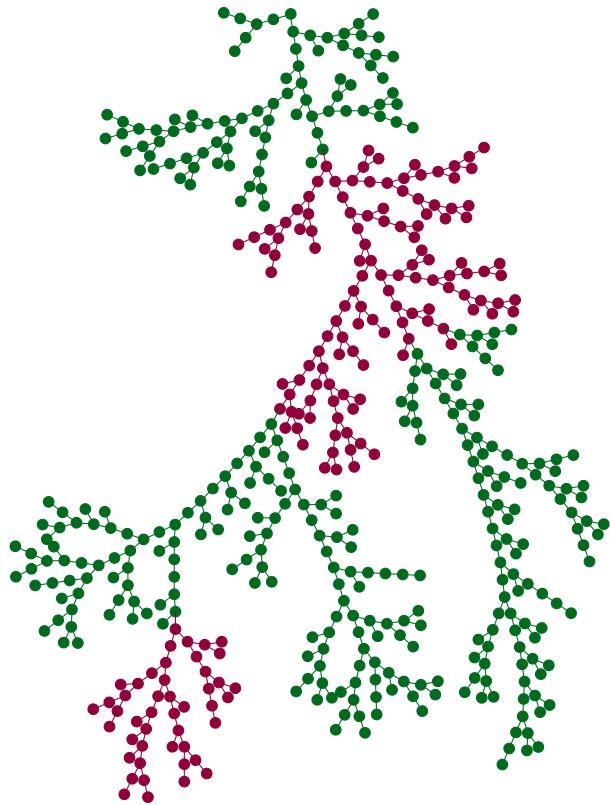


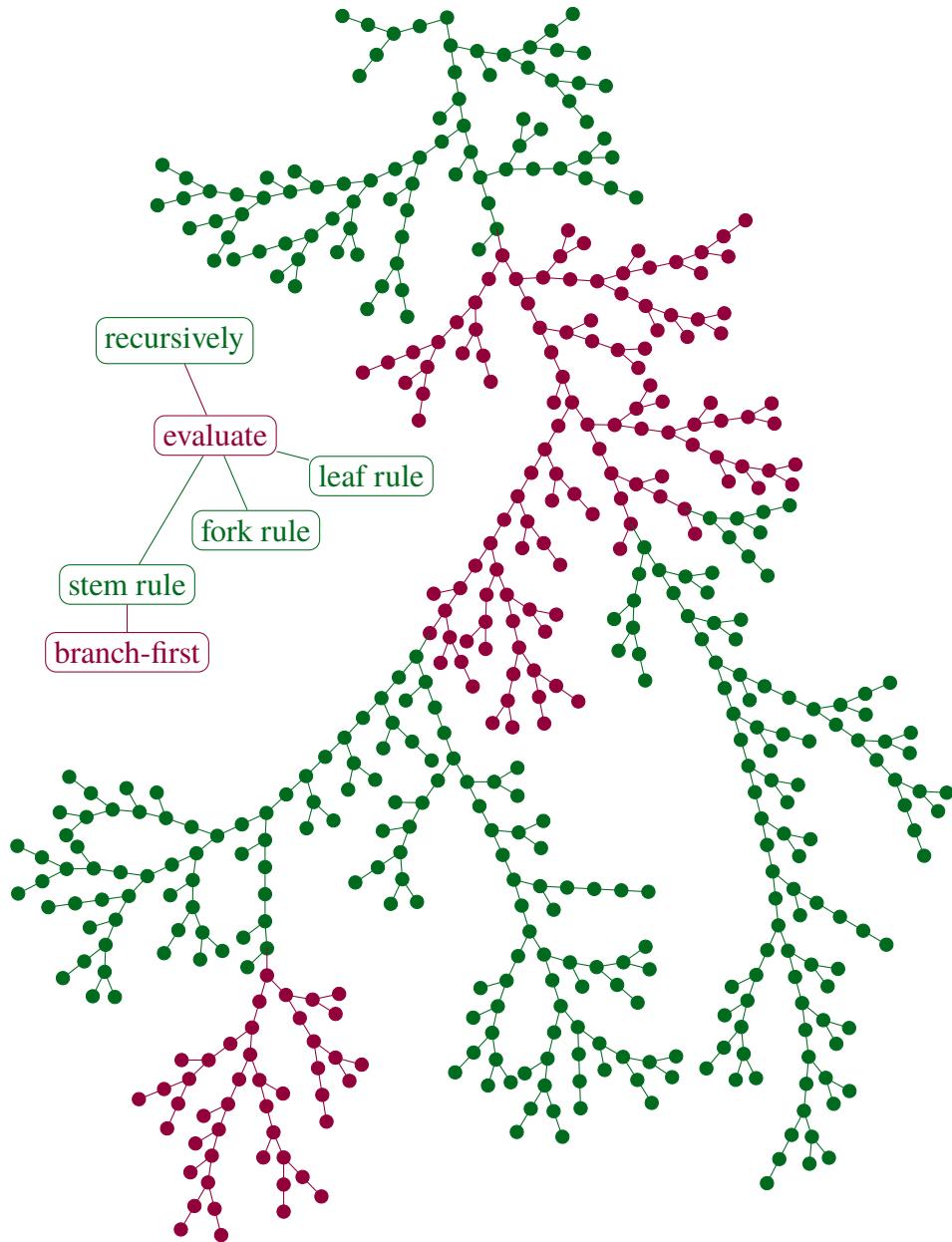
Reflective Programs in Tree Calculus



Barry Jay

©Barry Jay 2021

A BRANCH-FIRST EVALUATOR



To Cathy

who is always there for me

Preface

Tree calculus is seeded by a single operator, whose self-application gives rise to powerful reflective programs including a size program that can compute its own size, an equality program that can decide its own equality, and a self-evaluator that can evaluate itself. This is achieved without any of the usual outside machinery, such as the Gödel numbering of Turing machines, the quotation of lambda-abstractions or serialisation of programs. The resulting theory of computation is both simple and powerful: simple because it is based on three equations only; powerful because one program can query the internal structure of another, using a mix of extensional and intensional techniques.

Trees have long been fundamental to computing practice. Compilers parse strings of characters to syntax trees which are analysed and transformed in various ways before being converted into bits or numbers that can be executed by appropriate hardware. However, syntax trees are quite artificial, with nodes labelled by keywords that vary from one language to another, which makes them unsuitable for a general theory. So we avoid artifice by using *natural trees*, without labels, which are as natural as the numbers but whose branching provides support for program structure; extra machinery such as Gödel numbers or quotation is no longer necessary.

Comparison with tree calculus reveals new insights about traditional models of computation. For example, combinatory logic supports fixpoint functions that are in normal form, that do not reduce until given an argument. Again, there is a new variant of lambda-calculus, called VA-calculus, whose operators represent variables and abstractions, without any explicit mention of scoping rules.

The relative power of tree calculus is shown by giving *meaningful translations* to it from traditional models of computation. Conversely, there is no such translation from tree calculus to combinatory logic. In this sense, combinatory logic is weaker than tree calculus, which conflicts with some versions of the Church-Turing Thesis. The conflict is resolved in the appendix, written a few years ago with my student Jose Vergara.

By eliminating such outside machinery, tree calculus eliminates the boundary between programs and their interpreters, so that compilation techniques may be freely used during program execution without serialisation or other artifice. It will be interesting to see how this develops. More broadly, tree calculus may provide a better foundation for mathematics, in which data structures and functions are uniformly represented.

This book is intended for anyone who is interested in the theory of computation, from students to experts. It is in two parts. Part I introduces tree calculus and explores its expressive power, culminating in the production of self interpreters, using no more than equational reasoning and a little structural induction. Part II considers its relationship to other models of computation, using some basic rewriting theory.

None of the key results in this book have been previously published. Those of Part I and the appendix have been submitted to multiple conferences and journals but spirited debate has never quite resulted in a decision to publish, even when the theorems have been formally verified.

Sydney, Australia
May 2021

Barry Jay

Acknowledgements

Tree calculus emerged as a response to an anonymous negative review which asked, who cares about programs in your *SF*-calculus? This was the spur that pushed me to consider natural trees instead of artificial ones. More generally, every comment in every review has helped to sharpen these ideas.

Thanks go to all those who have read or commented on drafts of this work, including Justin Anderson, Thomas Given-Wilson, Aslak Gronflatén, Ourian Monseigny and Alberto Vergara. Xuanyi Chew and Eric Torreborre read the second draft with great care, and made many valuable observations. I would especially like to thank my old friend Paul Pilipowski. An electrical engineer by profession, Paul tested my claim that high school mathematics was sufficient background, by patiently reading the whole of the first draft. Among other things, he pushed me to include a glossary. Adding illustrations, as a way of lightening the mood, was discussed with Benjamin Marks and Rodney Marks. Simultaneously, Cathy Lockhart assisted at the birth of Shasta and Fir, as I wibble-wobbled my way through the illustrations.

Aside from the appendix written with Alberto, all remaining errors are my sole responsibility.

Contents

I Tree Calculus	1
1 Introduction	3
1.1 Reflective Programs	4
1.2 Tree Calculus	4
1.3 Other Theories of Computation	6
1.4 The Church-Turing Thesis	7
1.5 How to Read this Book	8
2 Equational Reasoning	11
2.1 Addition	12
2.2 Inductive Definitions	14
2.3 Notational Conventions	15
2.4 Equivalence Relations	16
2.5 Congruences	16
2.6 Collecting Like-Terms	17
2.7 Rules for Operators	17
2.8 Arithmetic Expressions	17
2.9 Translation	21
3 Tree Calculus	23
3.1 Syntax Trees	24
3.2 Natural Trees	24
3.3 Tree Calculus	28
3.4 Programs	32
3.5 Propositional Logic	32
3.6 Pairs	34
3.7 Natural Numbers	34
3.8 Fundamental Queries	35
4 Extensional Programs	39
4.1 Combinators, Combinations and Terms	40
4.2 Variable Binding	41
4.3 Fixpoints	43
4.4 Waiting	44

4.5 Fixpoint Functions	46
4.6 Arithmetic	48
4.7 Lists and Strings	48
4.8 Mapping and Folding	49
5 Intensional Programs	53
5.1 Intensionality	54
5.2 Size	54
5.3 Equality	55
5.4 Tagging	57
5.5 Simple Types	58
5.6 More Queries	60
5.7 Triage	60
5.8 Pattern Matching	61
6 Reflective Programs	63
6.1 Reflection	64
6.2 Evaluation Strategies	64
6.3 Branch-First Evaluation	65
6.4 A Branch-First Self-Evaluator	66
6.5 Root Evaluation	68
6.6 Root-and-Branch Evaluation	70
6.7 Root-First Evaluation	70
II Other Models of Computation	73
7 Rewriting	77
7.1 Proving Negative Results	78
7.2 Rewriting	78
7.3 Normal Forms	80
7.4 The Diamond Property	82
7.5 Confluence of Tree Calculus	84
7.6 The Halting Problem	86
7.7 Standard Reduction	86
7.8 Converse Properties of the Self-Evaluators	88
8 Incompleteness of Combinatory Logic	91
8.1 Logic Without Variables	92
8.2 SK-Calculus	92
8.3 Combinators in SK-Calculus	93
8.4 Incompleteness of Combinatory Logic	94
8.5 Meaningful Translation	97
8.6 No Tree Calculus in Combinatory Logic	99

9 Lambda-Abstraction in VA-Calculus	101
9.1 First-Class Abstraction	102
9.2 VA-Calculus	105
9.3 Combinators	106
9.4 Incompleteness	107
9.5 Translation to Tree Calculus	108
9.6 Tagging	108
9.7 Translation from Tree Calculus	110
10 Divide-and-Conquer in SF-Calculus	113
10.1 Divide and Conquer	114
10.2 SF-Calculus	114
10.3 Intensional Examples	115
10.4 Translation from Tree Calculus	117
10.5 Translation to Tree Calculus	119
11 Concluding Remarks	123
11.1 Recapitulation	124
11.2 Program Completeness	124
11.3 Extensional, Intensional and Reflective Programs	125
11.4 Translations to Tree Calculus	126
11.5 Copying is Meta-Theoretic	126
11.6 Implications for Programming Languages	127
11.7 Implications for Logic	127
11.8 Implications for Mathematics	128
11.9 Implications for Traditional Computing Theory	129
A Confusion in the Church-Turing Thesis	131
A.1 Introduction	133
A.2 Confusion Leads to Error	136
A.3 Church's Thesis	137
A.4 Turing's Thesis	139
A.5 The Church-Turing Thesis	140
A.6 Models of Computability	141
A.7 Comparison of Models	142
A.8 Programming Language Design	145
A.9 Extensional and Intensional Computation	146
A.10 SF-Calculus	147
A.11 Conclusions	149
B Shasta and Fir	157
Glossary	159

List of Figures

1.1	Polish Notation for the Branch-First Evaluator	7
2.1	Table of Decimal Addition	12
2.2	Table of Binary Addition	13
2.3	Derivations of Two Roman Numerals	15
3.1	A Word Represented by a Tree of Letters	26
3.2	Tree Equations	28
3.3	The Combinations K, I and D	29
3.4	Two Representations of S	30
3.5	Evaluating $Sxyz$ to $xz(yz)$	31
3.6	Some Boolean Operations	33
3.7	The Zero Test and the Predecessor	35
3.8	Parametrising the Fundamental Queries	37
3.9	The Fundamental Queries	37
4.1	The Program W	45
4.2	The Program $Y_2\{f\}$	47
4.3	The Program plus	48
4.4	The String “small”	49
4.5	Mapping over Lists	50
4.6	Folding over Lists	51
5.1	The Programm size	55
5.2	The Program equal	56
5.3	Tagging	58
5.4	The Program type_check	59
5.5	Pattern Matching	62
6.1	Branch-First Evaluation	65
6.2	A Branch-First Evaluator	66
6.3	Quotation	69
6.4	Root Evaluation	69
6.5	A Root Self-Evaluator	70
6.6	Root-and-Branch Evaluation	70

7.1	Reduction may Duplicate Redexes	84
7.2	Simultaneous Reduction in Tree Calculus	85
7.3	Simultaneous Reduction can Handle Duplicate Redexes	86
7.4	Branch-First Evaluation Using the Fork Rule	89
7.5	Branch-First Evaluation Using the Stem Rule	90
9.1	An Invalid Argument	104
9.2	Informal Rules of VA-Calculus	105
9.3	Reduction Rules of VA-Calculus	105
9.4	The Translation of V to Tree Calculus	109
9.5	The Translation of Δ to VA-Calculus: Circles are V and Squares are A	111
10.1	Reduction rules of SF -calculus	115
10.2	The Translation of Δ to SF -Calculus: Circles are S and Squares are F	118
10.3	$\text{ternary_op}\{f\}$ is Used to Translate S and F to Tree Calculus	121
A.1	A faulty argument, with premises quoted from Kleene [46]	136
A.2	Recoding is not λ -definable	144
A.3	Recoding is SF -definable but not λ -definable	148

Part I

Tree Calculus

Chapter 1

Introduction



Abstract

This chapter provides an overview of the book. Part I shows how tree calculus supports extensional, intensional and reflective programs. Part II provides meaningful translations from other models of computation to tree calculus.

1.1 Reflective Programs

Although the theory of computing began with the study of programs that act on numbers, with numerical analysis, the arrival of compilers and interpreters introduced programs that act on other programs, or program analysis, which is altogether more challenging. In particular, to the degree that an interpreter is written in the same language as its inputs, so that the interpreter can be applied to itself, the language supports *reflective programs* that are able to look in the mirror and see their own structure.

In such a language, programs have a dual nature, as both functions and data structures. This dual nature is hard to capture within the traditional theories of computation, which emphasise one or the other. If programs are defined to be data structures, e.g. to be sequences of instructions, then some external machinery has been required to make them function. For example, Alan Turing showed how to represent a program as a sequence of symbols on a tape, but then some other machinery (a universal machine) is required to interpret it as a function. Alternatively, if a program was defined to be a function, as in the lambda-calculus of Alonzo Church, then external machinery has been required to interpret functions as data structures, e.g. by *serialisation* for their Gödel number or *quotation* for their syntax tree. In all cases, "self-interpretation" is a misnomer, as we must do the interpretation before evaluation can begin. More generally, reflection has required some additional machinery, functionality or theory that lies outside of the theory proper.

In this sense, all of the traditional theories of computation are lacking in some manner. No matter how good they are at data analysis, program analysis lies outside of the theory.

Traditionally, a program analysis is *not* a reflective program.

In practice, this forces most program analysis to be done statically. Source programs are given by data structures, usually syntax trees, which data analysis converts to executable functions. Dynamic analysis is limited by the difficulty of recovering a source program from the executable. More important is that application developers are unable to make their own static analysis, as this is the work of the language implementer. Thus, every optimisation of the evaluation strategy or refinement of type inference (which is, after all, a form of program analysis) requires that the whole programming language be rebuilt. Full support for reflective programs, where source programs and executables can be identified, will allow developers with standard tools to add program analyses, including type systems and evaluation strategies, without having to redefine or re-implement the programming language.

1.2 Tree Calculus

In tree calculus, a program analysis is a reflective program.

Tree calculus computations are unlabelled, finitely-branching trees, called *natural trees*. Like the natural numbers, they are pure structure, but their branching supports modularity in a way that numbers cannot. The slogan is:

Compute with trees, not numbers.

Algebraically, the nodes of a tree are instances of the *node operator* Δ while the edges are left-associative *applications*. Their functionality arises by treating the node constructor Δ as a ternary operator with the three evaluation rules

$$\begin{aligned}\Delta\Delta yz &= y \\ \Delta(\Delta x)yz &= (yz)(xz) \\ \Delta(\Delta wx)yz &= zwx.\end{aligned}$$

Evaluation proceeds according to whether the first argument is a leaf Δ with no branches, a stem Δx with one branch x or a fork with two branches w and x . The *programs* and the *values* they compute, are then the binary trees, in which every node is a leaf, stem or fork.

Although the rules are simple, they are powerful. For example, the frontispiece shows a reflective program for branch-first self-evaluation in which branches of a node are evaluated before the node itself. The main goal of Part I is to develop the machinery required for reflective programs like this. The development proceeds in five steps, in five chapters.

Chapter 2 introduces the equational reasoning tools required to define a self-interpreter. This amounts to little more than the ability to substitute equals for equals, much as in high school algebra, though the settings will be novel. It also introduces inductive definitions, which will be used passively throughout, though an inductive proof will feature at the end of Part I.

Chapter 3 introduces tree calculus and shows how to build tests for being a stem, leaf or fork. It defines the programs and normal forms to be the binary trees. When reduction rules are introduced in Part II, the normal forms will prove to be exactly the irreducible terms.

Chapter 4 considers functions that treat their argument t as a function whose internal structure is hidden. Thus, the only information which can be revealed about t concerns its input-output relation, i.e. its *extension*. Such extensional functions are the traditional preserve of functional programming so it is convenient to develop some machinery for binding parameters (as in lambda-abstraction) and also for defining recursive functions as fixpoints. Unlike traditional accounts, the fixpoint construction here preserves normal forms, and so preserves programs in the sense above.

Chapter 5 considers functions that treat their argument t as a data structure, whose internal structure can be revealed by queries. In particular, queries can discover the algorithm that t implements, its *intension*. Examples include programs that compute the size of a program or test the equality of programs. It also shows how a program may be *tagged* with extra information that does not affect its extensional behaviour, its functionality, but can be recovered by intensional means, using queries. Of course, tags can be used to record the intentions of the programmer, too, including type information.

Chapter 6 produces some self-interpreters, or more precisely, *self-evaluators* for tree calculus. The first uses a *branch-first* evaluation strategy: the last uses a *node-first* evaluation strategy. These are reminiscent of eager and lazy evaluation respectively. Such *reflective* programs use all of the extensional and intensional machinery developed

earlier. Previous accounts of self-interpretation had to work with programs that were not normal forms, that were unstable. Stability was imposed by first *quoting* the program to produce a data structure, by putting on some make-up. In tree calculus, the programs are already data structures, so that no pre-processing is required; both of the self-evaluators above act on the program and its input directly. In short, tree calculus supports honest reflection without make-up.

