monotonic in each argument:

\[
\begin{align*}
A \leq A' & \quad \frac{A * B \leq A' * B}{A * B \leq A' * B} \\
B \leq B' & \quad \frac{A * B \leq A * B'}{A * B \leq A * B'}
\end{align*}
\]
For example, the word \((p \ast (q \ast r)) \ast s\) is below the word \(p \ast (q \ast (r \ast s))\) in the Tamari order:

\[
\begin{array}{ccc}
p & q & r & s \\
\downarrow & \downarrow & \downarrow & \downarrow \\
p & q & r & s \\
\end{array}
\rightarrow
\begin{array}{ccc}
p & q & r & s \\
\downarrow & \downarrow & \downarrow & \downarrow \\
p & q & r & s \\
\end{array}
\rightarrow
\begin{array}{ccc}
p & q & r & s \\
\downarrow & \downarrow & \downarrow & \downarrow \\
p & q & r & s \\
\end{array}
\]

The variables \(p, q, \ldots\) 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 unlabelled binary trees. Since such trees are enumerated by the ubiquitous Catalan numbers (there are \(C_n = \binom{2n}{n}/(n+1)\) distinct binary 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 [8], triangulations of a polygon [20], or Dyck paths [2].

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”.

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” [22, 17].

### 1.2 A Lambekian analysis of the Tamari order

In this paper we will introduce and study a surprisingly elementary presentation of the Tamari order as a sequent calculus in the spirit of Lambek [10, 11]. This calculus consists of just one left rule and one right rule:

- **Left rule:**
  \[
  \frac{A, B, \Delta \rightarrow C}{A \ast B, \Delta \rightarrow C} \ast L
  \]

- **Right rule:**
  \[
  \frac{\Gamma \rightarrow A, \Delta \rightarrow B}{\Gamma, \Delta \rightarrow A \ast B} \ast R
  \]

together with two structural rules:

- **Identity rule:**
  \[
  \frac{\Theta \rightarrow A}{A \rightarrow A} \text{ id}
  \]

- **Cut rule:**
  \[
  \frac{\Gamma, \Theta, \Delta \rightarrow B}{\Gamma, \Theta, \Delta \rightarrow B} \text{ cut}
  \]
Here, $\Gamma$, $\Delta$, and $\Theta$ 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 [10], except for the $*$L rule which is a restriction of his left rule for products. Lambek’s original rule looked like this:

\[
\begin{align*}
\Gamma, A, B, \Delta &\rightarrow C \\
\Gamma, A \ast B, \Delta &\rightarrow C \ast L_{\text{amb}}.
\end{align*}
\]

That is, Lambek’s left rule allowed the formula $A \ast B$ to appear anywhere in the context, whereas our more restrictive rule $\ast L$ 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 $(p \ast (q \ast r)) \ast s \leq p \ast (q \ast (r \ast s))$ (we write L and R as short for $\ast L$ and $\ast R$, and don’t bother labelling instances of id):

\[
\begin{align*}
q &\rightarrow q \\
r &\rightarrow r \\
s &\rightarrow r \ast s \\
q, r, s &\rightarrow q \ast (r \ast s) \\
p \rightarrow p \\
p \ast q \ast r, s &\rightarrow p \ast (q \ast (r \ast s)) \ast L \\
p \ast (q \ast r), s &\rightarrow p \ast (q \ast (r \ast s)) \ast R \\
(p \ast (q \ast r)) \ast s &\rightarrow p \ast (q \ast (r \ast s)) \ast L.
\end{align*}
\]

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

\[
\begin{align*}
p \rightarrow p \\
q &\rightarrow q \\
r &\rightarrow r \\
s &\rightarrow s \\
q, r, s &\rightarrow q \ast (r \ast s) \\
p \rightarrow p \\
p, q, r, s &\rightarrow p \ast (q \ast (r \ast s)) \ast L_{\text{amb}} \\
p, q \ast r \ast s &\rightarrow (p \ast (q \ast (r \ast s))) \ast s \ast L_{\text{amb}} \\
p \ast (q \ast (r \ast s)) &\rightarrow (p \ast (q \ast (r \ast s))) \ast s \ast L_{\text{amb}} \\
(p \ast (q \ast r)) \ast s &\rightarrow (p \ast (q \ast (r \ast s))) \ast s \ast L_{\text{amb}}.
\end{align*}
\]

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

**Claim 1.** $A \rightarrow B$ is derivable using the rules $\ast L$, $\ast R$, 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 [7]. We will prove the following somewhat stronger form of cut-elimination:

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

\[
\begin{align*}
\Gamma^{\text{init}} &\rightarrow A \\
A &\rightarrow B \\
\Delta &\rightarrow A \ast B \\
\ast R^{\text{loc}} &\rightarrow p \rightarrow p \\
\text{id}_{\text{init}} &\rightarrow
\end{align*}
\]

where $\Gamma^{\text{init}}$ ranges over contexts that don’t have a product $C \ast D$ 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 \( *L, *R^{boc} \), 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 careful analysis of these three rules confirms that any sequent \( \Gamma \rightarrow A \) has at most one focused derivation.

Combining this property with Claims 1 and 2, we therefore have

**Claim 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.

### 1.3 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 [3] proved the following surprising formula for the number of *intervals* in the Tamari lattice \( Y_n \):

\[
\frac{2(4n+1)!}{(n+1)!(3n+2)!}
\]

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) [21]. 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 [27] 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 [28].

Sparked by Chapoton’s observation, Bernardi and Bonichon [2] found an explicit bijection between intervals of the Tamari order and 3-connected rooted planar triangulations, and quite recently, Fang [5] 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).

Meanwhile, in [32], Alain Giorgetti and I gave a bijection between rooted planar maps (possibly containing bridges) and a simple fragment of linear lambda calculus consisting of the \( \beta \)-normal planar terms. As with Chapoton’s result, this connection between maps and lambda calculus was found using hints from the OEIS, since the sequence enumerating rooted

---

2 A *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. [12].)
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 \((\frac{2(2n)!}{n!(n+2)!})\). Now, let us say that a term is indecomposable if it has no closed proper subterms. Although this property may be unfamiliar to some readers, indecomposability turns out to be very natural to consider in a linear context – for example, in [30] it was used to give a lambda calculus reformulation of the Four Color Theorem, based on a characterization of bridgeless rooted trivalent maps as indecomposable linear lambda terms. In any case, it is not difficult to check that the bijection described in [32] restricts to a bijection between bridgeless rooted planar maps and indecomposable \( \beta \)-normal planar terms, and therefore that formula (1) also enumerates indecomposable \( \beta \)-normal planar terms by size. It is then a natural question whether there exists a direct bijection between such terms and intervals of the Tamari order.

An explicit bijection between indecomposable \( \beta \)-normal planar terms and Tamari intervals was given in an earlier, longer version of this paper [31], and the proof of its correctness relies in an essential way on the sequent calculus and coherence theorem. That bijection is omitted here in part due to space constraints, and in part because I would like to give a fuller and more conceptual account of the combinatorics, connecting it to the duality between skew-monoidal categories [24] and skew-closed categories [23]. However, even without this original application, I believe that the sequent calculus considered here is of intrinsic mathematical interest: Tamari lattices and related associahedra have been studied for over sixty years, so the fact that such an elementary and natural proof-theoretic characterization of semi-associativity has been seemingly overlooked is surprising. Moreover, this characterization is clearly productive, since it leads to a much more structured proof of Chapoton’s seminal result – an unexpected link between important ideas in proof theory (such as cut-elimination and focusing) and an active research topic in combinatorics.

## 2 A sequent calculus for the Tamari order

### 2.1 Definitions and terminology

For reference, we recall here the definition of the sequent calculus introduced in 1.2, and clarify some notational conventions. The four rules of the sequent calculus are:

\[
\begin{align*}
A, B, \Delta & \to C \\
A \ast B, \Delta & \to C & *L \\
\Gamma, \Delta & \to A \ast B & +R \\
\Gamma, \Delta & \to A & id \\
\Gamma, \Theta, \Delta & \to B & cut
\end{align*}
\]

Uppercase Latin letters \((A, B, \ldots)\) range over formulas, which can be either compound \((A \ast B)\) or atomic (ranged over by lowercase Latin letters \(p, q, \ldots\)). Uppercase Greek letters \(\Gamma, \Delta, \ldots\) 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 [10] but in contrast to Gentzen’s original sequent calculus [7], 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 \(\Gamma\) and a formula \(A\). Abstractly, a derivation is a tree (technically, a rooted planar tree with boundary) 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 1.2, when constructing derivations we sometimes write \( L \) and \( R \) as shorthand for \( \ast L \) and \( \ast R \), and usually don’t bother labelling the instances of \( id \) and \( cut \) since they are clear from context.