1.3 Other Theories of Computation

The main goal of Part II is to compare tree calculus to three other theories of computation: combinatory logic, λ -calculus and factorisation calculus. The translation of λ (or, more precisely of the abstraction operator A of VA-calculus) is given by the tree at the beginning of Part II.

All of the models support the same numerical analysis; they are all, in this sense, Turing complete. Traditionally, this was enough to consider them equivalent, but some do not support program analysis or meaningful translations from other theories.

Chapter 7 introduces *rewriting theory* and recasts tree calculus as a rewriting system. This allows us to prove some basic properties of tree calculus, especially its *soundness*, e.g. that forks are never leaves, and will facilitate the comparisons that follow.

Chapter 8 considers *combinatory logic*. Since it is extensional, distinct functions which have the same extent, the same input-output relation, cannot be distinguished. It follows that combinatory logic is unable to decide equality of programs, and so is incomplete for program analysis.

This chapter also introduces *meaningful translations* between rewriting systems, to ensure that translations preserve the rewriting rules, as well as the results of computations. It is easy to give a meaningful translation from combinatory logic to tree calculus. Much harder is to prove that there is *no* meaningful translation from tree calculus to combinatory logic. However, meaningful translations preserve the ability to distinguish, or separate, identity programs. Since tree calculus can, but combinatory logic cannot, there is no meaningful translation.

Chapter 9 introduces *VA-calculus*, an equational variant of λ -calculus built from operators for variables and abstraction. It has evolved from my earlier creation, *closure calculus*, which has been further explored by Xuanyi Chew. It avoids the usual difficulties with binding or scoping by restricting substitution to the environments of abstractions, which are not bound.

Like combinatory logic, it is incomplete for program analysis since it cannot define equality of programs, and there is a meaningful translation from VA-calculus to tree calculus. Interestingly, VA-calculus supports tagging by storing tags as additional bindings in the environments of abstractions. These can never modify the body of the abstraction, but can be recovered by querying the environment, in ways that are not available in most lambda-calculi. In turn, tagging supports a meaningful translation from tree calculus to VA-calculus. It follows that equality of translated programs is definable, even though equality of arbitrary programs is not.

Chapter 10 recalls the factorisation calculus that is *SF-calculus*, developed by Thomas Given-Wilson and me. It supports the same sorts of program analysis as tree

```

212101021202120210102121212021121010212121212110200200201020112121
202121010210102121102002002010200200200202120212021212110201020102
121202121212120212121212121200212121200212121202101021211020020
0200200200200201120202121202101021202021010212021010021211020020020
020020102011212020021212120200212121212020021212120200212121202101021
211020020020020020020111020020020020020102120021212002101020020102
01021212021202101212121212121202002121212020021212121102002002002
00200200201102002002010201121212110200200201020020021202110201002121
20202002120002002120002121202020021202010020021211020020021210102101
0

```

Figure 1.1: Polish Notation for the Branch-First Evaluator

calculus, and there are meaningful translations in both directions. It gives a more direct account of divide-and-conquer algorithms, which simplifies the treatment of data structures and pattern-matching. However, it is less attractive than tree calculus for two related reasons: it is more complicated and; it uses artificial trees (with nodes labelled by S and F) instead of natural trees.

Some other models of computation have *not* been considered for detailed treatment. For example, the μ -recursive functions act on numbers, not μ -recursive functions and so cannot support reflection. Again, Turing machines act on tapes, not Turing machines. Some elaborate encoding is required to patch over these difficulties, with the effect of pushing the computational effort out of sight. For example, Steven Cole Kleene introduced *partial combinatory algebras* to allow numbers to be functions, but their evaluation is not axiomatic, as it requires decoding of numbers to functions. Since it doesn't explain anything, I consider it a model of computation that is not a theory of computation. To illustrate, Figure 1.1 shows the encoding of the branch-first self-evaluator of the frontispiece in Polish notation using ternary numerals (so that, for example, the ternary numeral 201 is the same as the decimal $2 * 3^2 + 1 = 19$). Reconstruction of the function from the numeral must recover all of the tree structure, which is a non-trivial computation. Again, Turing machines can be encoded as tapes, but then the application of one tape to another requires Turing's *universal machine* to do the interpretation.

1.4 The Church-Turing Thesis

In a tradition dating back to Alonzo Church in 1936, the various theories of computation were all deemed to be equivalent, and complete in the sense that they could represent all known algorithms. And yet it was also well understood that algorithms for deciding equality of programs were not definable! To maintain these conflicting beliefs, has required a sensitivity to context similar to that required for the wave-particle duality in quantum physics, where something is either a wave or a particle, depending on the observation being made. The traditional rationale, as given by Steven Cole Kleene, goes something like this.

Any comparison of theories requires a translation which is outside of them, which is meta-theoretic. These translations can reduce all symbolic computation to numerical analysis, at which all the theories are equally competent.

The weakness of this argument is that the hard computational work of inspecting the program structure has been done by the translation. Complaining about this is met with a shrug: since translation is necessarily meta-theoretic, you are free to dislike it, but cannot disqualify it on theoretical grounds.

However, there is no need to consider translation at all! It is enough to consider expressive power within a single theory. If program analysis is within the theory, if a program analysis is admitted as an effectively calculable procedure, then the incompleteness of the traditional theories is confirmed. The only escape is to exclude program analysis from consideration. Given the existence of compilers for programming languages, the centrality of program analysis to computer science, and now the existence of tree calculus, this would be a pretty desperate claim, but it can be made, as follows.

Since the Halting Problem is insoluble, since there is no algorithm for deciding if a computation halts or not, programs are not legitimate inputs to programs. Program analysis is necessarily meta-theoretic.

The gap in this argument is the identification of programs with computations. This identification is traditional in lambda-calculus and combinatory logic, since recursive functions are habitually represented by terms that are themselves non-terminating computations. However, this non-termination is a weakness of the theory, not an argument for ignoring program analysis.

For example, in combinatory logic it is possible to represent recursive functions as computations which have halted, as combinators in normal form. These can be thought of as binary trees with forks labelled by S and with stems and leaves labelled by S or K . If these binary trees are admitted as legitimate inputs then combinatory logic is incomplete as it cannot decide their equality. The only escape is to deny the legitimacy of these trees somehow.

Labelled trees are artificial, not natural like numbers, so can be ignored.

Of course, this claim is at odds with the claim that translation of symbolic expressions to numbers is harmless, but let that pass. The direct answer is that natural trees are every bit as natural as the natural numbers, and so program analysis in tree calculus cannot be ignored.

The first appendix gives a detailed critique of the Church-Turing Thesis. Written some years ago with Jose Vergara, it references *SF*-calculus rather than tree calculus but is otherwise up to date.

1.5 How to Read this Book

By and large, each chapter of the book builds on those before it so, from a logical point of view, there is not much opportunity to skip ahead. For example, reflective programs are

built using a combination of extensional and intensional techniques. Also, the examples, proofs and exercises begin simply but complexity grows significantly. Fortunately, all the proofs of the named theorems have been verified in Coq and can be found among the materials accompanying the book at <https://github.com/barry-jay/personal/tree-calculus/>. These formal proofs are substantial, being larger than the book itself, so perhaps your first reading should focus on the ideas, leaving formal proofs for later.

This book is intended for all those who are interested in the theory of computation, from students to experts. Students do not need any particular exposure to computing theory or mathematics though some awareness of general principles will be useful; there is a glossary of all the main terms mentioned in this introduction. All ideas required for the technical development will be introduced as required but with minimal exposition. Each chapter after the introduction includes some exercises; where appropriate, solutions have been included with the proofs in Coq. Part I begins with an account of equational reasoning and inductive definitions. This is enough to follow all of Part I except the last proof, about the properties of a self-evaluator, which uses induction within a proof. Part II opens with a chapter on rewriting theory, that will be used to compare models of computation. By the end of the book, you will have learned about a wide variety of models of computation, and also some of the standard data structures, including booleans, natural numbers, tuples, lists and strings.

Experts in computing theory, especially in rewriting, may have two possible concerns. On the one hand, many constructions, e.g. of fixpoints, will be familiar, so there is the risk of boredom. On the other hand, many of the theorems conflict with conventional opinions about the nature of computation, so there is the fear of pollution. To counter boredom, I have included something new in every chapter. Here is a partial list.

Chapter 2 uses Roman numerals as a running example of a computing system. They will be familiar to many readers and yet provide a convenient setting in which to explore some of the hidden assumptions of elementary algebra. For example, when the numerals *IV* and *VI* are different numbers (four and six) then the commutativity of juxtaposition does not apply.

Chapter 3 introduces tree calculus, and defines the fundamental queries, for being a leaf, stem or fork.

Chapter 4 introduces the idea of *waiting*. An “application” $\text{wait}\{M, N\}$ must wait for another argument P before applying M to N to get $\text{wait}\{M, N\}P = MNP$. In turn, this is used to define fixpoint functions in normal form, that must wait for their argument before evaluation.

Chapter 5 shows how to tag a function with additional, recoverable information without changing its functionality.

Chapter 6 introduces a self-evaluator without the need for any of the usual outside machinery, such as quotation.

Chapter 7 shows that the halting problem is insoluble within tree calculus, without appealing to external machinery. Also, it shows that the self-evaluators of Chapter 6 do no more evaluation than allowed by the corresponding evaluation strategy.

Chapter 8 shows that there is no meaningful translation of tree calculus to combinatory logic.

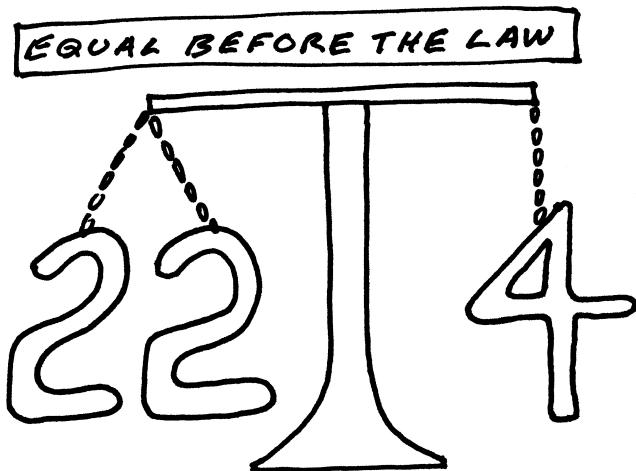
Chapter 9 introduces an equational variant of λ -calculus, in which abstraction is represented by the free-standing operator A .

Chapter 10 recalls SF -calculus, and shows how there is a meaningful translation of it into tree calculus, even though the translation of F is *not* a factorisation function in tree calculus.

The final concern is the fear of pollution. If this book conflicts with what you already know then it is natural to fear its corrupting influence. If your first goal is to consider the clash of ideas or if you want some historical context, then you could begin with the appendix. Written with Jose Vergara in 2015 but refused my most referees, it analyses the weaknesses inherent in the traditional, verbal, account of computation. However, I encourage you to save that till the end. After all, there is nothing to fear from a few simple axioms, some equational reasoning and some pictures. At worst, you will have some new tricks: at best, you will see computation in a new light.

Chapter 2

Equational Reasoning



Abstract

Some of the core issues of computation can be seen in arithmetic. This chapter uses addition to illustrate the fundamentals of computation by equational reasoning, for Roman numerals and then for a small language of arithmetic expressions. The numerals are given by inductive definitions. Their equality will be defined by introducing a sequence of rules, from the general rules for equivalence relations and congruences, to the special rules that carry the meaning of these numerals. Proofs about the translation from Roman numerals to arithmetic expressions proceed by structural induction.

+	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9	10
2	2	3	4	5	6	7	8	9	10	11
3	3	4	5	6	7	8	9	10	11	12
4	4	5	6	7	8	9	10	11	12	13
5	5	6	7	8	9	10	11	12	13	14
6	6	7	8	9	10	11	12	13	14	15
7	7	8	9	10	11	12	13	14	15	16
8	8	9	10	11	12	13	14	15	16	17
9	9	10	11	12	13	14	15	16	17	18

Figure 2.1: Table of Decimal Addition

2.1 Addition

Some of the core issues of computation can be seen in arithmetic. This chapter uses addition to illustrate the fundamentals of computation by equational reasoning, for Roman numerals and then for a small language of arithmetic expressions. The numerals are given by inductive definitions. Their equality will be defined by introducing a sequence of rules, from the general rules for equivalence relations and congruences, to the special rules that carry the meaning of these numerals. Proofs about the translation from Roman numerals to arithmetic expressions proceed by structural induction.

When children begin adding systematically, they commonly begin by memorising a table, as in Figure 2.1. Of course, memorisation can only take us so far, but the power of zero allows us to extend this to a general algorithm, for performing long addition, that uses positional notation and the principle of “carrying one” to add arbitrarily large numbers. For example, we have

$$\begin{array}{r}
 & 2 & 0 & 4 & 7 \\
 + & 1 & 2 & 0_1 & 5 \\
 \hline
 & 3 & 2 & 5 & 2
 \end{array}$$

in which the addition of 7 and 5 produces 12; put down the 2 and carry the 1. Note how the positioning of the zeroes is used to distinguish the 4 tens from the 2 thousands.

This approach works well enough but has two drawbacks. The first is the need to memorise the table. The second is the reliance on positional notation, which is a geometric aid, rather than a logical one.

The size of the table can be reduced from 10×10 to 2×2 by shifting from arithmetic in base ten to base two, or binary arithmetic, in which the only digits are 0 and 1. Now the table of addition is given in Figure 2.2 where 10 now indicates that there is one 2 and no 1’s. Equally, we could use ternary arithmetic, as in Figure 1.1 or some other base.

The tables can be eliminated altogether by using *tallies*. Now $7 + 4$ is represented

+	0	1
0	0	1
1	1	10

Figure 2.2: Table of Binary Addition

by

$$\text{|||||} + \text{|||} = \text{|||||||}$$

in which the lists of tallies are simply merged. Of course, this makes our eyes dance, and becomes completely impractical for long addition.

The Romans improved on the tallies by introducing some symbols for larger numbers, as in the following table

I	V	X	L	C	D	M
1	5	10	50	100	500	1000

Now $7 + 4$ yields

$$VII + IIII = XI .$$

but we don't have to memorise this. Instead, we just need to memorise some definitions

$$\begin{aligned} V &= IIII \\ X &= VV \\ L &= XXXX \\ C &= LL \\ D &= CCCC \\ M &= DD . \end{aligned}$$

Then we have

$$VII + IIII = VIIIIII = VVI = XI .$$

That is, addition is once again given by merging, but then followed by some tidying up. For the sake of simplicity, we will follow the earlier Roman convention in which IV is the same as VI so that

$$I + V = IV = VI .$$

Our long addition example works similarly, in that $2047 + 1205$ becomes

$$\begin{aligned} MMXXXXVII + MCCV &= MMXXXXVIIMCCV \\ &= MMMCXXXVII \\ &= MMMCXXXXII \\ &= MMCCLII . \end{aligned} \tag{2.1}$$

The second equation above groups together all of the M s, all of the C s etc. That is, it collects *like-terms*. The third and fourth equations apply the definitions of X and L

to contract the result. Of course, we could apply the equations differently, to produce 3252 copies of I or $ILCCMM$ but the given answer is traditional.

This system proved to be a good compromise that retains most of the simplicity of the tally system, and most of the power of the Arabic numerals in Figure 2.1. One still has to memorise a little, and it doesn't work for numbers over a few thousand, but overall it's pretty good.

2.2 Inductive Definitions

When studying computation, it is usual to define the set of terms inductively, by specifying some constants and some rules for building new terms from old.

For example, unary numerals are given by a single constant, zero and single rule: if n is a numeral then successor n is a numeral. This information can be expressed in tree form, by

$$\frac{}{\vdash n} \quad \frac{\vdash \text{zero} \quad \vdash \text{successor } n}{\vdash \text{successor } n} .$$

The syntax $\vdash n$ is a *judgment* that n is a (well-formed) numeral. The horizontal lines are *inferences*, that the *conclusion* below the line follows from the *premises* above the line. The whole tree is called a *derivation*. In a more compressed form, we can express the constructions as a Backus-Naur Form, or BNF for short, as follows

$$n ::= \text{zero} \mid \text{successor } n .$$

It asserts that an n is either zero or the successor of some (other) n . In a sense, unary arithmetic adds zero to the system of tallies.

Here is a BNF for the Roman numerals:

$$r, s ::= I \mid V \mid X \mid L \mid C \mid D \mid M \mid (r \ s) .$$

It is to be read as follows. The symbols r and s denote Roman numerals. The possible forms for Roman numerals come after the symbol $::=$ and are separated by vertical bars $|$. There are eight possible forms. Seven of them are the constants I, V etc. The final form is the *concatenation* $(r \ s)$ of two Roman numerals r and s . To support reasoning about Roman numerals, it is important to ensure that each numeral is constructed in a unique manner, so that the order in which concatenation is performed matters. For example, the numerals

$$((XV)I) \quad \text{and} \quad (X(VI))$$

are distinct representations of the number sixteen, just as $15 + 1$ and $10 + 6$ are. The differences can be highlighted by examining their derivations in Figure 2.3.

Of course, the profusion of brackets makes the notation heavy, so let us adopt the conventions that outermost brackets may be dropped and that concatenation associates to the left, so that a numeral of the form

$$r \ s \ t$$

$$\begin{array}{c}
 \frac{\overline{\vdash X} \quad \overline{\vdash V}}{\vdash (XV)} \quad \frac{}{\vdash I} \\
 \hline
 \vdash ((XV)I)
 \end{array}
 \qquad
 \begin{array}{c}
 \frac{\overline{\vdash X}}{\vdash X} \quad \frac{\overline{\vdash V} \quad \overline{\vdash I}}{\vdash (VI)} \\
 \hline
 \vdash (X(VI))
 \end{array}$$

Figure 2.3: Derivations of Two Roman Numerals

is actually

$$((r\ s)\ t) .$$

For example, $((XV)I)$ may be written XVI and $(X(VI))$ may be written $X(VI)$.

We can modularise the inductive definition of Roman numerals by separating the constants, or operators, from the constructed numerals by defining

$$\begin{aligned}
 O &::= I \mid V \mid X \mid L \mid C \mid D \mid M \\
 r, s &::= O \mid r\ s .
 \end{aligned}$$

That is, O represents an arbitrary operator. Roman numerals are either operators or concatenations of Roman numerals. In this manner, we can add or remove operators without changing the general nature of the numerals. When two Roman numerals r and s are identical, and not merely equal according to our rules, we may write this as $r \equiv s$. The same convention applies for any other inductively-defined terms.

2.3 Notational Conventions

This seems like a convenient point at which to mention notational conventions. Upper-case letters such as O, I, C, M may be used to denote operators but we will also use $M, N, P, Q \dots$ to denote combinations of operators. We rely on the context to make clear which meaning is intended. The lower case letters $m, n, p, q, r, s, t, u, v$ will be used to denote terms appearing in some BNF. Other lower case letters x, y, z may be used to denote variables in two distinct ways. For example, when upper case letters represent operators or particular combinations then lower case letters will be used to represent their arguments. For example, if K is an operator its evaluation rule may be given by

$$Kxy = x .$$

Alternatively, we may wish to bind a variable x in some arbitrary term t as in

$$\lambda^* x. t .$$

In this case, the syntactic class of terms will include a sub-class of variables such as x which are *not combinations*. Since these situations tend not to overlap, we will rely on the context to make clear which conventions are being used.

Words in sans-serif font, such as zero, successor, quote, depthfirst are used to represent programs. Further, curly braces may be used to indicate the parameters of a program, e.g. $\text{wait}\{M, N\}$. In general, spaces between terms represented by single letters may be elided if this eases legibility, so that $r\ s$ may be written rs . Of course, we may *not* write quotes for quote s or unquote for $u\ n$ quote.

2.4 Equivalence Relations

Now let us justify the equational reasoning of (2.1) by defining the equality relation $=$ on the inductively defined Roman numerals above. This will proceed in four steps, as we introduce the rules for: equivalence relations; congruences; collecting like-terms; and characterising the operators. We will quickly cover all of this material here, and use it to understand the Roman numerals, but if you haven't seen any of it before then you may want to consult a textbook on discrete mathematics.