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

**Proposition 4.** Suppose that \( \Gamma \rightarrow A \). Then
1. (Refinement) \( fr(\Gamma) = fr(A) \).
2. (Relabelling) \( \sigma \Gamma \rightarrow \sigma A \), where \( \sigma \) is any function sending atoms to atoms.

### 2.2 Completeness

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

**Theorem 5** (Completeness). If \( A \leq B \) then \( A \rightarrow B \).

**Proof.** We must show that the relation \( A \rightarrow B \) 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:

\[
\frac{A \rightarrow A \quad B \rightarrow B \quad C \rightarrow C}{A \ast B \rightarrow (B \ast C)} \quad R
\]

\[
\frac{A, B, C \rightarrow A \ast (B \ast C)}{A \ast B, C \rightarrow A \ast (B \ast C)} \quad R
\]

\[
\frac{(A \ast B) \ast C \rightarrow A \ast (B \ast C)}{(A \ast B) \ast C \rightarrow A \ast (B \ast C)} \quad L
\]

4. Monotonicity:

\[
\frac{A \rightarrow A' \quad B \rightarrow B'}{A \ast B \rightarrow A' \ast B'} \quad L
\]

\[
\frac{A \rightarrow A \quad B \rightarrow B'}{A, B \rightarrow A \ast B'} \quad R
\]

\[
\frac{A \ast B \rightarrow A \ast B'}{A \ast B \rightarrow A \ast B'} \quad L
\]

**2.3 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

\[
A_1, A_2, \ldots, A_n \rightarrow B
\]
(where \( n \geq 1 \)) is interpreted as an entailment of the form
\[
(\cdots (A_1 \ast A_2) \ast \cdots) \ast A_n \leq B
\]
in the Tamari order. Visualizing everything in terms of binary trees, the sequent can be interpreted like so:

\[
\begin{array}{c}
A_1 \\
\downarrow \\
\vdots \\
A_n \\
\rightarrow \\
B
\end{array}
\]

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 \( \Gamma \) to a formula \( \phi[\Gamma] \) by the above interpretation. The operation is defined by the following equations:

\[
\phi[A] = A \quad \phi[\Gamma, A] = \phi[\Gamma] \ast A
\]

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

\[\textbf{Proposition 6.} \quad \phi[\Gamma, \Delta] \leq \phi[\Gamma] \ast \phi[\Delta] \text{ for all non-empty contexts } \Gamma \text{ and } \Delta.\]

\[\textbf{Proof.} \quad \text{By induction on } \Delta. \text{ The case of a singleton context } \Delta = A \text{ is immediate. Otherwise, if } \Delta = (\Delta', A), \text{ we have}\]

\[
\phi[\Gamma, \Delta', A] = \phi[\Gamma, \Delta'] \ast A \leq (\phi[\Gamma] \ast \phi[\Delta']) \ast A \leq \phi[\Gamma] \ast (\phi[\Delta'] \ast A) = \phi[\Gamma] \ast \phi[\Delta', A]
\]

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 \( \Delta \ast A \) of an arbitrary context on a formula, where this action is defined by the following equations:

\[
\begin{align*}
A \ast - &= A \\
A \ast (\Delta, B) &= (A \ast \Delta) \ast B
\end{align*}
\]

We will make use of a few simple properties of \( - \ast \Delta \):

\[\textbf{Proposition 7.} \quad \phi[\Gamma, \Delta] = \phi[\Gamma] \ast \Delta \text{ for all non-empty contexts } \Gamma \text{ and arbitrary contexts } \Delta.\]

\[\textbf{Proposition 8 (Monotonicity).} \quad \text{If } A \leq A' \text{ then } A \ast \Delta \leq A' \ast \Delta.\]

\[\textbf{Proof.} \quad \text{Both properties are immediate by induction on } \Delta, \text{ where in the case of Prop. 8 we apply monotonicity of the operations } - \ast B. \]

We are now ready to prove soundness.

\[\textbf{Theorem 9 (Soundness).} \quad \text{If } \Gamma \rightarrow A \text{ then } \phi[\Gamma] \leq A.\]

\[\textbf{Proof.} \quad \text{By induction on the (closed) derivation of } \Gamma \rightarrow A. \text{ There are four cases, corresponding to the four rules of the sequent calculus:}\]
(Case \(*L\)): The derivation ends in
\[
\begin{align*}
    A, B, \Delta & \rightarrow C \\
    A \ast B, \Delta & \rightarrow C
\end{align*}
\]
By induction we have \(\phi[A, B, \Delta] \leq C\), but by Prop. 7 we have
\(\phi[A \ast B, \Delta] = \phi[A, B] \ast \Delta = (A \ast B) \ast \Delta = \phi[A, B, \Delta]\).

(Case \(*R\)): The derivation ends in
\[
\begin{align*}
    \Gamma & \rightarrow A \\
    \Gamma, \Delta & \rightarrow A \ast B
\end{align*}
\]
By induction we have \(\phi[\Gamma] \leq A\) and \(\phi[\Delta] \leq B\), hence
\(\phi[\Gamma, \Delta] \leq (\phi[\Gamma] \ast \phi[\Delta]) \ast \Delta = (\phi[\Gamma] \ast A) \ast \Delta = \phi[\Gamma, A, \Delta] \leq B\).

2.4 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 10. A context \(\Gamma\) is said to be reducible if its leftmost formula is compound, and irreducible otherwise. A sequent \(\Gamma \rightarrow A\) is said to be:
- left-inverting if \(\Gamma\) is reducible;
- right-focusing if \(\Gamma\) is irreducible and \(A\) is compound;
- atomic if \(\Gamma\) is irreducible and \(A\) is atomic.

Proposition 11. Any sequent is either left-inverting, right-focusing, or atomic.

Definition 12. A closed derivation \(D\) 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 \(\Gamma^{irr}\) to indicate that a context \(\Gamma\) is irreducible.

Proposition 13. 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):
\[
\begin{align*}
    \Gamma^{irr} & \rightarrow A \\
    \Gamma^{irr}, \Delta & \rightarrow A \ast B
\end{align*}
\]

Example 14. One way to derive \(((p \ast q) \ast r) \ast s \rightarrow p \ast ((q \ast r) \ast s)\) is by cutting together
the two derivations
\[
\begin{align*}
    SA_{p, q, r} \\
    (p \ast q) \ast r & \rightarrow p \ast (q \ast r) \\
    (p \ast q) \ast r, s & \rightarrow (p \ast (q \ast r)) \ast s
\end{align*}
\]
and
\[
\begin{align*}
    SA_{p, q, r, s} \\
    (p \ast (q \ast r)) \ast s & \rightarrow p \ast ((q \ast r) \ast s)
\end{align*}
\]
where $\mathcal{S}_A, A, B, C$ is the derivation of the semi-associative law $(A * B) * C \rightarrow A * (B * C)$ from the proof of Theorem 5. 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:

\[
\frac{q \rightarrow q}{q, r \rightarrow q * r} \quad \frac{q * r \rightarrow \overline{r}}{R} \quad \frac{s \rightarrow s}{R}
\]

\[
\frac{p \rightarrow p}{p, q, r, s \rightarrow p * ((q * r) * s)} \quad \frac{p * q \rightarrow R}{L} \quad \frac{p * q \rightarrow R}{L}
\]

\[
\frac{(p * q) \rightarrow R}{L} \quad \frac{(p * q) \rightarrow R}{L}
\]

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

**Theorem 15 (Focusing completeness).** If $\Gamma \rightarrow A$ then $\Gamma \rightarrow A$.

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. [19]), which entails the admissibility of id, then show cut and $*$R in turn.

**Lemma 16 (Deduction).** If $\Gamma^{\text{itr}} \rightarrow A$ implies $\Gamma^{\text{itr}} \rightarrow B$ for all $\Gamma^{\text{itr}}$, then $A, \Delta \rightarrow B$. In particular, $A \rightarrow A$.

**Proof.** By induction on the formula $A$:

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

  \[
  \frac{A_1, A_1, \Delta \rightarrow B}{A_1 * A_2, \Delta \rightarrow B} *L
  \]

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

  \[
  \mathcal{D}_1 \quad \frac{A_1 \rightarrow A_2}{D_2} \quad \frac{\mathcal{D}_1 \rightarrow A_1 \rightarrow A_2}{D} *R
  \]

  where we apply the i.h. on $A_2$ to obtain $D_2$. Finally, applying the assumption to $\mathcal{D}$ (with $\Gamma^{\text{itr}} = \Gamma^{\text{itr}}_1, A_2$) we obtain the desired derivation of $\Gamma^{\text{itr}}_1, A_2, \Delta \rightarrow B$.