Let T be a set, whose elements are to be thought of as terms. Let \sim be a binary relation on T that is written *infix*, so that if s and t are elements of T then $s \sim t$ asserts that s and t are related by \sim . Then \sim is an *equivalence relation* if it satisfies the following three rules:

(Reflexive) $t \sim t$ for all t in T .

(Symmetric) If $s \sim t$ then $t \sim s$ for all s and t in T .

(Transitive) If $r \sim s$ and $s \sim t$ then $r \sim t$ for all r and s and t in T .

In particular cases, we may use the symbol $=$ instead of \sim . For example, the chain of equalities in (2.1) implies

$$MMXXXXVIIMCCV = MMMCCLII$$

by transitivity.

2.5 Congruences

Where T above is inductively defined, then an equivalence relation is a *congruence* if equals can be replaced by equals in all of the inductive constructions. For Roman numerals, there is only one (non-trivial) construction, so we make $=$ a congruence by adding the following rule

(Congruence) If $r_1 = r_2$ and $s_1 = s_2$ then $r_1s_1 = r_2s_2$ for any Roman numerals r_1, r_2, s_1 and s_2 .

For example,

$$LLVV = CX$$

since $LL = C$ and $VV = X$.

2.6 Collecting Like-Terms

Also, we need rules for collecting like terms. The rules for associativity and commutativity are

- (i) (Associative) $(r s) t = r (s t)$.
- (ii) (Commutative) $r s = s r$.

In our example, these rules are used as follows.

$$\begin{aligned}
 & (MMXXXVII)(MCCV) \\
 &= (MXXXVII)M(MCCV) \quad (\text{commutativity, congruence}) \\
 &= (MXXXVII)(MMCCV) \quad (\text{associativity}) \\
 &= \dots \\
 &= MMMCCXXXVII .
 \end{aligned} \tag{2.2}$$

2.7 Rules for Operators

Having collected the like-terms, we finish up with some rules for simplification, that characterise the operators in terms of I :

$$\begin{aligned}
 V &= IIII \\
 X &= VV \\
 L &= XXXX \\
 C &= LL \\
 D &= CCCC \\
 M &= DD .
 \end{aligned}$$

In our example, this supports

$$MMMCCXXXVII = MMMCCXXXII = MMCCLII .$$

This completes our description of the equality of Roman numerals: three rules for the equivalence relation; one for the congruence; two for collecting like-terms; and six for characterising the operators.

2.8 Arithmetic Expressions

The equality of Roman numerals requires, as well as the rules for congruence, eight rules for collecting like terms and characterizing the operators. This adds eight cases to simple proofs about equality, and dozens of steps to proofs that consider pairs of equations.

We can simplify the system by representing the numbers as unary numerals, built from zero and successor, while the concatenation of Roman numerals will be represented by addition.

Define the *arithmetic expressions* by

$$m, n ::= \text{zero} \mid \text{successor } m \mid (m + n)$$

As before, we eliminate superfluous brackets, on the understanding that the infix operators $+$ associates to the left, and that `successor` binds more tightly than $+$. For example,

$$\text{successor}(\text{successor } m) + n + p = (((\text{successor}(\text{successor } m)) + n) + p).$$

The arithmetic expressions that are constructed from `zero` and `successor` only, without addition, are the *unary numerals*.

The equality relation $m = n$ is defined as follows. First, it is an equivalence relation (reflexive, symmetric and transitive). Second, it is a congruence. Since there are three non-trivial ways to build arithmetic expressions, this requires the following three rules.

- (i) $\text{zero} = \text{zero}$.
- (ii) If $m = n$ then $\text{successor } m = \text{successor } n$.
- (iii) If $m_1 = n_1$ and $m_2 = n_2$ then $m_1 + m_2 = n_1 + n_2$.

Note that there are two rules asserting that `zero` equals itself, namely the reflexive rule and the congruence rule. Finally, we need rules that characterise addition.

$$\begin{aligned} \text{zero} + n &= n \\ \text{successor } m + n &= m + (\text{successor } n) \end{aligned}$$

These rules cover all the possible ways of adding numerals, but do not give rules for expressions in general. For example, there is no specific rule for simplifying $m + n + p$. Rather, we must first simplify $m + n$ to become a numeral, so that the rules above will apply.

To check that this is always possible requires some proofs. We begin by considering the addition of numerals.

Theorem 1 *If m and n are unary numerals then there is a unary numeral p such that $m + n = p$.*

Proof. The proof is by *mathematical induction* on m . In order to show that a property is true of all unary numerals, it is enough to show two things:

- (i) `zero` has the property.
- (ii) If m has the property then so does `successor` m .

For this theorem, the property of m is that:

For all numerals n there is a numeral p such that $m + n = p$.

Suppose m is zero and let n be any numeral. Then choose p to be n . Then $m + n$ is zero + n which equals n by the first rule for addition.

Suppose m is of the form successor m_1 and let n be any numeral. Then m_1 has the property with respect to the numeral successor n . Hence, there is a numeral p_1 such that $m_1 + \text{successor } n = p_1$. Choose p to be p_1 . Then $m + n$ is successor $m_1 + n$. Now successor $m_1 + n = m_1 + \text{successor } n$ by the second rule for addition, and $m_1 + \text{successor } n = p$ so, by transitivity, we have $m + n = p$ as required. \square

Theorem 2 *Every arithmetic expression is equal to a numeral.*

Proof. Let e be an arithmetic expression. The proof is by *structural induction*, i.e. by induction on the structure of the expression e . This is a generalization of mathematical induction. In order to show that a property is true of all arithmetic expressions, it is enough to show three things:

- (i) zero has the property.
- (ii) If m has the property then so does successor m .
- (iii) If m and n have the property then so does $m + n$.

For this theorem, the property of e is that there is a numeral n such that $e = n$. Here are the three cases.

If e is zero then choose n to be zero. Then zero = zero by reflexivity.

If e is successor e_1 and e_1 has the property, then there is a numeral n_1 such that $e_1 = n_1$. Choose n to be successor n_1 . Then $e = n$ by congruence.

If e is of the form $e_1 + e_2$ and e_1 and e_2 have the property then there are numerals n_1 and n_2 such that $e_1 = n_1$ and $e_2 = n_2$. By Theorem 1 there is a numeral n such that $n_1 + n_2 = n$. Then

$$e = e_1 + e_2 = n_1 + n_2 = n$$

by reflexivity, congruence, and the equality above, so $e = n$ by transitivity. \square

Here are some further results which strengthen our confidence in this account of addition.

Theorem 3 *For all numerals m and n we have*

$$m + \text{successor } n = \text{successor } (m + n).$$

Proof. The proof is by induction on m . If m is zero then we have zero + successor n = successor n = successor (zero + n) by two applications of the first rule for addition, and congruence.

If m is of the form successor m_1 then

$$\text{successor } m_1 + \text{successor } n = m_1 + \text{successor } (\text{successor } n)$$

by the first rule of addition. Now the induction principle for m_1 with respect to successor n yields

$$m_1 + \text{successor } (\text{successor } n) = \text{successor } (m_1 + \text{successor } n)$$

which is equal to successor (successor $m_1 + n$) or successor ($m + n$) by congruence and the first rule for addition. \square

Theorem 4 zero is a unit for addition of unary numerals. That is, for all unary numerals n we have

$$\text{zero} + n = n = n + \text{zero} .$$

Proof. The first equation holds by the first law of addition. For the second equation, proceed by induction on the structure of n .

If n is zero then we have $n + \text{zero} = \text{zero} + \text{zero} = \text{zero}$ by the first rule for addition.

If n is some successor successor n_1 then

$$\begin{aligned} n + \text{zero} &= \text{successor } n_1 + \text{zero} \\ &= n_1 + \text{successor zero} \\ &= \text{successor } (n_1 + \text{zero}) \\ &= \text{successor } n_1 = n . \end{aligned}$$

by the first rule of addition, and Theorem 3 and the induction hypothesis. \square

Theorem 5 Addition is commutative. That is, for all arithmetic expressions e_1 and e_2 we have

$$e_1 + e_2 = e_2 + e_1 .$$

Proof. By Theorem 2, we may assume that e_1 and e_2 are numerals n_1 and n_2 respectively. Now proceed by induction with respect to n_1 , to establish the following property:

For all numerals n_2 , we have $n_1 + n_2 = n_2 + n_1$.

If n_1 is zero then apply Theorem 4. If n_1 is some successor successor n_3 then we have

$$\begin{aligned} n_1 + n_2 &= \text{successor } n_3 + n_2 \\ &= n_3 + \text{successor } n_2 \\ &= \text{successor } n_2 + n_3 \\ &= n_2 + \text{successor } n_3 = n_2 + n_1 . \end{aligned}$$

\square

Theorem 6 Addition is associative. That is, for all arithmetic expressions e_1, e_2 and e_3 , we have

$$(e_1 + e_2) + e_3 = e_1 + (e_2 + e_3) .$$

Proof. Without loss of generality, e_1, e_2 and e_3 are unary numerals n_1, n_2 and n_3 respectively. The proof is by induction on the structure of n_1 . If n_1 is zero then the

result is straightforward. If n_1 is some successor successor n_4 then we have

$$\begin{aligned}
 (n_1 + n_2) + n_3 &= (\text{successor } n_4 + n_2) + n_3 \\
 &= (n_4 + \text{successor } n_2) + n_3 \\
 &= n_4 + (\text{successor } n_2 + n_3) \\
 &= n_4 + (n_2 + \text{successor } n_3) \\
 &= n_4 + \text{successor } (n_2 + n_3) \\
 &= \text{successor } (n_4 + (n_2 + n_3)) \\
 &= ((\text{successor } n_4) + (n_2 + n_3)) \\
 &= n_1 + (n_2 + n_3).
 \end{aligned}$$

□

2.9 Translation

In what sense do the Roman numerals and the unary numerals represent the same things? They both represent whole numbers, but the Roman numerals do not represent zero, so there are some differences. The relationship can be expressed by providing a translation $[-]$ from the Roman numerals to the arithmetic expressions that preserves all of the important information. Deciding what is important will be discussed more thoroughly in Part II, where translations play a central role but as the Roman numerals will not feature there, it seems best to develop this example immediately.

In this setting, a translation should map $r s$ to $[r] + [s]$ and it should preserve equality. There is a strong argument for requiring that I is translated by successor zero too. Here is a translation which will prove to have these properties.

$$\begin{aligned}
 [I] &= \text{successor zero} \\
 [V] &= [I] + [I] + [I] + [I] + [I] \\
 [X] &= [V] + [V] \\
 [L] &= [X] + [X] + [X] + [X] + [X] \\
 [C] &= [L] + [L] \\
 [D] &= [C] + [C] + [C] + [C] + [C] \\
 [M] &= [D] + [D] \\
 [r s] &= [r] + [s].
 \end{aligned}$$

The first and last of the three properties above are true by definition.

Theorem 7 *Let r and s be two Roman numerals. If $r = s$ then $[r] = [s]$.*

Proof. Once again, the proof is by structural induction. However, the proof is *not* by induction on the structure of r or s . Rather, it is by induction on the structure of the

proof of $r = s$. It may help to think of this proof as a tree, so that it is the structure of the tree which underpins the induction.

Recall that there are twelve rules for equality of Roman numerals: three for equivalence; one for congruence; two for collecting like terms; and six for characterising the operators. However, the proof is not so long.

Equality of arithmetic expression is also an equivalence relation, so the first three cases are immediate.

For congruence, suppose that $r_1 = s_1$ and $r_2 = s_2$. Then we have

$$[r_1 r_2] = [r_1] + [r_2] = [s_1] + [s_2] = [s_1 s_2]$$

where the middle equation holds by induction.

The rules for associativity and commutativity are preserved by Theorems 6 and 5.

The rules for characterizing the operators are preserved by the way in which the translation was defined. For example,

$$[X] = [V] + [V] = [VV].$$

□

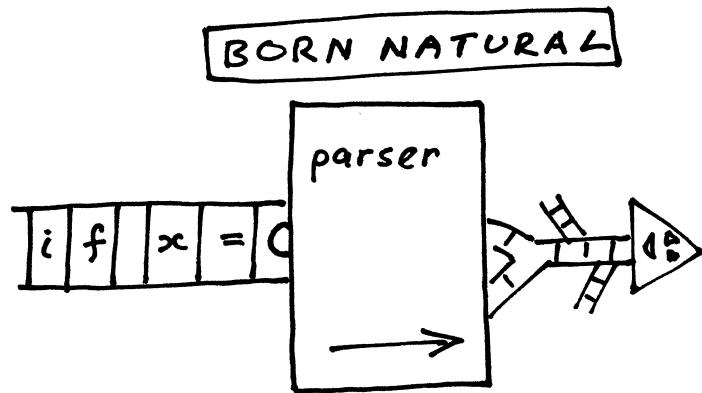
We do not expect there to be a translation in the opposite direction since there is no Roman numeral for zero. To prove this, it is enough to show that there is *no* Roman numeral r such that $rI = I$. However, such proofs are not easy: even proving that II is not equal to I is difficult. We shall return to this challenge after developing some more machinery, in Chapter 7.

Exercises

1. Convert the decimal numbers 2047 and 1205 to binary notation, add them together, and confirm that the result is the binary notation for the decimal numeral 3252.
2. Write out the 3×3 addition table for ternary arithmetic in which $2 + 2$ is 11.
3. Produce the tree derivation of the Roman numeral $CCLX$. Hint: add all the missing brackets to the numeral first.
4. Expand the equational reasoning in (2.1) so that each equation is justified according to the rules for equivalence, congruence etc. Don't forget the uses of reflexivity and symmetry. If you wish, use a tree of derivations.
5. Define one to be successor zero and two to be successor one and three to be successor two and four to be successor three. Now prove that two + two = four.
6. Using the theorems in the text, prove that $m + n + p = p + n + m$ for all arithmetic expressions m, n and p . Hint: add all the superfluous brackets first.
7. (Hard) In most presentations of Roman numerals, IV represents four, rather than six, so that there is a hint of positional notation involved. Describe how you would adapt the developments in this chapter to cope with this.

Chapter 3

Tree Calculus



Abstract

Central to the life-cycle of a program is its representation as a syntax tree. By abstracting away the syntax, we are left with the unlabelled trees, the natural trees. These can be viewed as computations, with the binary trees playing the roles of data structures, values and programs. This leads to tree calculus, which has one operator and three evaluation rules. Some simple examples are given.

3.1 Syntax Trees

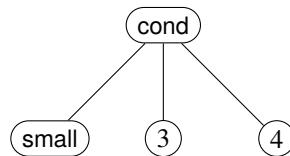
The standard life-cycle of a program begins with a sequence of characters which are chunked into a sequence of tokens, parsed to form a syntax tree, analysed, optimised and then converted to an executable for some machine. For example, consider the following program fragment. The sequence of characters

[i,f, ,s,m,a,l,l, ,t,h,e,n, ,3, ,e,l,s,e, ,4]

becomes the sequence of tokens

[if,small,then,3,else,4]

and then parsing produces the syntax tree



in which the if-then-else is represented by a node of the tree that has three branches, each of which has a label. If analysis shows that `small` is always true then this tree may be optimised to 3 but if `small` is a variable then the resulting executable will include the conditional.

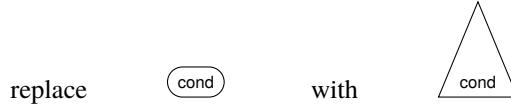
Reflective programming allows this sort of program analysis to be done after the program has been converted to an executable, by recovering the syntax tree from the executable, optimising and then converting back to executable form. However, this is not well presented in traditional theories of computation. In the best case scenario, the language of the program being analysed is the same as the language in which the analysis is written but even then, the conversion functions lie outside of the language in some sort of meta-theory. For example, if the programs are terms of a traditional λ -calculus (see Chapter 9) then *quotation* is required to convert abstractions into trees.

Tree calculus will adopt a different approach, by avoiding the conversions from trees to executables altogether. Rather, the trees will be themselves executables, will be functions as well as data structures. Then the challenge is to represent function application as an operation on trees.

3.2 Natural Trees

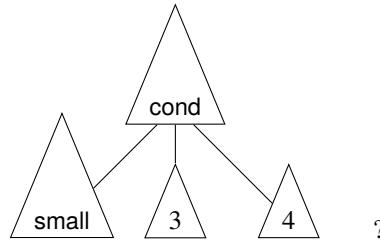
The first issue then is to decide which tokens will be used to label the nodes of our trees. In the example of the previous section, the nodes were labelled by the keyword `cond`, the identifier `small`, and the numerals 3 and 4. In general, there will be many different syntactic constructions and so many different sorts of labels. Fortunately, we do not need to commit to some choice, since the labelled trees can be represented by unlabelled trees, here called *natural trees*, by analogy with the natural numbers. In

pictures,

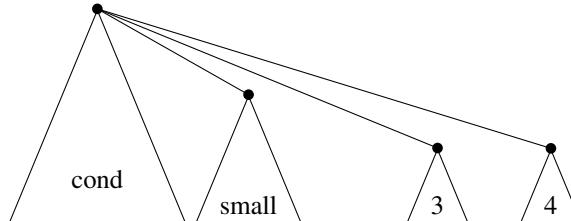


so that the label *cond* (represented by the rounded rectangle) is replaced by a tree (represented by the triangle) whose name is *cond*. In the same manner, we can replace the labels *small* and 3 and 4 with little trees, represented by named triangles.

That done, should the display of the original program fragment be



This would suggest that the subordinate trees have the status of leaves of *cond* when in fact they are branches, as was clear when *cond* was a mere node with a label. To avoid this suggestion, their status as branches will be emphasised by adding a dot to represent the root of each named tree, and drawing the edges between the roots of the trees. Thus, the program fragment becomes



so that we can see that the sub-trees have been grafted to the root of the main tree.

Since there are only finitely many keywords, it is easy to designate some particular, small trees to represent them. However, there is an unbounded collection of numbers and identifiers to be represented too. Happily, these can be represented using a finite alphabet. For example, the identifier *small* could be represented by the tree in Figure 3.1. Note that the large dots in the picture are actual, free-standing nodes, while the small dots denote the roots of the various sub-trees.

All of the natural trees can be built from unlabelled nodes by grafting new branches in the style developed above. While the dot notation for nodes is good in graphics, it reads poorly in text, so we will there use a small triangle Δ to represent a node. Although its name is "node" you may think of this as the Greek letter delta (or Δ) on the understanding that, like a tree, the delta of the river Nile has branches. The grafting of a new branch *N* to an existing tree *M* is represented by the *application* *MN*. Application associates to the left, so that *MNP* is the same as $(MN)P$ but *not* the same

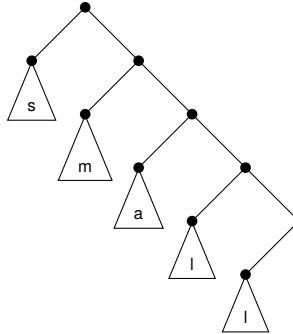
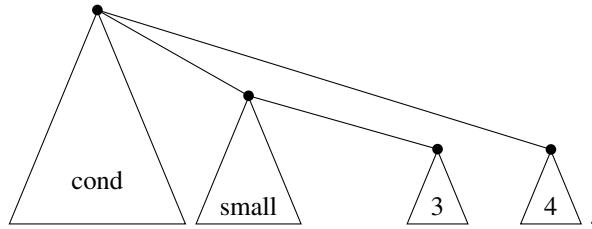


Figure 3.1: A Word Represented by a Tree of Letters

as $M(NP)$. For example, the iterated application $\text{cond small } 3\ 4$ is pictured above, while $\text{cond } (\text{small } 3)\ 4$ is pictured by



In this manner, syntax trees can be represented by natural trees so the question becomes how to compute with them.

The natural trees can be represented algebraically by the following BNF:

$$M, N ::= \Delta \mid MN .$$

This asserts that every tree given by the BNF is either Δ or an application MN of one tree to another. Application is left-associative. Pictorially, each instance of Δ provides a node of the tree, while each application MN adds an edge from the root of M to the root of N .