**Lemma 17 (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. 4) 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 \( id_{\text{atm}} \)): 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

\[
\mathcal{E} \Rightarrow \begin{array}{c}
\Gamma_{\text{irr}}, A, \Delta_1 \Rightarrow B_1 \\
\Gamma_{\text{irr}}, A, \Delta_2 \Rightarrow B_2 \\
\end{array} \quad *R
\]

2. \( \exists \Gamma_{\text{irr}} \) such that \( \Gamma_{\text{irr}} = \Gamma \) and

\[
\mathcal{E} \Rightarrow \begin{array}{c}
\Gamma_{\text{irr}}, \Gamma_{\text{irr}}, A, \Delta \Rightarrow B_1 \\
\end{array} \quad *R
\]

In the first case, we cut \( \mathcal{D} \) with \( \mathcal{E}_1 \) to obtain \( \Gamma_{\text{irr}}, \Theta, \Delta \Rightarrow B_1 \), 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

\[
\mathcal{E} \Rightarrow \begin{array}{c}
C_1, C_2, \Gamma', A, \Delta \Rightarrow B \\
\end{array} \quad *L
\]

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

2. \( \Gamma = \cdot \) and

\[
\mathcal{E} \Rightarrow \begin{array}{c}
A_1, A_2, \Delta \Rightarrow B \\
\end{array} \quad *L
\]

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

\[
\mathcal{D} \Rightarrow \begin{array}{c}
C_1, C_2, \Theta' \Rightarrow A_1 * A_2 \\
\end{array} \quad *L
\]

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

- (Case \( id_{\text{atm}} \)): Impossible.
- (Case \( *R_{\text{foc}} \)): \( \exists \Theta_{\text{irr}} \) s.t. \( \Theta_{\text{irr}} = \Theta_{\text{irr}} \) and

\[
\mathcal{D} \Rightarrow \begin{array}{c}
\Theta_{\text{irr}} \Rightarrow A_1 \\
\Theta_{\text{irr}} \Rightarrow A_2 \\
\end{array} \quad *R
\]

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).

\[\blacktriangleright\]

\textbf{Lemma 18} \( (*R \ \text{admiss.}) \). \textit{If} \( \Gamma \Rightarrow A \) \textit{and} \( \Delta \Rightarrow B \) \textit{then} \( \Gamma, \Delta \Rightarrow A * B \).
Proof. Let $D : \Gamma \Rightarrow \rightarrow A$ and $E : \Delta \Rightarrow \rightarrow B$. We proceed by induction on $D$. If $\Gamma$ is irreducible then we directly apply $*R^{\text{foc}}$. Otherwise, $D$ must be of the form

$$D = \frac{D', C_1, C_2, \Gamma' \rightarrow A}{C_1 \ast C_2, \Gamma' \rightarrow A} *L$$

so we apply the i.h. to $D'$ with $E$, and reapply the $*L$ rule.

Proof of Theorem 15. 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 $*R$.

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

Proposition 19 (Frontier invariance). Let $\sigma$ be any function sending atoms to atoms. Then $\Gamma \Rightarrow \rightarrow A$ if and only if $\text{fr}(\Gamma) = \text{fr}(A)$ and $\sigma \Gamma \Rightarrow \rightarrow \sigma A$.

Proof. The forward direction is Prop. 4. For the backward direction we use induction on focused derivations, which is justified by Theorem 15. The only interesting case is $*R^{\text{foc}}$, where we can assume $\text{fr}(\Gamma, \Delta) = \text{fr}(A \ast B)$ and $\sigma \Gamma \Rightarrow \rightarrow \sigma A$ ($\sigma \Gamma$ irreducible) and $\sigma \Delta \Rightarrow \rightarrow \sigma B$.

By Prop. 4 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 $\text{fr}(\Delta) = \text{fr}(A)$, from which $\Gamma \Rightarrow \rightarrow A$ ($\Gamma$ irreducible) and $\Delta \Rightarrow \rightarrow B$ follow by the induction hypothesis, hence $\Gamma, \Delta \Rightarrow \rightarrow A \ast B$.

If we let $\sigma = _\rightarrow \mapsto p$ be any constant relabelling function, then speaking in terms of the Tamari order, Proposition 19 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 20. We say that an irreducible context $\Gamma^{\text{irr}}$ is a maximal decomposition of $A$ if $\Gamma^{\text{irr}} \Rightarrow \rightarrow A$, and for any other $\Theta^{\text{irr}}, \Theta^{\text{irr}} \Rightarrow \rightarrow A$ implies $\Theta^{\text{irr}} \Rightarrow \rightarrow \phi[\Gamma^{\text{irr}}]$.

Proposition 21. If $\Gamma^{\text{irr}}$ is a maximal decomposition of $A$, then $A, \Delta \Rightarrow \rightarrow B$ if and only if $\Gamma^{\text{irr}}, \Delta \Rightarrow \rightarrow B$.

Proof. The forward direction is by cutting with $\Gamma^{\text{irr}} \Rightarrow \rightarrow A$, the backwards direction is by the deduction lemma (16) and the universal property of $\Gamma^{\text{irr}}$.

Proposition 22. Let $\psi[A]$ be the irreducible context defined inductively by:

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

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

Proof. We construct $\psi[A] \Rightarrow \rightarrow A$ by induction on $A$, and prove the universal property of $\psi[A]$ by induction on focused derivations of $\Theta^{\text{irr}} \Rightarrow \rightarrow A$.

Proposition 23. $\phi[\psi[A]] = A$ and $\psi[\phi[\Theta^{\text{irr}}]] = \Theta^{\text{irr}}$.

The maximal decomposition $\psi[A]$ of $A$ is essentially the same thing as what Chapoton [3] calls a “décomposition maximale” of a binary tree. The logical characterization expressed in Defn. 20 is quite general, though, and is familiar from studies of focusing in other settings (cf. [29]).
2.5 The coherence theorem

We now come to our main result:

**Theorem 24 (Coherence).** Every derivable sequent has exactly one focused derivation.

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

**Lemma 25.** For any context $\Gamma$ and formula $A$, there is at most one focused derivation of $\Gamma \rightarrow A$.

**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 $|A * B| = 1 + |A| + |B|$ and $|p| = 0$. (That is, $|A|$ counts the number of multiplication operations occurring in $A$.) Then any sequent $A_1, \ldots, A_n \rightarrow B$ induces a multiset of sizes $(\biguplus_{i=1}^n |A_i|) \uplus |B|$, and at each step of our induction this multiset will decrease in the multiset ordering.

There are three cases:

- (A left-inverting sequent $A * B, \Delta \rightarrow C$): Any focused derivation must end in $*L$, so we apply the i.h. to $A, B, \Delta \rightarrow C$.

- (A right-focusing sequent $\Gamma_{irr} \rightarrow A * B$): Any focused derivation must end in $*R$, and decide some splitting of the context into contiguous pieces $\Gamma_{irr}^1$ and $\Delta_2$. However, $\Gamma_{irr}^1$ and $\Delta_2$ are uniquely determined by frontier refinement ($fr(\Gamma_{irr}^1) = fr(A)$ and $fr(\Delta_2) = fr(B)$) and the equation $\Gamma_{irr} = \Gamma_{irr}^1, \Delta_2$. So, we apply the i.h. to $\Gamma_{irr}^1 \rightarrow A$ and $\Delta_2 \rightarrow B$.

- (An atomic sequent $\Gamma_{irr} \rightarrow p$): The sequent has exactly one focused derivation if and only if $\Gamma_{irr} = p$.

**Proof of Theorem 24.** By Theorem 15 and Lemma 25.

2.6 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 $Y_n$ [26, 6]. A natural question is whether the sequent calculus can be used to better understand and further simplify the proofs (cf. [8] [15, §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.3 Interestingly, Lambek, who originally presented his “syntactic calculus” as a tool for mathematical linguistics, also introduced a fully non-associative version [11] (cf. [16, Ch. 4]); one may wonder whether there are any linguistic motivations for semi-associativity, as an intermediate point between these two extremes.