One way of computing with natural trees is to treat them all as potential outputs, but then the functions which act on trees, i.e. the programs, and the machinery for evaluation must be elsewhere. This leads to a traditional separation of functions from data, which will not help to build reflective programs. Rather, let us consider how the trees themselves can be the programs, can be functions, so that the computation of program M on input N is given by the application MN of M to N . In this sense, the trees also play the role of *computations* that require evaluation. The question is then to decide the nature of the results produced by evaluation and, indeed, what constitutes a program or its inputs.

Since tree calculus is to be reflective, the programs must also qualify as inputs. Since it is important to be able to compose programs, the inputs should be the same as

the outputs. So allow the programs, inputs and results form a single sub-class of trees that we may call the *values* or *programs*.

To discuss the choice of values it will be useful to introduce some terminology. A natural tree which is a node without any branches is a *leaf* or sometimes a *kernel*. A natural tree with one branch is a *stem*. A natural tree with two branches is a *fork*. Since the order of the branches is important, a fork has a *left branch* and a *right branch*.

If the only value is a leaf then no distinctions between values are possible. If the values are expanded to allow stems built from other stems then the values are *chains* of the form

$$\Delta(\Delta(\Delta \dots (\Delta\Delta))) .$$

These correspond to the natural numbers. More conveniently, let us define

$$M^k(N) = M(M \dots (MN) \dots) \quad \text{where } M \text{ is applied } k \text{ times} .$$

Then we can identify the natural number k with $\Delta^k \Delta$.

This approach leads to a traditional account of computation built around natural numbers. The difficulty with this is that chains have no obvious internal structure, so that elaborate encoding and decoding is required to represent program structure. These operations are themselves computations which then require their own explanation, so this cannot be considered progress.

The next simplest option is for the values to be the *binary trees*: the leaves, stems and forks whose branches are also binary trees. As the forks are easily able to represent internal structure, this is the choice made for tree calculus. It follows that trees containing a node with three or more branches is a computation that is not a value, and so must be evaluated. Symbolically, this implies that any tree of the form ΔMNP requires evaluation. If M, N and P are values then ΔMN is a program which is applied to the input P . In other words, the node Δ is a ternary operator, i.e. one that requires three arguments to trigger evaluation whose goal is to produce a binary tree.

Now let us consider how we might formulate the computation rules. If the evaluation does not query the internal structure of its arguments x, y and z then we are led to produce a single rule of the form

$$\Delta xyz = t$$

for some tree t built from the trees x and y and z . However, this would limit the expressive power to that of combinatory logic, which cannot support reflection (see Chapter 8), so we must *query* the internal structure of the arguments. This opens up a world of possibilities. The first argument x could be a leaf, stem or fork, whose branches could also take one of three forms, etc. Similar remarks apply for y and z . Even if we limit the inspection to just the first layer of structure in each argument, there are $3 \times 3 \times 3 = 27$ possible left-hand sides for equations! Fortunately, there is no need to explore too far. Complex queries can be broken down into a sequence of simple queries, with each one considering just one level of structure of just one argument. Since English is written from left to right, it is customary to inspect arguments from left to right as well, so we will consider the internal structure of the first argument x above.

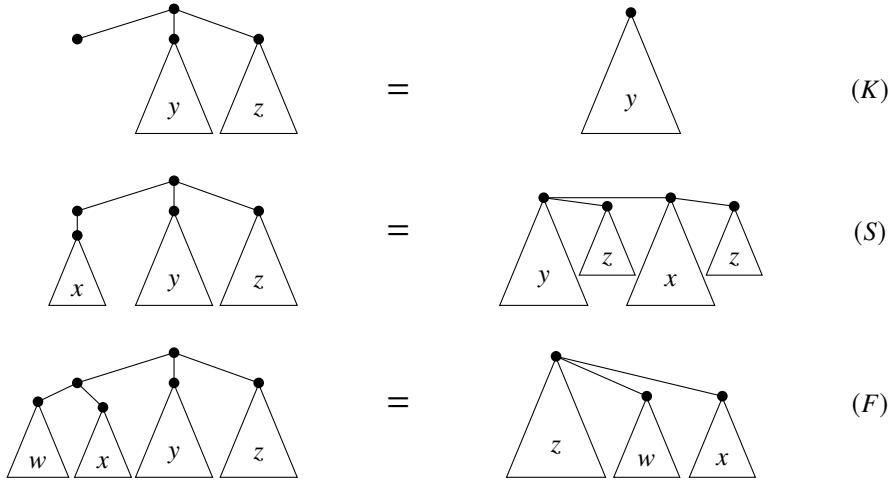


Figure 3.2: Tree Equations

Thus, the left-hand sides of the three rules take the form

$$\begin{aligned}\Delta\Delta yz &= \dots \\ \Delta(\Delta x)yz &= \dots \\ \Delta(\Delta wx)yz &= \dots\end{aligned}$$

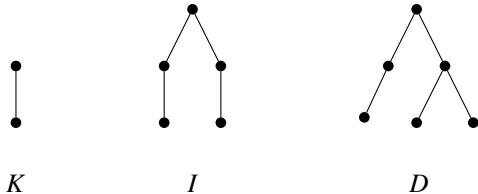
3.3 Tree Calculus

The combinations of tree calculus are the natural trees, built from the operator Δ and application, as described by the BNF for natural trees in the previous section. Its evaluation rules are inspired by earlier calculi, specifically, the operators K and S of combinatory logic (Chapter 8) and the operator F of SF -calculus (Chapter 10). Happily, the operator K inspires the *kernel rule* (K), the operator S inspires the *stem rule* (S) and the operator F inspires the *fork rule* (F). The rules are:

$$\begin{array}{lll}\Delta\Delta yz & = & y \\ \Delta(\Delta x)yz & = & yz(xz) \\ \Delta(\Delta wx)yz & = & zwx\end{array}\begin{array}{r}(K) \\ (S) \\ (F)\end{array}$$

These rules can be expressed pictorially, using graphs, in Figure 3.2.

Note that (K) deletes the third argument, (S) duplicates the third argument, and (F) decomposes the first argument to expose its branches. These rules will prove to be enough to represent the calculi that inspired them. For example, we have the

Figure 3.3: The Combinations K , I and D

combinations

$$\begin{aligned} K &= \Delta\Delta \\ I &= \Delta(\Delta\Delta)(\Delta\Delta) \\ D &= \Delta(\Delta\Delta)(\Delta\Delta\Delta). \end{aligned}$$

Their pictures are given in Figure 3.3. They satisfy the equations

$$\begin{aligned} Kyz &= \Delta\Delta yz = y \\ Ix &= \Delta(\Delta\Delta)(\Delta\Delta)x = \Delta\Delta x(\Delta x) = x \\ Dxxyz &= \Delta(\Delta\Delta)(\Delta\Delta\Delta)xyz \\ &= \Delta\Delta\Delta x(\Delta x)yz \\ &= \Delta(\Delta x)yz \\ &= yz(xz). \end{aligned}$$

Since Dx already simplifies to $\Delta(\Delta x)$ it is useful to define

$$d\{x\} = \Delta(\Delta x)$$

to make combinations easier to read.

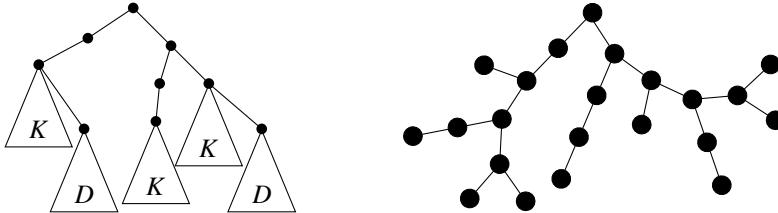
As well as duplicating its third argument, the stem rule (and D) permute the order in which the first two arguments occur, which is not the obvious choice. However, this approach has the advantage of being able to decide if a program is a stem or not, since $yz(xz)$ in the rule for stems allows us to use the behaviour of y to expose the fact that x appeared in a stem, while the behaviour of $xz(yz)$ is determined by that of the unknown x . This will be important for identifying the leaves, stems and forks.

Also, the permutation is easily programmed away, to support the other traditional operator S of combinatory logic (see Chapter 8). It is enough to find a tree S such that

$$Sxyz = xz(yz).$$

The equation can be solved by eliminating the variables one at a time, as follows.

$$\begin{aligned} Sxyz &= xz(yz) \\ &= Dyxz \end{aligned}$$

Figure 3.4: Two Representations of S

so we must solve $Sxy = Dyx$ by

$$\begin{aligned} Sxy &= Dyx \\ &= Dy(Kxy) \\ &= D(Kx)Dy . \end{aligned}$$

These equations are more easily checked from right-to-left, but the given order moves copies of y ever further to the right, to be collected at the argument position, from which it can be eliminated. It remains to solve $Sx = D(Kx)D$. For this step, we must represent $D(Kx)$ and D as functions of x by

$$\begin{aligned} Sx &= D(Kx)D \\ &= (KDx)(Kx)(KDx) \\ &= DK(KD)x(KDx) \\ &= D(KD)(DK(KD))x \end{aligned}$$

so we have S being $D(KD)(DK(KD))$ which evaluates to

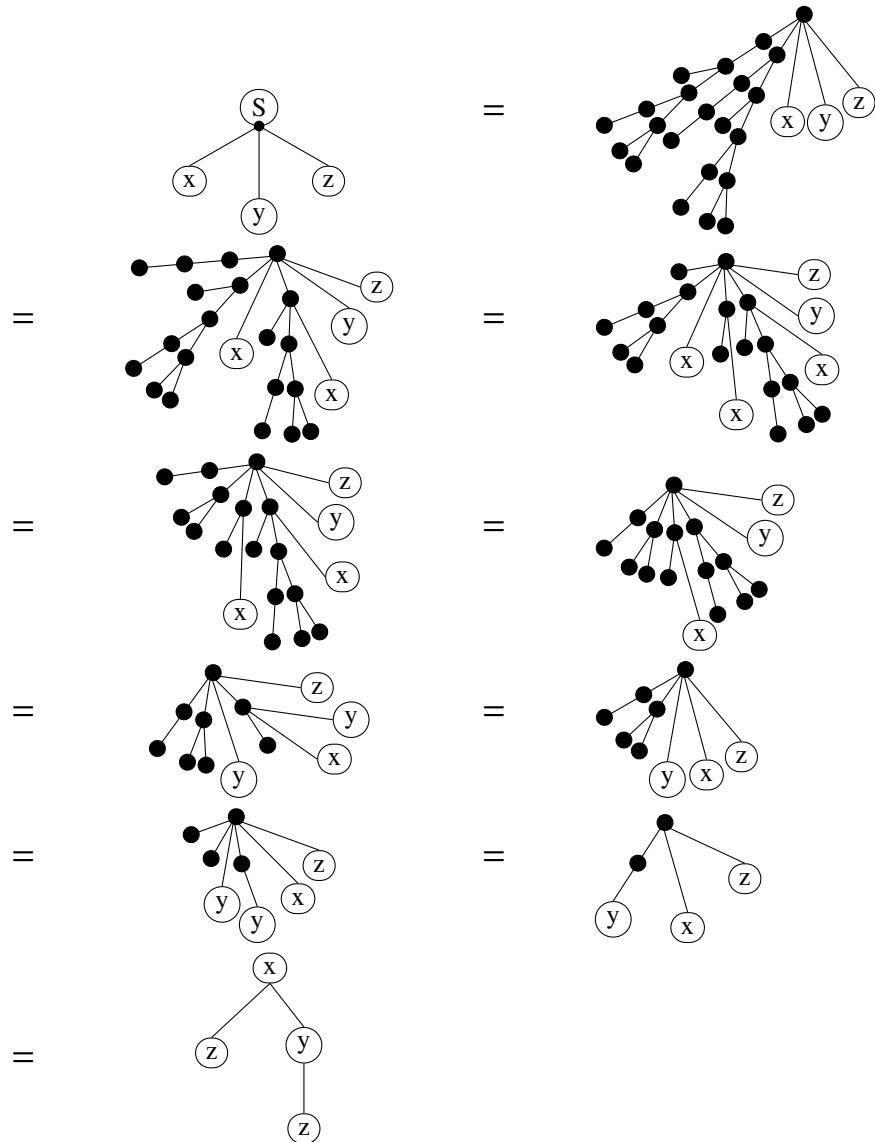
$$S = d\{KD\}(d\{K\}(KD)) . \quad (3.1)$$

Figure 3.4 shows two representations of the tree S . That on the left uses a mix of bare nodes and named branches. That on the right uses bare nodes only. Each has its advantages. Note that the precise angles of forks, and the relative weight given to nodes and edges may have a big impact on the look of a tree but they have no semantic significance.

We can check this account of S by evaluating its application to three arguments:

$$\begin{aligned} Sxyz &= d\{KD\}(d\{K\}(KD))xyz \\ &= d\{K\}(KD)x(KDx)yz \\ &= KDx(Kx)(KDx)yz \\ &= D(Kx)Dyz \\ &= Dy(Kxy)z \\ &= Dyxz \\ &= xz(yz) . \end{aligned}$$

These equations can also be visualised, in Figure 3.5.

Figure 3.5: Evaluating $Sxyz$ to $xz(yz)$

3.4 Programs

We define the *programs* of tree calculus to be the binary trees. They can be given by the BNF

$$p, q ::= \Delta \mid \Delta p \mid \Delta pq .$$

More precisely, every program *is* a computation of tree calculus, so this BNF describes those computations which are programs.

It is not immediately obvious that this definition accords with the usual intuitions about programs. To the extent that programs are distinct from arbitrary terms of a calculus, they should be more easily analysed and optimised by, say, a compiler, while still maintaining the desired expressive power of the underlying calculus. Certainly, it is easy to work with binary trees, especially as they are stable; none of the evaluation rules can be applied from left to right. The important question is whether they can maintain expressive power. This will emerge throughout the book, as programs are defined for equality of arbitrary programs, for the operators of abstraction calculus, for pattern-matching, etc. These are all examples of *recursive programs*. Traditionally, recursive programs that act on other programs are *not* stable, so the identification of programs with binary trees may be surprising. Nevertheless, it will all work out. The definition of programs is given here as it will simplify the development. Its suitability will emerge as we go along.

3.5 Propositional Logic

Propositional logic determines the truth of compound propositions, built using conjunction, disjunction, implication, negation and equivalence, from the truth of their components. That is, it computes elementary functions of the boolean values true and false.

These boolean values are specified by the requirements

$$\begin{aligned} \text{true } x \ y &= x \\ \text{false } x \ y &= y . \end{aligned}$$

Thus we can define true to be $K = \Delta\Delta$ above and false to be KI or $\Delta\Delta(\Delta(\Delta\Delta)(\Delta\Delta))$ since

$$KIxxy = Iy = y$$

as required.

The basic boolean functions are defined by

$$\begin{aligned} \text{and } &= d\{K(KI)\} \\ \text{or } &= d\{KK\}I \\ \text{implies } &= d\{KK\} \\ \text{not } &= d\{KK\}(d\{K(KI)\}I) \\ \text{iff } &= \Delta(\Delta I \text{ not})\Delta . \end{aligned}$$

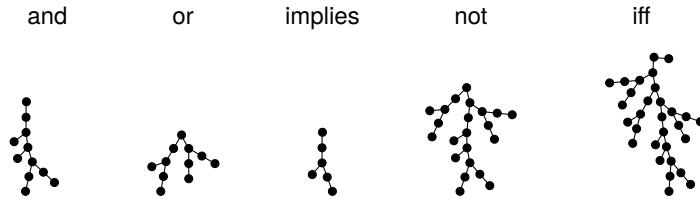


Figure 3.6: Some Boolean Operations

Their graphs are given in Figure 3.6.

Now

$$\text{and } x y = d\{K(KI)\}xy = xy(K(KI)y) = xy(KI) .$$

There are four cases to check for conjunction, according to whether x and y are true or false. The first two are

$$\begin{aligned} \text{and true true} &= \text{true true (KI)} \\ &= K \text{ true (KI)} \\ &= \text{true} \end{aligned}$$

and

$$\begin{aligned} \text{and true false} &= \text{true false (KI)} \\ &= K \text{ false (KI)} \\ &= \text{false} . \end{aligned}$$

The other two cases are left as exercises.

Similarly, we have

$$\begin{aligned} \text{or } x y &= d\{KK\}Ixy \\ &= Ix(KKx)y \\ &= xKy . \end{aligned}$$

If x is true then we get

$$\text{or true } y = KKy = K = \text{true}$$

and if x is false then we get

$$\text{or false } y = KIKy = y$$

as required. The properties of implies, not and iff are left as exercises.

3.6 Pairs

The term $\Delta x y$ can be used to represent the pair of x and y so that we have

$$\begin{aligned}\text{Pair} &= \Delta \\ \text{first}\{p\} &= \Delta p \Delta K \\ \text{second}\{p\} &= \Delta p \Delta (KI)\end{aligned}$$

which implies

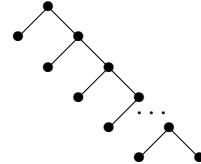
$$\begin{aligned}\text{first}\{\text{Pair } x y\} &= \Delta(\Delta xy) \Delta K = Kxy = x \\ \text{second}\{\text{Pair } x y\} &= \Delta(\Delta xy) \Delta (KI) = (KI)xy = y.\end{aligned}$$

3.7 Natural Numbers

In Section 3.3 the natural numbers were represented by chains of the form $\Delta^n \Delta$. Working with such chains makes constant reference to the rule (S) which is the most complex of the three rules. To avoid this, let us represent the number n by

$$K^n \Delta$$

or



so that positive numbers are always forks. Now zero is given by Δ and the *successor* of n is given by Kn so that we have $\text{successor} = K$. Now the test for being zero is

$$\text{isZero} = d\{K^4 I\} (d\{KK\} \Delta).$$

To see this, observe that

$$\begin{aligned}\text{isZero } n &= d\{K^4 I\} (d\{KK\} \Delta) n \\ &= d\{KK\} \Delta n ((K^4 I) n) \\ &= \Delta n (KKn) (K^3 I) \\ &= \Delta n K (K^3 I).\end{aligned}$$

So we have

$$\begin{aligned}\text{isZero } \Delta &= \Delta \Delta K (K^3 I) \\ &= K = \text{true}\end{aligned}$$



Figure 3.7: The Zero Test and the Predecessor

and

$$\begin{aligned}
 \text{isZero}(Kn) &= \Delta(Kn)K(K^3I) \\
 &= \Delta(\Delta\Delta n)K(K^3I) \\
 &= K^3I\Delta n \\
 &= KI = \text{false}
 \end{aligned}$$

as required. Now the *predecessor* of a positive number $n_1 = \text{successor } n$ is recovered by $n_1\Delta$ since

$$\text{successor } n \Delta = Kn\Delta = n.$$

Unfortunately, this account doesn't work for the predecessor of zero, which should be zero, but $\Delta\Delta$ is not Δ . The solution is to first test n by solving

$$\text{predecessor } n = \Delta n \Delta (KI)$$

to get

$$\text{predecessor} = d\{K^2I\}(d\{K\Delta\}\Delta). \quad (3.2)$$

since

$$\begin{aligned}
 \text{predecessor } n &= d\{K^2I\}(d\{K\Delta\}\Delta)n \\
 &= d\{K\Delta\}\Delta n (K^2In) \\
 &= \Delta n(K\Delta n)(KI) \\
 &= \Delta n\Delta(KI)
 \end{aligned}$$

as required. The graphs of the zero test and the predecessor are given in Figure 3.7.

Further arithmetic functions, such as addition and subtraction, will be introduced in Chapter 4 after more machinery has been produced.

3.8 Fundamental Queries

It is not quite trivial to distinguish the leaves, stems and forks, since a test ΔxMN of the structure of the argument x has only two terms M and N to separate the three possibilities. The fundamental queries provides the machinery to do this. Here are the details.

A test for being a leaf must map leaves to K and stems and forks to KI . There are two other possibilities for the other two tests. To solve all three problems together, let

`is0` be the result of testing a leaf, and `is1` and `is2` be the results of testing stems and forks, respectively. Now we must solve three equations:

$$\begin{aligned}\text{query}\{\text{is0}, \text{is1}, \text{is2}\} \Delta &= \text{is0} \\ \text{query}\{\text{is0}, \text{is1}, \text{is2}\}(\Delta x) &= \text{is1} \\ \text{query}\{\text{is0}, \text{is1}, \text{is2}\}(\Delta xy) &= \text{is2}.\end{aligned}$$

Since the goal is to test x , we look for a solution of the form $\Delta x f g$ or $\Delta x f g h$ or $\Delta f g h k$ or \dots . A first attempt might be something like $\Delta x \text{ is0 } \text{ is1 } \text{ is2}$ but I can't solve the problem using just three arguments, so consider solving by $\Delta x f g h k$. The three possibilities for x yield

$$\begin{aligned}f h k &= \text{is0} \\ f g _ h k &= \text{is1} \\ g _ _ h k &= \text{is2}\end{aligned}$$

where the wildcards $_$ represent arguments over which we have no control, or information. The naive solutions for f and g are $K^2(\text{is0})$ and $K^4(\text{is2})$. This leaves the middle equation, which simplifies to

$$\text{is0 } h \ k = \text{is1}.$$

If looking for a leaf then we must solve $K h k = K I$ so that h is $K I$. The rest are solved by taking k to be `is1`. So we have

$$\text{query}\{\text{is0}, \text{is1}, \text{is2}\} x = \Delta x (K^2 \text{is0})(K^4 \text{is2})(K I) \text{is1}$$

Finally, an equation of the form

$$qx = \Delta x f g h k$$

can be solved by

$$\begin{aligned}qx &= \Delta x f g h k \\ &= D(Kf)\Delta x g h k \\ &= D(Kg)(D(Kf)\Delta)x h k \\ &= D(Kh)(D(Kg)(D(Kf)\Delta))x k \\ &= D(Kk)(D(Kh)(D(Kg)(D(Kf)\Delta)))x\end{aligned}$$

so, after simplifying, we have

$$\text{query}\{\text{is0}, \text{is1}, \text{is2}\} = d\{K \text{ is1}\}(d\{K^2 I\}(d\{K^5 \text{is2}\}(d\{K^3 \text{is0}\}\Delta)))$$

which is represented in Figure 3.8.

Hence, we can define the *fundamental queries*

$$\begin{aligned}\text{isLeaf} &= \text{query}\{K, K I, K I\} \\ \text{isStem} &= \text{query}\{K I, K, K I\} \\ \text{isFork} &= \text{query}\{K I, K I, K\}\end{aligned}$$

which are tests for being a leaf, a stem or a fork, respectively. They are represented in Figure 3.9.

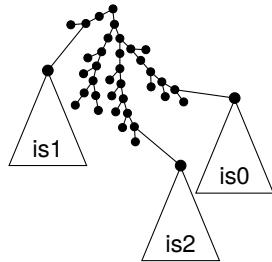


Figure 3.8: Parametrising the Fundamental Queries

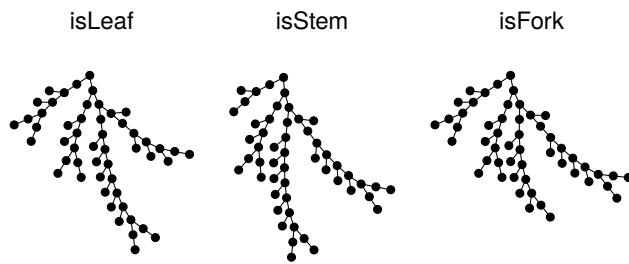


Figure 3.9: The Fundamental Queries

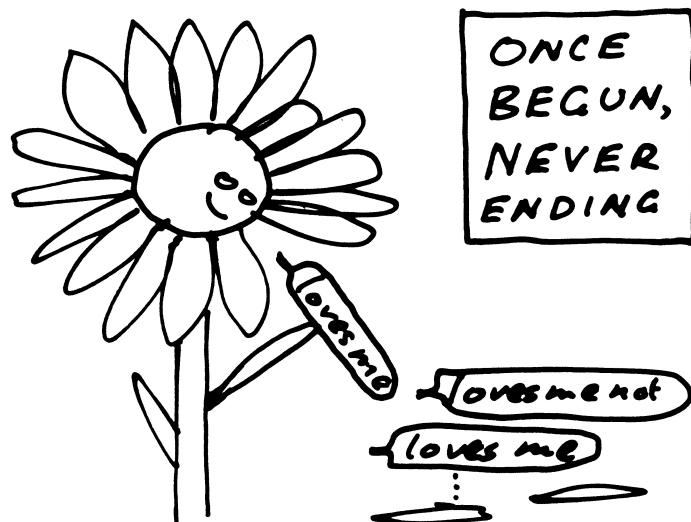
Exercises

1. Define D in terms of K and S just as S can be defined in terms of K and D in (3.1).
2. Check the evaluation of $\text{and } x \text{ } y$ where x is false.
3. Check the properties of the boolean operators or, implies, not and iff.
4. Find trees fst and snd that represent the projections from pairs in general, instead requiring a specific pair. That is, solve $\text{fst } p = \text{first}\{p\}$ and $\text{snd } p = \text{second}\{p\}$. Draw their graphs.
5. Complete the proof that the definition of the predecessor in (3.2) meets its specification.
6. (Hard) Rebuild the arithmetic of this chapter where the number n is represented by $\Delta^n \Delta$ instead of $K^n \Delta$.
7. Write out the binary tree for isLeaf . Use it to evaluate $\text{isLeaf } \Delta$ and $\text{isLeaf } K \Delta$ and $\text{isLeaf } (K \Delta)$ from first principals.
8. (Hard) The *depth* of a binary tree is the length of the longest chain within it that starts at the root. How many binary trees are there with depth at most three, or

four, or five? How many binary trees are there with depth exactly three, or four, or five? Can you give a general formula for these?

Chapter 4

Extensional Programs



Abstract

This chapter develops some extensional functions, i.e. that do not query the internal structure of their arguments. Often these will be defined using variable binding, on the understanding that it is notation only. As well as the usual fixpoint functions, there is *waiting* where a function application waits for a second argument for evaluation begins. In turn, this can be used to define fixpoint functions that are programs, as they must wait for another argument before evaluation can begin.

4.1 Combinators, Combinations and Terms

The *extensional* or *functional behaviour* of a program is the behaviour which is revealed by applying the program to various arguments. In particular, two programs which have the same input-output relation are extensionally equivalent. Haskell Curry characterised this by identifying the extensional functions with the combinators. A *combinator M* is a function of the form

$$Mx_1x_2 \dots x_n = t \quad (4.1)$$

where the x_i are *variables* and t is some *combination* of the x_i built by application. That is, M is a function which, when applied to some combinators N_1, N_2, \dots, N_n produces the result of substituting N_i for x_i in t . More generally, we can denote the substitution of N for x in t by $\{N/x\}t$ so that

$$MN_1N_2 \dots N_n = \{N_1/x_1\}\{N_2/x_2\} \dots \{N_n/x_n\}t$$

or, removing superfluous brackets, that

$$MN_1N_2 \dots N_n = \{N_1/x_1, N_2/x_2, \dots, N_n/x_n\}t.$$

For examples of combinators, we have

$$\begin{aligned} Ix &= x \\ Kxy &= x \\ Sxyz &= xz(yz) \\ Bfxy &= fyx \\ Cgfx &= g(fx) \\ Dxyz &= yz(xz). \end{aligned} \quad (4.2)$$

Thus KMN is equal to the result of substituting M for x and N for y in x , i.e. M .

In turn, the combinators form a calculus that has no need of variables, as they can be applied to each other. For example, if M and N are combinators then

$$(KI)MN = (KIM)N = IN = N$$

shows that KI returns its second argument, just as K returns its first argument. In this manner, the combinators provide good support for computation. Actually, the first combinators were introduced by Moses Schönfinkel as a functional calculus ("funktion calcul") with the goal of eliminating variables and their binding from the logic of predicate calculus. From this point of view, Curry's definition of combinators is not quite satisfactory, as it references the variables that we are trying to eliminate. So let us start with the operators.

In developing a calculus of combinators, we have various choices for the primitive operators, as the collection S, K, I, B, C or S, K, I or S, K . Further, we have the primitive operator Δ of tree calculus. So let us develop a general account based on combinations of some given class O of operators. That is, the O -combinations are given by the Backus-Naur Form (or BNF, see Chapter 2.2)

$$M, N ::= O \mid MN$$

where MN is the *application* of M to N and application is left-associative. Usually, the operators come with some evaluation rules. Where the choice of rules is given, the collection of combinations and evaluation rules is called O -calculus. For example, we may call tree calculus Δ -calculus, on the understanding that the evaluation rules are (K) , (S) and (F) .

Although the goal of combinatory logic was to eliminate the variables, it is difficult, if not impossible to follow through, because variables M, N, P, Q are required to describe all of the evaluation rules considered so far. Instead, we must be content to eliminate variable binding, or at least to give an account of it in terms of more fundamental operations. Thus, to present the evaluation rules requires variables x, y, z, \dots . We will not here concern ourselves with the exact nature of the variables. They could be given by strings of characters, by numbers or something else; it is enough that they be plentiful and that we are able to decide if two variables are the same or not. Then the O -terms are described by the BNF

$$t, u ::= v \mid O \mid tu$$

where v here stands for an arbitrary variable. The *closed terms* are the subset of the terms corresponding to the O -combinations.

For example, the combinations of variables t used to define the combinators in Equation (4.1) are O -terms where O is the empty collection of operators.

The *substitution* $\{u/x\}t$ of a term u for a variable x in a term t is defined by

$$\begin{aligned}\{u/x\}x &= u \\ \{u/x\}y &= y \quad (y \neq x) \\ \{u/x\}O &= O \\ \{u/x\}(st) &= \{u/x\}s \{u/x\}t.\end{aligned}$$

4.2 Variable Binding

In tree calculus, every combinator M as in Equation (4.1) can be represented by a combination denoted by

$$M = [x_1][x_2] \dots [x_n]t$$

where the *bracket abstraction* $[x]t$ represents the function that, when applied to x yields t . It is defined as follows:

$$\begin{aligned}[x]x &= I \\ [x]y &= Ky \quad (y \neq x) \\ [x]O &= KO \\ [x]uv &= d\{[x]v\}([x]u).\end{aligned}$$

Theorem 8 (bracket_beta) *For all variables x and terms t and u we have*

$$([x]t)u = \{u/x\}t. \tag{\beta}$$

Hence the combinator M in Equation (4.1) is represented by $[x_1][x_2] \dots [x_n]t$.

Proof. The proof of the β -rule is by induction on the structure of t . There are three cases, according to whether t is a variable, operator or application. The proof of the second statement is by induction on the number n of variables. Details are left as exercises, or can be found in the Coq proofs. \square

Since all combinators can now be represented in tree calculus, we say that tree calculus is *combinatorially complete*. Note that this does not imply that Δ can be represented as a combinator, as will be proved in Chapter 8.

A curious property of bracket abstractions is that they are *always* stable, in the sense that none of the evaluation rules can be applied from left to right. For example, although Kxx reduces to x , we have

$$[x]Kxx = d\{[x]x\}([x]Kx) = d\{I\}(\Delta(\Delta I)(KK))$$

which does *not* reduce to $[x]x = I$. That is, bracket abstraction does *not* preserve the equality induced by the evaluation rules. In other words, it does not preserve the meaning of terms.

In this sense, bracket abstraction is not a “first-class” means of constructing terms, as these should preserve the equality based on the evaluation rules. This has long been seen as a weakness of combinatory logic. Could we add evaluation rules so that, for example, $[x]Kxx$ reduces to $[x]x$? Could we derive

$$d\{I\}(d\{I\}(KK)) = I ?$$

Despite various attempts, no set of additional rules has been found that will support arbitrary evaluation under bracket abstraction. This abstraction problem will be solved in VA-calculus in Chapter 9. For now, this “bug” is better regarded as a feature since the stability of bracket abstraction (and star abstraction below) provides fine control over reduction, as we will see in Section 4.5.

Unfortunately, the abstraction $[x]t$ is usually larger than necessary. For example, if M is a combinator then KM satisfies the β -rule just as well as $[x]M$ since KMu reduces to M in one step. However, $[x]M$ is much bigger than M . If M is a combinator of size k (e.g built from k copies of Δ) then $[x]M$ has size $5k - 2$. For example, if M is I then we have

$$[x]I = [x](\Delta(\Delta\Delta)(\Delta\Delta))$$

which has size twenty-three instead of KI whose size is seven. Similarly, $[x](Mx) = d\{I\}([x]M)$ is bigger than M even though M also satisfies the β -rule. This five-fold expansion is especially problematic when abstractions are nested.

To solve this, it is necessary to keep track of variable occurrences in terms. Recall that we were trying to avoid variables, but this is easier said than done! Define $x \in t$ by induction on the structure of t as follows:

$$\begin{aligned} x \in y &= \text{ if } x \text{ and } y \text{ are the same variable} \\ x \in \Delta &= \text{ false} \\ x \in tu &= x \in t \text{ or } x \in u . \end{aligned}$$

The solution is to optimise the bracket abstraction to exploit variable occurrence, by defining *star abstraction* as follows

$$\begin{aligned}\lambda^* x.t &= Kt & (x \notin t) \\ \lambda^* x.tx &= t & (x \notin t) \\ \lambda^* x.x &= I \\ \lambda^* x.tu &= d\{\lambda^* x.u\}(\lambda^* x.t) & (\text{otherwise}).\end{aligned}$$

The reference to λ , the Greek letter lambda, comes from λ -calculus (see Chapter 9). If you are already familiar with λ -calculus then this notation will be suggestive but if not then it is safe to treat star abstraction as an independent concept.

Theorem 9 (star_beta) *For all variables x and terms t and u we have*

$$(\lambda^* x.t)u = \{u/x\}t . \quad (\beta)$$

Proof. If t is of the form s or sx where $x \notin s$ then the result is immediate. Otherwise, the proof is as for bracket abstraction. \square

Note that star abstraction still breaks some evaluations (since $\lambda^* x.Kxx = d\{I\}K$ which is not I). Unlike bracket abstractions, star abstractions may require evaluation, since any combinator M is also $\lambda^* x.Mx$.

4.3 Fixpoints

One of the fundamental concepts in computation is *iteration*, in which a function f is applied a given number of times. For example, to iterate f three times over an argument x is to produce $f(f(fx))$. More generally, given a number n then we have

$$f^n(x) = f(f(\dots(fx)\dots)) \quad (\text{where } f \text{ is applied } n \text{ times.})$$

Iteration is good enough when the number of iterations is known in advance, but there are many algorithms for searching or approximation where there is no prior bound on how long the search will take. These functions require the more powerful use of *recursion* which can be represented in *fixpoint functions*. This can be achieved by a combinator Y such that, for any f , we have

$$Yf = f(Yf) .$$

For example, if f is KI then

$$Y(KI)x = KI(Yf)x = x$$

so that $Y(KI)$ is an identity function. On the other hand, if f is I then we get

$$YIx = I(YI)x = YIx$$

which does *not* have a value. Some searches never find a solution: some approximations never converge: they go on forever. Here is a concrete example of a combinator whose evaluation never ends. For any x we have

$$(\lambda^* x. xx)(\lambda^* x. xx) = \{\lambda^* x. xx/x\}(xx) = (\lambda^* x. xx)(\lambda^* x. xx).$$

Expressed using operators, this abstraction becomes

$$DII = \Delta(\Delta(\Delta(\Delta\Delta)(\Delta\Delta)))(\Delta(\Delta\Delta)(\Delta\Delta)).$$

and we have

$$(DII)(DII) = I(DII)(I(DII)) = (DII)(DII).$$

Since any evaluation of it can be further evaluated to recover the original itself, $(DII)(DII)$ can never be stabilised.

A variant of this can be used to define Y . Let

$$\omega = \lambda^* z. \lambda^* f. f(zzf)$$

and define $Y = \omega\omega$. Then we have

$$\begin{aligned} Yf &= \omega\omega f \\ &= (\lambda^* z. \lambda^* f. f(zzf))\omega f \\ &= \{\omega/z\}(\lambda^* f. f(zzf))f \\ &= (\lambda^* f. \{\omega/z\}(f(zzf)))f \\ &= (\lambda^* f. f(\omega\omega f))f \\ &= \{f/f\}(f(\omega\omega f)) \\ &= f(\omega\omega f) \\ &= f(Yf) \end{aligned}$$

as required. Note how the substitution of ω for z has been passed under the star abstraction with respect to f . On this occasion this is safe to do, but perhaps you would like to check by eliminating the star abstraction.

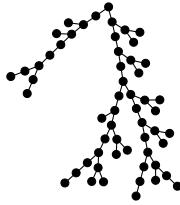
4.4 Waiting

The account of fixpoints above, where Yf evaluates to $f(Yf)$, means that there are always evaluation sequences that do not terminate, such as

$$Yf = f(Yf) = f(f(Yf)) = f(f(f(Yf))) = \dots$$

where the next step is always to evaluate Yf to $f(Yf)$. Using this approach, our recursions are never stable, which makes reflective programming difficult, if not impossible. However, there is a better way.

Since the goal is to produce recursive *functions*, there is no prior requirement to evaluate Yf ; it is enough to evaluate Yfx for any function argument x . All that is

Figure 4.1: The Program W

required is a mechanism to delay the application of Y to f until an argument x is given. The application of Y to f should *wait*.

Consider the combinations $W_0 = \lambda^*x.\lambda^*y.\lambda^*z.xyz$ and

$$W = \lambda^*x.\lambda^*y.[z]xyz$$

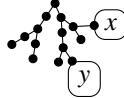
(see Figure 4.1). Both have the property that application to combinations M, N and P yields MNP but they differ when applied to M and N only. Since W_0 is actually I we have $W_0MN = IMN = MN$ in which MN can be reduced but

$$WMN = \{M/x, N/y\}[z]xyz = d\{I\}(d\{KN\}(KM))$$

in which M and N cannot yet interact. WMN must *wait* for another argument P before applying M to N . Note that, in general, WMN is *not* the same as $[z]MNz$ since the latter is always a program, even if M is not. Since it will be used often, let us define

$$\text{wait}\{x, y\} = d\{I\}(d\{Ky\}(Kx)) .$$

It is pictured by



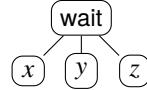
More abstractly, it can be represented by using three labelled triangles for wait and x and y . Even more convenient is to use rounded rectangles instead of triangles, as in



This is a much cleaner approach, which will be iterated to construct much larger programs. However, it does require implicit knowledge that *wait* takes two parameters. By contrast, the diagram



represents Kxy since K takes no parameters. Indeed, these two interpretations can be combined in the diagram



which denotes the application $\text{wait}\{x, y\}z$.

Here are some related forms of waiting. The value of Wx given by $\lambda^*y.\text{wait}\{x, y\}$ on the assumption that the variable y does not occur in x . Calculating the star abstraction allows this assumption to be eliminated to produce

$$\text{wait1}\{x\} = d\{d\{K(Kx)\}(d\{d\{K\}(K\Delta)\}(K\Delta))(K(d\{\Delta KK\}))\}.$$

Waiting for two more arguments is given by analysis of $\lambda^*z.[w].xyzw$ where w, x, y and z are distinct variables. This yields

$$\text{wait2}\{x, y\} = d\{d\{K(d\{Ky\}(Kx))\} d\{d\{K\}(K\Delta)\}(K\Delta))(K(d\{I\}))\}.$$

Then

$$\text{wait21}\{x\} = \lambda^*y.\text{wait2}\{x, y\} \quad (\text{y not in } x)$$

can be similarly expanded to avoid the side-condition.

Again, we can define $\text{wait3}\{x, y\}$ by analysis of $\lambda^*z.\lambda^*t.[w]xyztw$ and then define $\text{wait31}\{x\}$ by abstracting with respect to y .

4.5 Fixpoint Functions

The instability of the usual fixpoints is structural, in that the Y combinator takes only one argument. The first step in producing stable fixpoints is to make the the combinator wait for a second argument, by solving

$$Y_2fx = fx(Y_2f)$$

in a way that makes Y_2f stable if f is.