The name “coherence theorem” for Theorem 24 is inspired by the terminology from category theory and Mac Lane’s coherence theorem for monoidal categories [13]. Laplaza [14] extended Mac Lane’s coherence theorem to the situation (very close to Tamari’s) where

---

3 Since circulating an earlier version of this paper [31], I have learned that Jason Reed briefly considered precisely such an extension of the sequent calculus studied here (independently, of course), with left unit and right division [18]. Reed did not remark the connection with the Tamari order, but rather was interested in the apparent failure of the induced logic to satisfy Nuel Belnap’s display property.
there is no monoidal unit and the associator $\alpha_{A,B,C} : (A \otimes B) \otimes C \to A \otimes (B \otimes C)$ 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 [24, 9], cf. [23].) The precise relationship with our proof-theoretic coherence theorem remains to be clarified.

3 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 $\mathcal{Y}_n$ is given by Tutte’s formula (1) for planar triangulations. We will assume some basic familiarity with generating functions.

The problem of “counting intervals” is to compute the cardinality of the set $\mathcal{I}_n = \{ (A,B) \in \mathcal{Y}_n \times \mathcal{Y}_n | A \leq B \}$ as a function of $n$. By the soundness and completeness theorems as well as the frontier invariance property (Prop. 19), each $\mathcal{Y}_n$ 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 $L(z,x) = \sum_{k,n} \ell_{k,n} x^k z^n$ and $R(z,x) = \sum_{k,n} \ell_{k,n} x^k z^n$, 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 $fr(A) = p^{n+1}$.) We write $L_1(z)$ to denote the coefficient of $x^1$ in $L(z,x).

- Proposition 26. $L_1(z)$ 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 $A \implies B$, where $|B| = n$ (and hence $|A| = n$), so $\ell_{1,n} = |\mathcal{I}_n|$ by Theorem 24. ◼

- Proposition 27. $L$ and $R$ satisfy the equations:

\[
\begin{align*}
L(z,x) &= \frac{L(z,x) - xL_1(z)}{x} + R(z,x) \\
R(z,x) &= zR(z,x)L(z,x) + x
\end{align*}
\]

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

- The first summand in (2) corresponds to the contribution from the $*$-rule, which transforms any $A, B, \Gamma \implies C$ into $A \ast B, \Gamma \implies C$. The context in the premise must have length $\geq 2$ which is why we subtract the $xL_1(z)$ 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 $\ast R^{\text{foc}}$ rule, which transforms $\Gamma^{\text{irr}} \implies A$ and $\Delta \implies B$ into $\Gamma^{\text{irr}}, \Delta \implies A \ast B$: the length of the context in the conclusion is the sum of the lengths of $\Gamma^{\text{irr}}$ and $\Delta$, while the size of $A \ast B$ 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 $id^{\text{atm}} : p \implies p$. ◼

- Proposition 28. $L_1(z) = R(z,1).$
Proof. This follows algebraically from (2), but we can interpret it constructively as well. The coefficient of $z^n$ in $R(z, 1)$ is the formal sum $\sum_n r_{k,n}$, 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. 21–23, the operations $\phi[-]$ and $\psi[-]$ realize a 1-to-1 correspondence between derivable sequents of the form $\Gamma^{irr} \to B$ and ones of the form $A \to B$.

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 $R$:

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

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

\[
R(z,x)= x+x^2 z+(x^2+2x^3)z^2+(3x^2+5x^3+5x^4)z^3+(13x^2+20x^3+21x^4+14x^5)z^4+... \\
L_1(z)=R(z,1)=1+z+3z^2+13z^3+68z^4+399z^5+2530z^6+16965z^7+... 
\]

Theorem 29 (Chapoton [3]). $|\mathcal{I}_n| = \frac{2(4n+1)!}{(n+1)!(3n+2)!}$.

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 [4] 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).

Let’s take a moment to discuss Chapoton’s original proof of Theorem 29, which it should be emphasized has many commonalities with the one given here, despite being less structured. 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.\(^4\) By a combinatorial analysis, Chapoton derives the following equation for $\Phi$:

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

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 [4].

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 $R(z, x)$):

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

\(^4\) Or rather at the upper end, since Chapoton uses the dual ordering convention (cf. Footnote 1).
Acknowledgment. I thank Wenjie Fang, Paul-André Melliès, Jason Reed, Gabriel Scherer, and Tarmo Uustalu for interesting conversations and helpful comments about this work. I also thank the anonymous FSCD reviewers for their generous readings of the paper and some useful remarks.

References