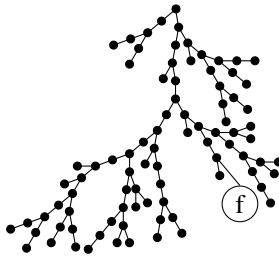
One solution begins by replacing $\omega\omega$ in Y by $\text{wait}\{\omega, \omega\}$ which is now a normal form. However, there are still two difficulties: when applied to some program f it becomes $\omega\omega f$ which does not have a normal form; it evaluates to $f(\omega\omega f)$ which is *not* $\text{wait}\{\omega, \omega\}f$. Continuing on this path will work, but there is a lot of waiting, and a lot of duplication. Instead, let us try a more direct approach.

To reduce duplication of combinations in the program, introduce

$$\text{self_apply} = \lambda^*w.ww = d\{I\}I$$

and replace $\omega\omega$ with **self_apply** ω . Then define

$$Z\{f\} = \text{wait}\{\text{self_apply}, d\{\text{wait1}\{\text{self_apply}\}(Kf)\}\}.$$

Figure 4.2: The Program $Y_2\{f\}$

This is almost the solution we seek since we have

$$\begin{aligned}
 Z\{f\}x &= \text{self_apply } d\{\text{wait1}\{\text{self_apply}\}(Kf)\}x \\
 &= d\{\text{wait1}\{\text{self_apply}\}(Kf)\}d\{\text{wait1}\{\text{self_apply}\}(Kf)\}x \\
 &= Kf(d\{\text{wait1}\{\text{self_apply}\}(Kf)\}) \\
 &\quad (\text{wait1}\{\text{self_apply}\}(d\{\text{wait1}\{\text{self_apply}\}(Kf)\}))x \\
 &= fZ\{f\}x.
 \end{aligned}$$

It remains to swap the order of arguments of f by defining

$$\text{swap}\{f\} = d\{Kf\}d\{d\{K\}(K\Delta)\}(K\Delta)$$

so that $\text{swap}\{f\}xy = fyx$. Now define

$$Y_2\{f\} = Z\{\text{swap}\{f\}\}.$$

Theorem 10 (fixpoint function) *For any function f the program $Y_2\{f\}$ is a fixpoint function for f that satisfies*

$$Y_2fx = fx(Y_2f)$$

for any x .

Two subtleties lie behind these calculations. First, even when $Y_2\{f\}$ is a program, this does *not* imply that $Y_2\{f\}x$ evaluates to a program. For example, $Y_2\{I\}$ is a program but $Y_2\{I\}x$ is not as it reduces to itself.

Second, it is tempting to replace $Y_2\{f\}$ by the application to f of the abstraction $\lambda^*f.Y_2\{f\}$. However, $(\lambda^*f.Y_2\{f\})f$ does not take the form of a program until simplified to $Y_2\{f\}$. Further, the proof of its properties becomes more complicated, since left-to-right application of the rules implies

$$(\lambda^*f.Y_2\{f\})fx = fx(Y_2\{f\})$$

but does not imply $(\lambda^*f.Y_2\{f\})fx = fx((\lambda^*f.Y_2\{f\})f)$. Rather, the fixpoint property requires a second left-to-right application to show $f((\lambda^*f.Y_2\{f\})f)fx = f(Y_2\{f\})x$. This will become more troublesome in Part II when the rules will acquire an orientation. So we will stick with the chosen definition.

The same techniques can be used to define fixpoints for functions for any number of arguments but in practice, Y_2 is enough to ensure stability.

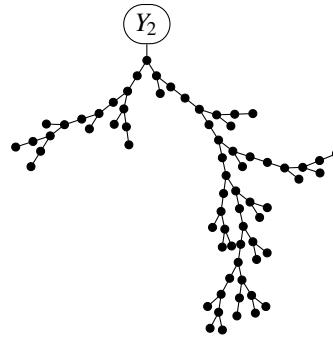


Figure 4.3: The Program plus

4.6 Arithmetic

As a simple application of fixpoints, we have addition of natural numbers given by

$$\text{plus} = Y_2\{\lambda^* m. \lambda^* p. \Delta m I(K(\lambda^* x. \lambda^* n. (K(pxn))))\}$$

which is presented in Figure 4.3.

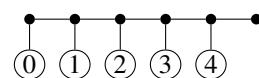
Similarly, subtraction can be defined by building on the predecessor instead of the successor. In turn, these can be used to define multiplication and division, exponentiation, etc.

The other main arithmetic task is to solve equations. In general, one can envisage all sorts of techniques for this but the brute force approach is to check each number in turn, starting with zero, to see if it is a solution. This process of *minimisation* can also be described using fixpoints, by fixing the function which maps n to n if it is a solution, and to the successor of n otherwise.

In this manner, Kleene specified the class of μ -recursive arithmetic functions, which arise in all the traditional models of computation. Indeed, they are all definable in tree calculus too. Since the details do not contribute much to the understanding of reflective programs, they are not given here in the text, but can be found in the formal development in Coq.

4.7 Lists and Strings

Lists can be represented by using a leaf to represent the empty or nil list, and a fork to combine the head and tail of a list. For example, the list with entries 0, 1, 2, 3, 4 is given by



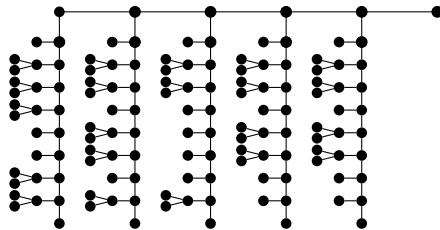
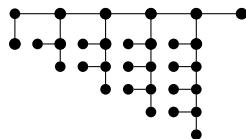
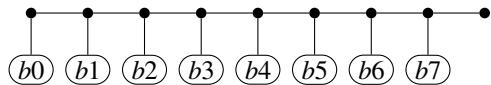


Figure 4.4: The String “small”

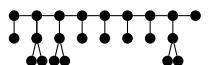
which expands to



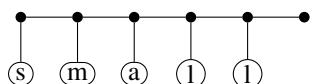
The *bits* 0 and 1 are represented by the natural numbers Δ and $K\Delta$ as before. The *bytes* are given by eight-tuples of bits $(b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7)$ which can be represented by



Hence, the ASCII characters can be represented by their bytecodes. For example, the character 'a' is given by the byte 01100001 which is given by



Then *strings* are given by lists of characters. For example, the string “small” is given by



To represent “small” as a winter tree without any labels, let us rotate the tree for each character clockwise by a quarter circle, to get the picture in Figure 4.4

4.8 Mapping and Folding

To *map* a function f over a list is to apply the function to each entry of the list. This time, the recursion is driven by the third argument of the fixpoint function. The direct solution would be to develop a function Y_3 for this. Although it can be done, the machinery becomes progressively heavier as more arguments are added. Fortunately,

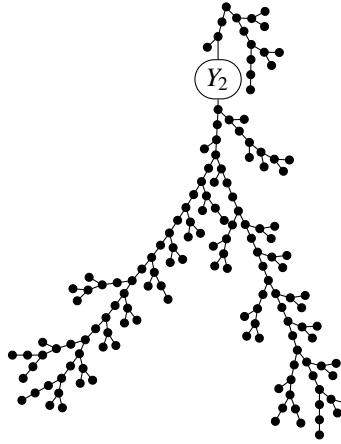


Figure 4.5: Mapping over Lists

Y_2 is enough, provided that the order of arguments is adjusted, so that the list comes first. This is defined by

$$\begin{aligned} \text{list_map_swap} &= \lambda^*x.\Delta x(K(K\text{ t_nil}))(\lambda^*h.\lambda^*t.\lambda^*m.\lambda^*f.\text{t_cons}(fh)(mf)) \\ \text{list_map} &= \text{swap}\{Y_2\{\text{list_map_swap}\}\} \end{aligned}$$

whose graph is in Figure 4.5. It follows that we have

$$\begin{aligned} \text{list_map } f \ x &= Y_2\{\text{list_map_swap}\} \ x \ f \\ &= \text{list_map_swap } x \ Y_2\{\text{list_map_swap}\} \ f \\ \text{list_map } f \ \text{t_nil} &= \text{t_nil} \\ \text{list_map } f \ (\text{t_cons } h \ t) &= \text{t_cons}(fh)(\text{list_map } f \ t) \end{aligned}$$

as required.

Similarly, the *left folding* of a function over a list is specified by

$$\begin{aligned} \text{list_foldleft } f \ x \ \text{t_nil} &= x \\ \text{list_foldleft } f \ x \ (\text{t_cons } h \ t) &= \text{list_foldleft } f \ (fxh) \ t . \end{aligned}$$

Now the list parameter comes fourth, so is handled by

$$\begin{aligned} \text{list_foldleft_aux} &= \lambda y.(\Delta y(K(KI))(\lambda h.\lambda t(\lambda r.\lambda f.\lambda x.rtf(fhx)))) \\ \text{list_foldleft} &= \lambda f.\lambda x.\lambda y.Y_2\{\text{list_foldleft_aux}\} \ y \ f \ x . \end{aligned}$$

Again, *right folding* of a function over a list is specified by

$$\begin{aligned} \text{list_foldright } f \ x \ \text{t_nil} &= x \\ \text{list_foldright } f \ x \ (\text{t_cons } h \ t) &= f \ h \ (\text{list_foldright } f \ x \ t) \end{aligned}$$

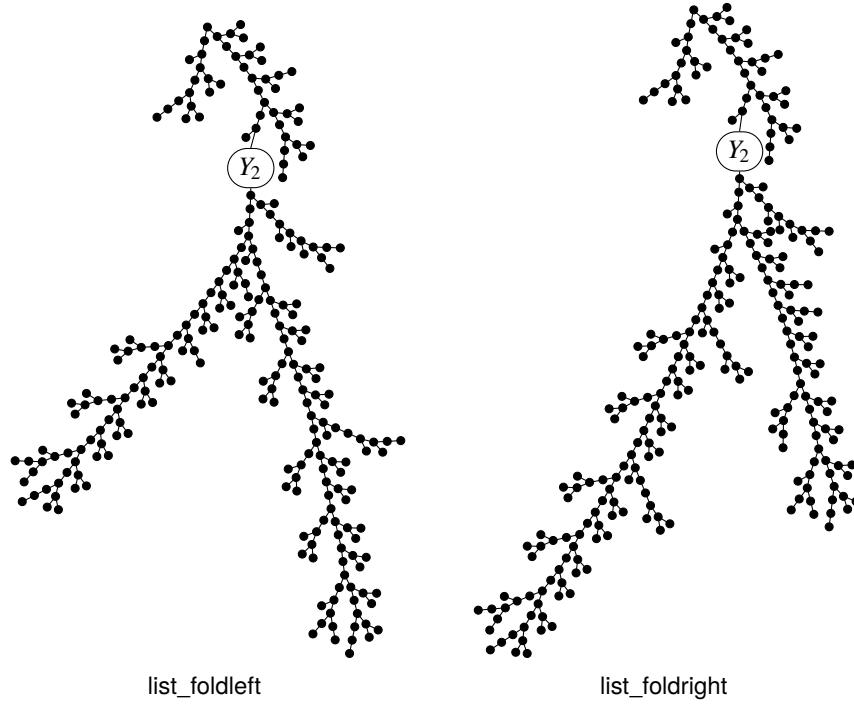


Figure 4.6: Folding over Lists

and realised in Figure 4.6 by

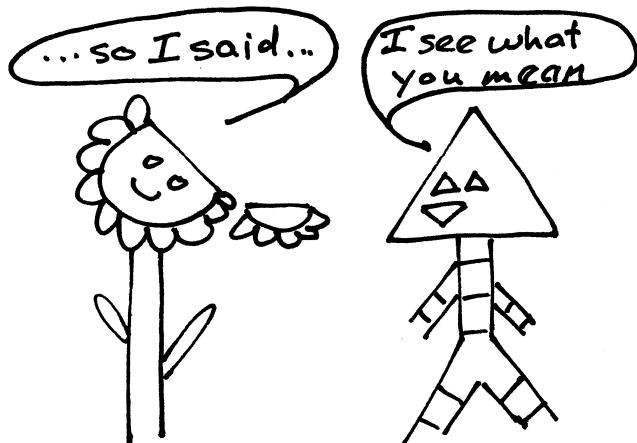
$$\text{list_foldright} = \lambda f. \lambda x. \lambda y. Y_2 \{ \lambda y. \lambda h. \lambda t. \lambda r. \lambda f. \lambda x. f h(r t f x) \},$$

Exercises

1. Show that if M is a combination of size k then $[x]M$ has size $5k - 2$.
2. Show that $\{\omega/z\}(\lambda^*f.f(z z f)) = \lambda^*f.f(\omega\omega f)$.
3. Define $\omega' = [x][f]f(x x f)$. Show that $\omega'xf$ has the same functional behaviour as ω .
4. Define multiplication of numerals and confirm its properties.
5. Define subtraction of numerals.
6. Using the Coq implementation of μ -recursion, show that the fixpoint of the successor function is a program (i.e. has a normal form) even though it fails to halt on every input.

Chapter 5

Intensional Programs



Abstract

Intensional programs query the internal structure of their arguments. Simple examples include programs for the size or equality of binary trees. More general is the ability to tag functions with additional information, such as comments or types.

5.1 Intensionality

Intensional programs query the internal structure of their arguments. Simple examples include programs for the size or equality of binary trees. More general is the ability to tag functions with additional information, such as comments or types.

The *intensional behaviour* of a program p is all behaviour which is revealed by making it the argument of some query, without trying to apply p to arguments. For example, any comments within a program reflect the intentions of the programmer. By design, they cannot impact on the program's functional behaviour but they may be revealed by program analysis. Similarly, the two programs which map a number x to either $0 + x$ or $x + 0$ have the same functional behaviour as the identity function, but may be distinguished if the programming language is sufficiently intensional.

Usually, calculi that allow functions to be applied to functions do not support intensional programming. For example, we will see in Chapter 8 that combinatory logic does not support comment extraction. By contrast, tree calculus supports intensional programming, in the sense that its programs can be tagged with comments which can be recovered during computation. These tags can be used to distinguish terms which have different meanings even though they have the same behaviour as functions. In particular, tags may include *type information* which can be used to check for type errors.

This chapter will begin with a couple of simple examples of intensional computation, of the size of a program and of the equality of programs. Then we will introduce tagging, and use it to define a simply-typed language of booleans and natural numbers.

5.2 Size

The size of a program p is computed easily enough, given the machinery already developed. If it is a leaf then its size is 1. If it is a stem Δx then its size is one more than that of x . If it is a fork Δxy then its size is one more than the sum of the sizes of x and y . Since the function `size` is a recursive function of one argument, it is given by an application of Y_2 . The simplest way to separate the leaves, stems and forks is to first test the argument x for being a stem. If so, then take the successor of the size of its branch. Otherwise, x must be a kernel or fork, whose cases can be separated by an application of Δx to suitable arguments. Hence, we have

$$\begin{aligned} \text{size} &= Y_2(\lambda^*x.\text{isStem } x \\ &\quad (\lambda^*s.\Delta(x\Delta)\Delta(\lambda^*x_1.K(K(s\ x_1)))) \\ &\quad (\lambda^*s.\Delta\ x\ (K\Delta)\ (\lambda^*x_1.(\lambda^*x_2.K(\text{plus}\ (s\ x_1)\ (s\ x_2))))))) . \end{aligned}$$

Although this works well enough, the use of `plus` introduces a second recursion, which is expensive. The representation of `size` is given in Figure 5.1.

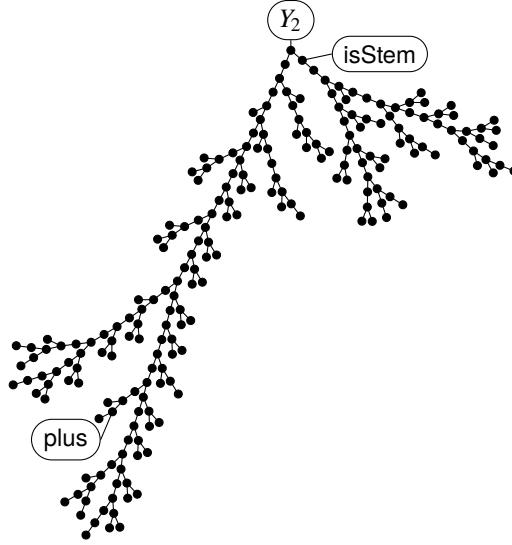


Figure 5.1: The Programm size

5.3 Equality

Equality of programs is defined in the same manner, by

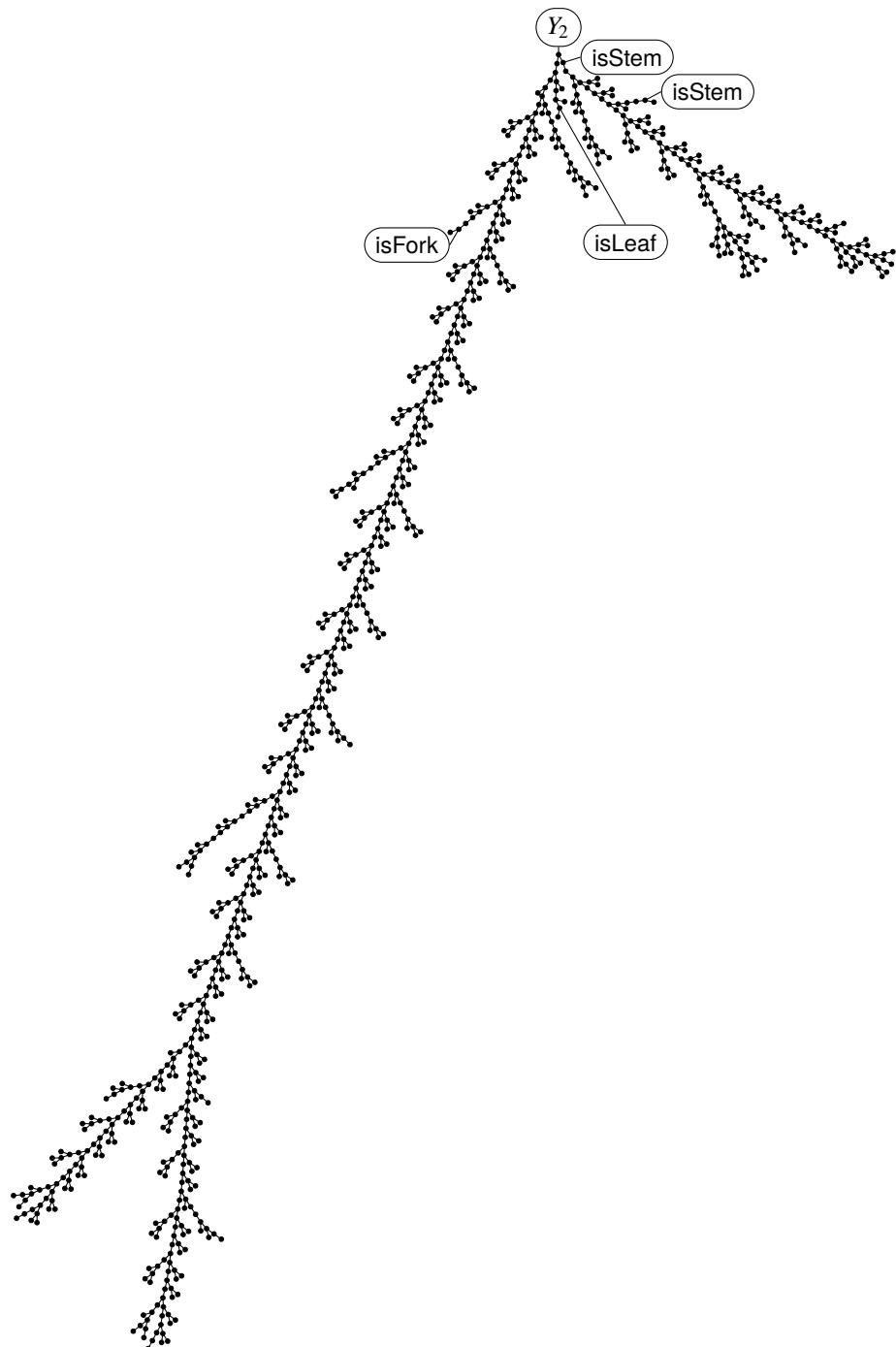
$$\begin{aligned}
 \text{equal} &= Y_2(\lambda^* x. \text{isStem } x \\
 &\quad (\lambda^* e. \lambda^* y. \text{isStem } y \\
 &\quad (e(x\Delta) (y\Delta)) \\
 &\quad (KI)) \\
 &\quad (\Delta x \\
 &\quad (\lambda^* e. \lambda^* y. \text{isLeaf } y) \\
 &\quad (\lambda^* e. \lambda^* y. \lambda^* x_1. \lambda^* x_2. \\
 &\quad \text{isFork } y \\
 &\quad (\Delta y \Delta (\lambda^* y_1. \lambda^* y_2. ex_1 y_1(ex_2 y_2)(KI))) \\
 &\quad (KI)))) .
 \end{aligned}$$

Theorem 11 (equal_programs) *For all programs M, equal M M = K*

Proof. The proof is by induction on the structure of M. \square

Theorem 12 (unequal_programs) *If M ≠ N then equal M N = KI for all programs M and N.*

Proof. The proof is by induction on the structure of M. \square

Figure 5.2: The Program `equal`

5.4 Tagging

A calculus supports *tagging* if there are term forms *tag* and *getTag* such that

$$\begin{aligned}\text{tag}\{t, f\} x &= f x \\ \text{getTag}(\text{tag}\{t, f\}) &= t.\end{aligned}$$

for all terms t and f . Thus, even though the tag t can be recovered by applying *getTag*, the tag has no effect on the functional behaviour, as the tag is discarded when the function is applied to some argument x . Note that *getTag* is defined to be a combination but *tag* is a binary function of combinations. The reasons for this are practical. In Part II we will define some programs using *tag* and *getTag*. That *getTag* is a combination simplifies the analysis, but if *tag* were defined to be a combination then its applications would not be programs.

Theorem 13 (tree_calculus_supports_tagging) *Tree calculus supports tagging.*

Proof. Define

$$\begin{aligned}\text{tag}\{t, f\} &= d\{t\}(d\{f\}(KK)) \\ \text{getTag} &= \lambda^* p.\text{first}\{\text{first}\{p\}\Delta\}.\end{aligned}$$

Then

$$\begin{aligned}\text{tag}\{t, f\} x &= d\{f\}(KK)x(tx) \\ &= KKx(fx)(tx) \\ &= K(fx)(tx) \\ &= fx.\end{aligned}$$

and

$$\begin{aligned}\text{getTag}(\text{tag}\{t, f\}) &= \text{first}\{\text{first}\{d\{t\}(d\{f\}(KK))\}\Delta\} \\ &= \text{first}\{\Delta t \Delta\} = t\end{aligned}$$

as required. \square

Further, we can recover the term f that has been tagged by

$$\text{untag} = \lambda^* x.\text{fst}\{\text{fst}\{\text{snd}\{x\}\}\Delta\}.$$

Trees for these constructions are in Figure 5.3

Tagging can be used to record the *intentions* of the programmer as executable comments. For example, it can be used to name things, so that the declaration `let small = True` can be replaced by `tag{small, True}`. We have seen how to define strings of characters in Chapter 4.7 so it is now easy to add comments to programs.

It will sometime be convenient to tag fixpoint functions. Such tags cannot be added after the fixpoint has been formed, as they must recur. Instead, define

$$Y_{2t}\{t, f\} = \text{tag}\{\text{wait}\{\text{self_apply}, d\{\text{tag_wait}\{t\}(Kf)\}\}\}$$

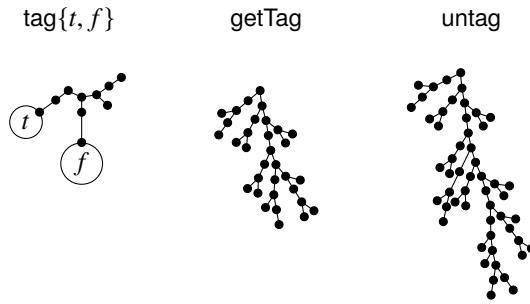


Figure 5.3: Tagging

where $\text{tag_wait}\{t\}$ is a tagged version of $\text{wait1}\{\text{self_apply}\}$ that appeared in the definition of Y_2 that is given by

$$\text{tag_wait}\{t\} = \lambda^* w. \text{tag}\{t, \text{wait}\{\text{self_apply}, w\}\} \quad (w \text{ not in } t).$$

As usual, the side-condition can be eliminated by expanding out the star-abstraction.

Below is another application of tagging.

5.5 Simple Types

Types are used to separate programs into meaningful classes, so that, for example, functions and numbers are kept apart. There is a rich collection of type systems and of typed programming languages. Some languages require every term to have a type, which improves clarity and safety but excludes some interesting programs. In others, types are desirable, but optional. This increases expressive power at the cost of some safety. This section will introduce a simple type system, in the style of William Tait, that supports booleans, natural numbers and functions, and use it to type a few terms of tree calculus by tagging terms with their types. Much remains to be done, even for simple types. With little effort, the type system can be extended to support other type constructions, e.g. for list types. With a greater effort, it should be possible to support more advanced type systems. However, these commonly require the full power of variable binding, which not be addressed until Part II. Furthermore, type systems deserve a comprehensive treatment that goes far beyond the scope of this book.

The *simple types* are given by the BNF

$$T ::= o \mid \iota \mid T \rightarrow T .$$

These are represented by the trees

$$\begin{aligned} [o] &= \triangle \text{ ``Bool''} \\ [\iota] &= \triangle \text{ ``Nat''} \\ [U \rightarrow T] &= \triangle [U] [T] . \end{aligned}$$

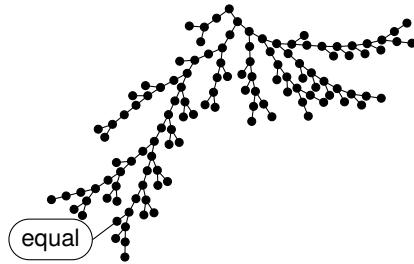


Figure 5.4: The Program type_check

That is, the boolean and number types are represented by stems and function types are represented by forks.

A *typed term* t^T is a term t decorated with a type T . It is represented by

$$t^T = \text{tag}\{T, t\} .$$

In this manner we can provide types for the booleans and some boolean operations, the natural numbers and some arithmetic operations, as follows.

$$\begin{aligned}
 \text{true} &= K^o \\
 \text{false} &= (KI)^o \\
 \text{and} &= (\Delta^2(K^2I))^{o \rightarrow o \rightarrow o} \\
 \text{or} &= (\Delta(\Delta(KK))I)^{o \rightarrow o \rightarrow o} \\
 \text{implies} &= (\Delta^2(KK))^{o \rightarrow o \rightarrow o} \\
 \text{not} &= (\Delta(\Delta(KK))(\Delta(\Delta(K(KI))))I)^{o \rightarrow o} \\
 \text{iff} &= (\Delta(\Delta \text{not}))^{o \rightarrow o \rightarrow o} \\
 \text{zero} &= \Delta^\iota \\
 \text{successor} &= K^{\iota \rightarrow \iota} \\
 \text{isZero} &= (\Delta(\Delta(K^4I))(\Delta(\Delta(KK))\Delta))^{\iota \rightarrow o} \\
 \text{plus} &= (Y_3\{\lambda^* p. \lambda^* x. \lambda^* y. \Delta yx(K(p(Kx)))\})^{\iota \rightarrow \iota \rightarrow \iota} .
 \end{aligned}$$

That done, we must show how to apply one typed term to another. The usual application will not do because the tags get in the way. Rather, we must tag the application of the untyped terms with its inferred type. Of course, if type inference fails then an error results. A type error is represented by $\text{error} = \Delta$.

Type checking of applications is defined by

$$\text{type_check} = \lambda^* x. \lambda^* y. \text{isFork } x (\Delta x \Delta (\lambda^* x_1. \lambda^* x_2. \text{equal } x_1 \ y \ x_2 \ \text{error})) \ \text{error} .$$

Then *typed application* is defined by

$$\text{typed_app} = \lambda^* f. \lambda^* x. \text{tag}\{\text{type_check} (\text{getTag } f) (\text{getTag } x), \text{untag}\{f\} \ \text{untag}\{x\}\}$$

The usual theorem to be proved of typed calculi is that evaluation preserves typing. This seems altogether plausible for our example, but we don't yet have the machinery required for the proof.

5.6 More Queries

The queries `isLeaf` and `isStem` and `isFork` of Chapter 3 produce the boolean values K or KI according to whether the argument is a leaf, stem or fork. However, this is not always the most efficient way of obtaining the information. An alternative is to use a leaf to represent falsehood and any fork to represent truth. Given such a boolean b then $\Delta b s t$ is a form of conditional that applies t to the branches of a fork b and returns s if b is a leaf. Here are some examples. The query

$$\text{isStem2} = \lambda^* z. \Delta z \Delta (K^2 \Delta)$$

maps leaves and forks to a leaf (false) and a stem Δx to the fork $\Delta(K^2 \Delta)(x(K^2 \Delta))$. Of course, this fork is not very useful, since it does not support a uniform means of recovering x .

Similarly, define

$$\text{isFork2} = \lambda^* z. \Delta z (KK)(K(K\Delta)) .$$

It maps forks to Δ and maps leaves and stems to forks.

5.7 Triage

The programs `size` and `equal` both test their arguments, to test if they are leaves, stems or forks. This sort of testing can be packaged up using *triage*:

$$\text{triage_comb} = \lambda^* f_0. \lambda^* f_1. \lambda^* f_2. \lambda^* z. \text{isStem2 } z (\Delta z f_0 f_2) (\Delta(z\Delta) \Delta(\lambda^* x. K(f_1 x))) .$$

If z is a stem Δx then $z\Delta$ is a fork which can be used to apply f_1 to x . Otherwise, if z is a leaf or fork then $\Delta z f_0 f_2$ applies f_0 or f_2 respectively. It follows that we have

$$\begin{aligned} \text{triage}\{f_0, f_1, f_2\} \Delta &= f_0 \\ \text{triage}\{f_0, f_1, f_2\} (\Delta x) &= f_1 x \\ \text{triage}\{f_0, f_1, f_2\} (\Delta xy) &= f_2 xy . \end{aligned}$$

In this manner, we may re-express `size` as

$$\text{size_variant} = Y_2\{\text{triage}\{K(K\Delta), \lambda^* y. \lambda^* s. K(sy), \lambda^* y. \lambda^* z. \lambda^* s. K(\text{plus}(s y) (s z))\}$$

and re-express `equal` as

$$\begin{aligned} \text{equal_variant} &= Y_2\{\text{triage}\{ & \\ & K \text{ triage}\{K, KI, KI\}, \\ & \lambda^* y. \lambda^* e. \text{triage}\{KI, ey, K(K(KI))\}, \\ & \lambda^* y_1. \lambda^* y_2. \lambda^* e. \text{triage}\{KI, K(KI), \lambda^* z_1. \lambda^* z_2. ey_1 z_1(ey_2 z_2)(KI)\} \\ & \} . \end{aligned}$$

Note that `equal_variant` is not the same as `equal` since triage induces a slightly different flow of control during execution.

More sophisticated examples require the recursive use of triage, which can be expressed as pattern-matching.

5.8 Pattern Matching

Pattern-matching functions take the form

$$\begin{array}{l} p_1 \Rightarrow s_1 \\ | \quad p_2 \Rightarrow s_2 \\ \dots \\ | \quad p_k \Rightarrow s_k \\ | \quad r \end{array}$$

where each $p_i \Rightarrow s_i$ is a *case*. An argument u is compared to each pattern in turn until a match is found or all patterns have been attempted. If u matches with p_i then the resulting substitution is applied to s_i . If no match is found then ru results.

Such pattern-matching functions can be expressed as iterated *extensions* of the form

$$p \Rightarrow s \mid r$$

where p is the *pattern*, s is the *body* and r is the *default function*, since r may itself be an extension that captures all the other cases. In turn, the extension could be interpreted as an application of the case $p \Rightarrow s$ to r . However, this would imply that extensions are never programs, so we must use waiting, and define

$$p \Rightarrow s \mid r = \text{wait}\{p \Rightarrow s, r\} .$$

The patterns for programs, or *program patterns* are given by the BNF

$$p, q ::= x \mid \Delta \mid \Delta p \mid \Delta pq .$$

That a variable occurs twice in a pattern is allowed but is not good style since only one occurrence will bind the body. A typical example of a pattern is $Y_2\{x\}$. Now define cases by induction on the structure of the pattern as shown in Figure 5.5.

Note that when the pattern is a variable then K is required to discard the default function. Also, when the pattern is a stem or fork then the default function r is used several times. Since this is commonly larger than the case itself, it is important to avoid its duplication. In turn, this requires auxiliary functions r_1 and r_2 to express the various uses of r when the pattern is a fork. This approach is able to show that, for example, we have $(Y_2\{x\} \Rightarrow x) I (Y_2\{K\}) = K$.

Exercises

1. Compute `size` ($K^n \Delta$).

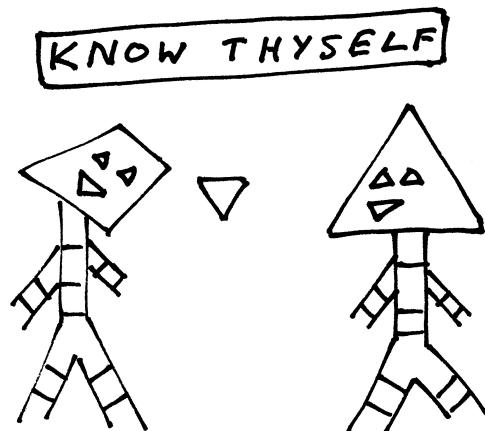
$$\begin{aligned}
x \Rightarrow s &= K(\lambda^* x.s) \\
\Delta \Rightarrow s &= \{s/z\}(\lambda^* r.\lambda^* x.\text{isLeaf } x z (r x)) \\
\Delta p \Rightarrow s &= \{p \Rightarrow s/z\}(\lambda^* r.\lambda^* x.\text{isStem } x (\Delta(x\Delta)\Delta(\lambda^* y.K(f(\lambda^* z.r(\Delta z))y))(r x))) \\
r_1 &= \lambda^* x_1.\lambda^* x_2.\lambda^* r.r(\Delta x_1 x_2) \\
r_2\{p\} &= \{p/z\}(\lambda^* x_2.(\lambda^* r.r(\Delta z x))) \\
\Delta pq \Rightarrow s &= \{p \Rightarrow \text{wait}\{q \Rightarrow Ks, r_2\{p\}\}/z\} \\
&\quad (\lambda^* r.(\lambda^* x.\text{isFork } x (\Delta x\Delta(\lambda^* x_1.\lambda^* x_2.\text{wait}\{z, r_1\}x_1 x_2 r)))) \\
\end{aligned}$$

Figure 5.5: Pattern Matching

2. Compute size W .
3. Compute size size.
4. Compute equal K ($K\Delta$) by applying Theorem 12.
5. Compute equal equal equal by applying Theorem 11.
6. Confirm that tag and getTag meet their specifications.
7. Check the type of the program and true false.
8. Evaluate size_variant K .
9. Evaluate equal_variant K ($K\Delta$).
10. Redefine size as a pattern-matching function and use it to compute the size of K .

Chapter 6

Reflective Programs



Abstract

Reflective programs are programs that can act on themselves to query their own structure. This chapter will consider self-evaluators as examples of reflective programs.

6.1 Reflection

Reflective programs are programs that can act on themselves to query their own structure. The querying is important: that the identity function can be applied to itself does not make it reflective. A simple example is the size function defined in Chapter 5. When applied to itself, the result is (the tree of) the number 508. The canonical examples of reflective programs are self-evaluators, or self-interpreters, which express an evaluation strategy as a program which can then be self-applied. Self-evaluation in a calculus provides good evidence for the ability to perform program analysis and optimisation within the calculus itself. Traditionally, self-interpreters were allowed to act on the syntax tree of a program, i.e. its *quotation*. This is necessary if programs are inherently unstable, but quotation performs all of the work necessary to discover program structure, performs all of the intensional computation. When quotation lies outside of the formal calculus then interpretation is separated from computation proper, so that some sort of staging is required.

6.2 Evaluation Strategies

So far, all of our calculations have been driven by reasonable use of the equations of tree calculus. However, a computing machine needs a strategy to make the rules useful. For example, consider the evaluation of an application tu . Clearly, we need some information from t so suppose that it evaluates to some $\Delta a y$. In turn, we must evaluate a so suppose that it is of the form Δx . Should we evaluate x or y or u next, or should we evaluate $yu(xu)$? Traditionally, this question has been framed in terms of eager evaluation versus lazy evaluation, etc. but the alignment with concepts of tree calculus is not exact so new terminology will be introduced. We will consider four strategies that are definable within tree calculus.

Branch-first evaluation assumes that t and u above are programs and produces the program resulting from their application (if it has one). Each branch that is created is evaluated before evaluating the root. This is conceptually simple, since it is enough to manipulate programs, but will sometimes evaluate branches that do not contribute to the result. In the worst case, the evaluation of the unneeded branch fails to produce a result, so that evaluation fails to find the value.

Root evaluation does just enough work to determine whether y above is a leaf, stem or fork. To facilitate discussion, define the *factorable forms* to be the terms which are leaves, stems or forks. If all nodes in a tree are factorable then it is a program, but this need not be so. So root evaluation produces factorable forms whose branches may contain unevaluated computations. To control the evaluation of these branches, they will be *quoted*, so that computations are represented by values. Thus, root evaluation maps a quotation to a factorable form of quotations.

Root-and-branch evaluation first performs root evaluation and then recursively evaluates the branches of the resulting factorable form. Thus, root-and-branch evaluation maps a quotation to a program.

Of course, the use of quotation introduces meta-theory to the system, since quotation of computations is not definable within tree calculus. However, quotation of programs

$$\begin{array}{c}
 \frac{}{\Delta, z \Rightarrow \Delta z} \quad \frac{}{\Delta y, z \Rightarrow \Delta yz} \quad \frac{}{\Delta\Delta y, z \Rightarrow y} \\
 \frac{y, z \Rightarrow v_1 \quad x, z \Rightarrow v_2 \quad v_1, v_2 \Rightarrow v}{\Delta(\Delta x)y, z \Rightarrow v} \quad \frac{z, w \Rightarrow v_1 \quad v_1, x \Rightarrow v}{\Delta(\Delta wx)y, z \Rightarrow v}
 \end{array}$$

Figure 6.1: Branch-First Evaluation

is definable by the program quote.

Root-first evaluation maps a pair of programs t and u to the program obtained by root-and-branch evaluation of $\Delta(\text{quote } t)$ (quote u). In this manner, no meta-theory is required.

6.3 Branch-First Evaluation

In branch-first evaluation, all branches must be evaluated before trying to evaluate at the root. The values are the programs, here denoted by v, v_1, v_2, \dots . The *evaluation*

$$v_1, v_2 \Rightarrow v$$

takes a pair of programs, the function v_1 and its argument v_2 , and seeks a program v that is the result of the application, by applying the rules in Figure 6.1. Each rule has a conclusion below the line and premises above. Unlike the pictures of trees of tree calculus, the root of the tree is now at the bottom.

For example, evaluation of $I\Delta$ proceeds as follows. First, place the pair to be evaluated below the line

$$\frac{?}{I, \Delta \Rightarrow ?}$$

where the question marks indicate missing parts of the derivation. Since I is $\Delta(\Delta\Delta)(\Delta\Delta)$ we can apply the stem rule to get

$$\frac{\Delta\Delta, \Delta \Rightarrow ? \quad \Delta, \Delta \Rightarrow ? \quad ?, ? \Rightarrow ?}{I, \Delta \Rightarrow ?}.$$

The first two premises follow from the first two rules, to get

$$\frac{\frac{\Delta\Delta, \Delta \Rightarrow \Delta\Delta\Delta}{\Delta, \Delta \Rightarrow \Delta\Delta} \quad \frac{?}{\Delta\Delta\Delta, \Delta\Delta \Rightarrow ?}}{I, \Delta \Rightarrow ?}.$$

Now apply the third to finish off with

$$\frac{\frac{\Delta\Delta, \Delta \Rightarrow \Delta\Delta\Delta}{\Delta, \Delta \Rightarrow \Delta\Delta} \quad \frac{?}{\Delta\Delta\Delta, \Delta\Delta \Rightarrow \Delta}}{I, \Delta \Rightarrow \Delta}.$$

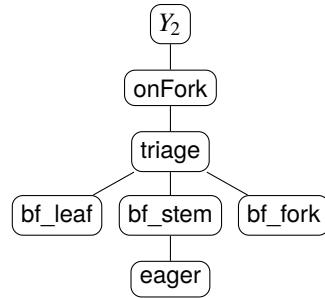


Figure 6.2: A Branch-First Evaluator

Note that for each function there is a unique rule that can be applied, so that the evaluation strategy is *deterministic*.

Theorem 14 (branch_first_eval_program) *If M and N are programs and there is a branch-first evaluation $M, N \Rightarrow P$ then P is a program.*

Proof. The proof is by induction on the structure of the evaluation. \square

6.4 A Branch-First Self-Evaluator

A *self-evaluator* or, if you like, a *self-interpreter* for an evaluation strategy is a program that can represent the strategy. For branch-first evaluation, this would be a program `bf` such that, for all programs M, N and P we have

$$M, N \Rightarrow P \quad \text{if and only if} \quad \text{bf } M \ N = P .$$

In this chapter, we will show the more important half of this result: that if $M, N \Rightarrow P$ then $\text{bf } M \ N = P$. That is, we can represent branch-first evaluation by a program. The proof of the converse, that `bf` does no more than the strategy allows, will require the rewriting theory of Chapter 7. Similar remarks apply to all the evaluators in this chapter.

The specification of `bf` is

$$Y_2\{\text{onFork}\{\text{triage}\{\text{bf_leaf}, \text{bf_stem}\{\text{eager}\}, \text{bf_fork}\}\}\}$$

as pictured in Figure 6.2. The term `onFork` will be defined so that leaves and stems are left unchanged, while f is applied to the branches of a fork. Then triage on the left branch will determine which of the three evaluation rules to apply. The strategic choice is captured by the use of `eager`, which will force the evaluation of branches before the root. Here are the details.

Define

$$\text{onFork}\{f\} = \lambda^* x. \Delta(\text{isFork2 } x)(\Delta x \Delta f)(K^3 x) \quad (x \text{ not in } f) .$$

If follows that $\text{onFork}\{f\} \Delta \text{bf} = K \Delta \text{bf} = \Delta$ as required. As usual, the side condition can be eliminated by unfolding the definition of star-abstraction. One solution for the three cases is given by

$$\begin{aligned}\text{bf_leaf} &=? \lambda^* y. \lambda^* b. \lambda^* z. y \\ \text{bf_stem}\{e\} &=? \lambda^* x. \lambda^* y. \lambda^* b. \lambda^* z. e(bxz)(b(byz)) \\ \text{bf_fork} &=? \lambda^* w. \lambda^* x. \lambda^* y. \lambda^* b. \lambda^* z. b(bzw)x\end{aligned}$$

Although this solution meets the specification, there are smaller solutions. And since bf appears in the frontispiece of the book, it is worth the effort to make it smaller. So define

$$\begin{aligned}D_b\{x\} &= d\{d\{Kx\}I\}(KD) \\ \text{bf_leaf} &= \lambda^* y. K(Ky) \\ \text{bf_stem}\{e\} &= \lambda^* x. \lambda^* y. d\{d\{K(Ke)\}D_b\{x\}\}(d\{d\{K\}D_b\{y\}\})(KD) \\ \text{bf_fork} &= \lambda^* w. \lambda^* x. K(d\{d\{K\}(d\{d\{Kw\}\})(KD))\}(K(d\{Kx\}))) .\end{aligned}$$

For example, if z is a program then

$$\begin{aligned}\text{bf } I z &= \text{onFork}\{\text{triage}\{\text{bf_leaf}, \text{bf_stem}\{\text{eager}\}, \text{bf_fork}\}\} I \text{bf } z \\ &= \text{triage}\{\text{bf_leaf}, \text{bf_stem}\{\text{eager}\}, \text{bf_fork}\} K K \text{bf } z \\ &= \text{bf_stem}\{\text{eager}\} \Delta K \text{bf } z \\ &= \text{eager} (\text{bf } \Delta z) (\text{bf } (\text{bf } K z)) \\ &= \text{eager } (\Delta z) (\text{bf } (\text{bf } K z)) \\ &= \text{bf } (\text{bf } K z) (\Delta z) \\ &= \text{bf } (Kz) (\Delta z) \\ &= \text{onFork}\{\text{triage}\{\text{bf_leaf}, \text{bf_stem}\{\text{eager}\}, \text{bf_fork}\}\} (Kz) \text{bf } (\Delta z) \\ &= \text{triage}\{\text{bf_leaf}, \text{bf_stem}\{\text{eager}\}, \text{bf_fork}\} \Delta z \text{bf } (\Delta z) \\ &= \text{bf_leaf } z \text{bf } (\Delta z) \\ &= z .\end{aligned}$$

Theorem 15 (branch_first_eval_to_bf) *If $M, N \Rightarrow P$ then $\text{bf } M \text{ } N = P$ for all programs M and N and combinations P .*

Proof. The proof is by induction on the structure of the derivation of $M, N \Rightarrow P$. There are five cases.

If M is a leaf then $P = \Delta N$ and $\text{bf } \Delta N = \Delta N$.

If M is a stem Δy then $P = \Delta y N$ and $\text{bf } (\Delta y) N = \Delta y N$.

If M is a fork of a leaf $\Delta \Delta y$ then $P = y$ and $\text{bf } (\Delta \Delta y) N = y$.

If M is a fork of a stem $\Delta(\Delta x)y$ then there are values v_1, v_2 and v_3 such that $y, N \Rightarrow v_1$ and $x, N \Rightarrow v_2$ and $v_1, v_2 \Rightarrow P$. Hence

$$\begin{aligned}\text{bf } (\Delta(\Delta x)y) N &= \text{eager } (\text{bf } x N) (\text{bf } (\text{bf } y N)) \\ &= \text{eager } v_2 (\text{bf } v_1) \\ &= \text{bf } v_1 v_2 \\ &= P\end{aligned}$$

by three applications of induction.

If M is a fork of a fork $\Delta(\Delta wx)y$ then there is a value v_1 such that $N, w \Rightarrow v_1$ and $v_1, x \Rightarrow P$. Hence

$$\begin{aligned}\text{bf } (\Delta(\Delta wx)y) z &= \text{bf } (\text{bf } N w) x \\ &= \text{bf } v_1 x \\ &= P\end{aligned}$$

by two applications of induction. \square

Hence, the self-evaluator can be applied to itself without requiring any additional machinery: if $\text{bf}, M \Rightarrow P$ then

$$\text{bf bf } M = P .$$

Although branch-first evaluation is quite efficient it does not find all results, since it may fail on an argument that is not required. Here is an example:

$$\begin{aligned}\Delta(\Delta \text{self_apply})(K(K\Delta)) \text{ self_apply} &= K(K\Delta) \text{ self_apply} (\text{self_apply self_apply}) \\ &= K\Delta (\text{self_apply self_apply}) .\end{aligned}$$

The natural progression is to simplify the whole combination to Δ . However, branch-first evaluation requires that the argument $\text{self_apply self_apply}$ be evaluated first, but this term has no value! In more detail, $\text{bf } \Delta(\Delta \text{self_apply})(K(K\Delta)) \text{ self_apply}$ evaluates to

$$\text{eager } (\text{bf self_apply self_apply}) (\text{bf } (\text{bf } (K\Delta))) .$$

Indeed, the role of *eager* in *bf* is precisely to ensure that all arguments are evaluated, as required by the branch-first evaluation strategy. It follows that this example requires a different strategy.

6.5 Root Evaluation

Root evaluation does just enough computation to determine the nature of the root of the tree, as a leaf, stem or fork, so that branches of a computation are evaluated only if needed. That is, the result of root evaluation must be able to represent unevaluated expressions even though these are not programs, are not values. This is achieved by *quotation*, which represents arbitrary computations by binary trees that have no stems.

The *quotation* of a combination is defined by

$$\begin{aligned}'\Delta &= \Delta \\ '(MN) &= \Delta('M)('N) .\end{aligned}$$

In tree calculus, quotation of arbitrary computations is not definable as a program since quotation does not preserve evaluation; that is, after all, the point. However, quotation of programs is definable, using

$$\begin{aligned}\text{quote} = Y_2\{\lambda^*x.\text{isStem } x &(\Delta(x\Delta)(K\Delta)(\lambda^*q.(\lambda^*x_1.K(qx_1)))) \\ &(\Delta x(K\Delta)(\lambda^*x_1.\lambda^*x_2.\lambda q.\Delta(K(qx_1))(qx_2)))\}\end{aligned}$$

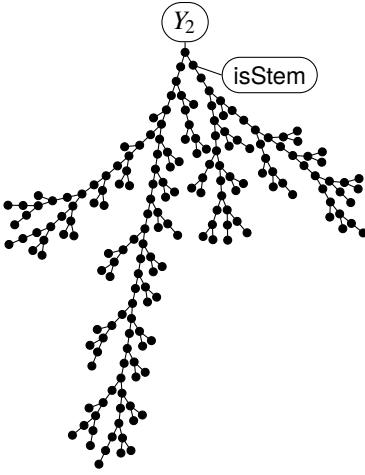


Figure 6.3: Quotation

The graph of quote is given in Figure 6.3.

Root evaluation is a relation $M \Rightarrow P$ between combinations whose rules are given in Figure 6.4. If M is a quotation then P is a factorable form whose branches are quotations. However, these conditions are not enforced by the rules since, for example, $\Delta\Delta(\Delta\Delta) \Rightarrow \Delta\Delta$ even though $\Delta\Delta$ is not a quotation.

The rules in Figure 6.4 are implemented by the program *root* given in Figure 6.5. Here is some explanation. It is easy to see that $\text{root } \Delta$ is an application of $\Delta\Delta\Delta$ which becomes Δ . Otherwise, the application of *root* to some quote Δfz becomes

onFork{rootn}(*root f*) *root z* .

Now consider *root f*. If this evaluates to a leaf or stem v_1 then the whole term evaluates

$$\begin{array}{c}
 \frac{}{\Delta \Rightarrow \Delta} \quad \frac{f \Rightarrow \Delta}{\Delta fz \Rightarrow \Delta z} \quad \frac{f \Rightarrow \Delta y}{\Delta fz \Rightarrow \Delta yz} \\
 \frac{f \Rightarrow \Delta ty \quad t \Rightarrow \Delta \quad y \Rightarrow v}{\Delta fz \Rightarrow v} \\
 \frac{f \Rightarrow \Delta ty \quad t \Rightarrow \Delta x \quad \Delta(\Delta yz)(\Delta xz) \Rightarrow v}{\Delta fz \Rightarrow v} \\
 \frac{f \Rightarrow \Delta ty \quad t \Rightarrow \Delta wx \quad \Delta(\Delta zw)x \Rightarrow v}{\Delta fz \Rightarrow v}
 \end{array}$$

Figure 6.4: Root Evaluation

$$\begin{aligned}
\text{rootl} &= \lambda^* r. \lambda^* y. \lambda^* z. ry \\
\text{roots} &= \lambda^* x. \lambda^* r. \lambda^* y. \lambda^* z. r(\Delta(\Delta yz)(\Delta xz)) \\
\text{rootf} &= \lambda^* w. \lambda^* x. \lambda^* r. \lambda^* y. \lambda^* z. r(\Delta(\Delta zw)x) \\
\text{rootn} &= \lambda^* t. \lambda^* y. \lambda^* r. \text{triage}\{\text{rootl}, \text{roots}, \text{rootf}\} (rt)ry \\
\text{root} &= Y_2\{\lambda^* a. \lambda^* r. \Delta a \Delta (\lambda^* f. \text{onFork}\{\text{rootn}\}(rf)r) .
\end{aligned}$$

Figure 6.5: A Root Self-Evaluator

$$\frac{x \Rightarrow \Delta}{x \Downarrow \Delta} \quad \frac{x \Rightarrow \Delta y \quad y \Downarrow v}{x \Downarrow \Delta v} \quad \frac{x \Rightarrow \Delta yz \quad y \Downarrow v \quad z \Downarrow w}{x \Downarrow \Delta vw}$$

Figure 6.6: Root-and-Branch Evaluation

to $v_1 z$ as required. If it evaluates to some fork $\Delta t y$ then the whole term evaluates to

$$\text{rootn } t \ y \ \text{root } z = \text{triage}\{\text{rootl}, \text{roots}, \text{rootf}\} (\text{root } t) \ \text{root } y \ z$$

Finally, triage on $\text{root } t$ triggers evaluations that express the last three rules of the strategy.

Theorem 16 (root_eval_to_root) *If $M \Rightarrow P$ then $\text{root } M = P$ for all combinations M .*

Proof. The proof is by induction on the structure of the derivation. \square

A converse to this theorem will be proved in Chapter 7.

6.6 Root-and-Branch Evaluation

Root-and-branch evaluation performs root evaluation and then recursively evaluates the branches to produce a normal form. The rules for this evaluation $M \Downarrow P$ are given in Figure 6.6. It is implemented by

$$\text{rb} = Y_2\lambda^* x. \lambda^* r. \text{triage}\{K\Delta, \lambda^* y. \lambda^* r. \Delta(ry), \lambda^* y. \lambda^* z. \Delta(ry)(rz)\}(\text{root } x) r .$$

Theorem 17 (rb_eval_implies_rb) *If ' $M \Downarrow P$ ' then ' $\text{rb}'M = P$ ' for all combinations M and programs P .*

Proof. The proof is by induction on the structure of the derivation. \square

6.7 Root-First Evaluation

Finally, root-first evaluation of the application of the program M to the program N is given by root-and-branch evaluation of ' (MN) '. Since quotation of programs is definable, root-first evaluation is implemented by

$$\text{rf} = \lambda^* f. \lambda^* z. \text{rb} (\Delta (\text{quote } f)(\text{quote } z)) .$$

Theorem 18 (root_first_eval_to_rf) *If ' $(MN) \Downarrow P$ ' then $\text{rf } M \ N = P$ for all programs M and N and P .*

As noted earlier, each of the theorems in this chapter has a converse, showing that if the self-evaluator produces a result then so does the evaluation strategy. However, the proofs require the expressive power of rewriting theory.

Exercises

1. Try to evaluate the following applications of programs using branch-first evaluation:

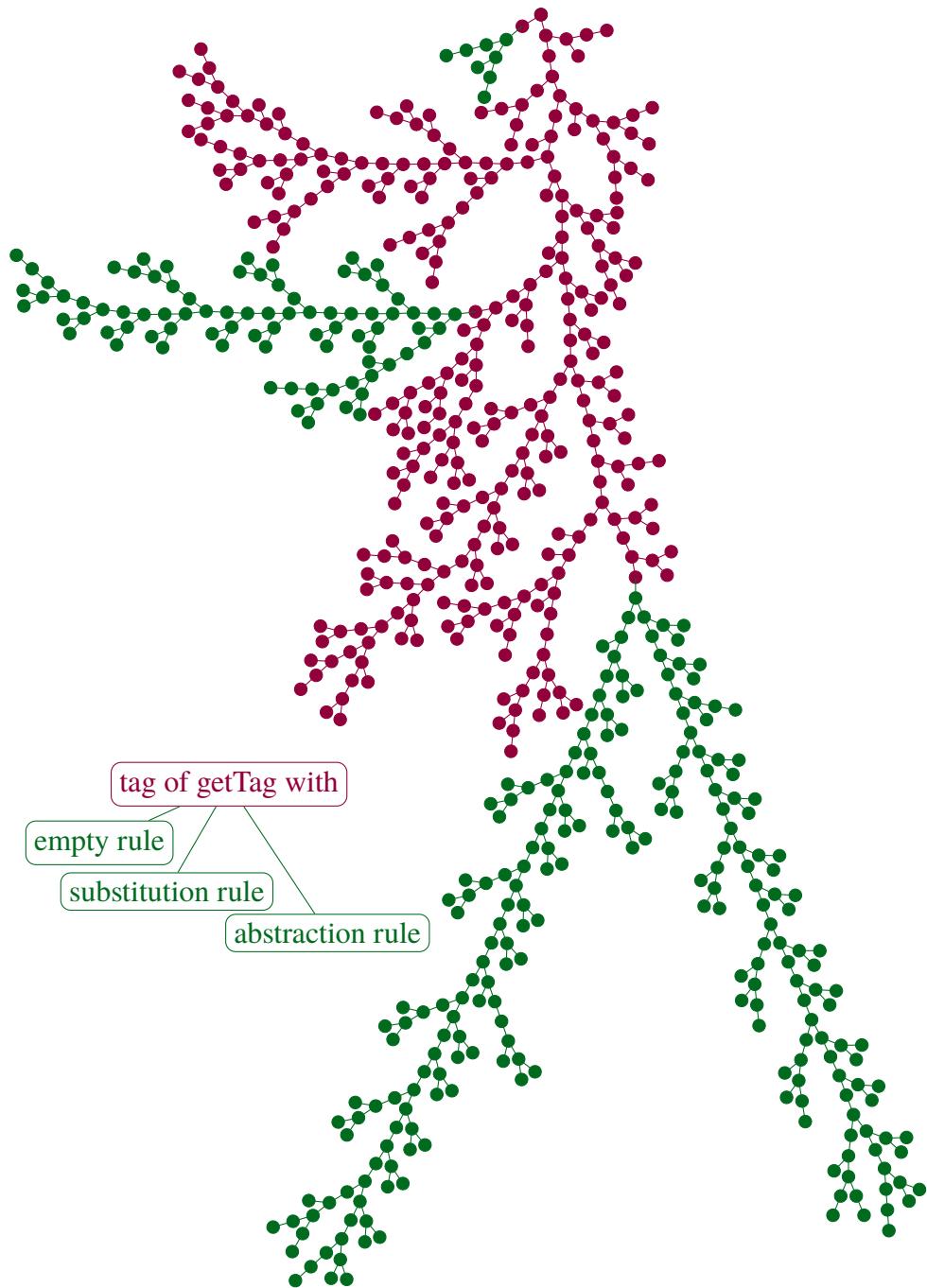
$$\begin{aligned} & K\Delta K \\ & KI\Delta K \\ & (\Delta(\Delta I)I)(\Delta(\Delta I)I) \\ & K(\Delta(\Delta I)I)(\Delta(\Delta I)I)K \\ & KI(\Delta(\Delta I)I)(\Delta(\Delta I)I)K . \end{aligned}$$

Hint: evaluate them by hand first, to look for traps.

2. Repeat the previous exercise using root-first evaluation. What has changed?

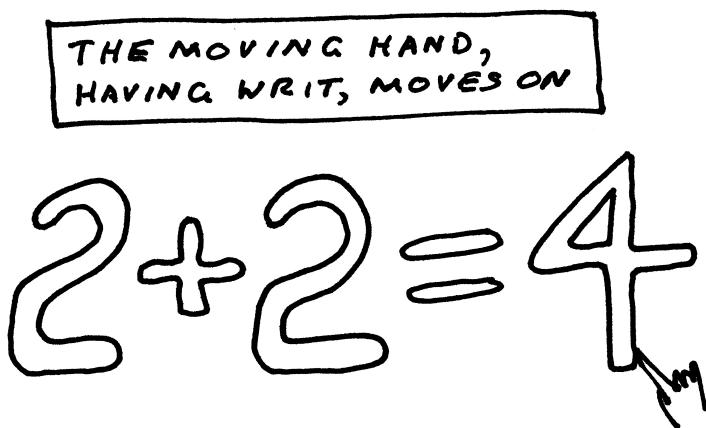
Part II

Other Models of Computation

(A TREE FOR λ)

Chapter 7

Rewriting



Abstract

Equational reasoning is not well suited to showing that bad things don't happen, e.g. that one is not equal to two. Imposing a direction on the equations yields a reduction relation, in which numerals cannot be reduced. By showing that equal terms have a common reduct, we can prove that one and two are not equal. These ideas will be applied to the arithmetic expressions and to tree calculus.

7.1 Proving Negative Results

Equational reasoning is well able to show that things do happen, e.g. that applying the program `bf` produces the same results as branch-first evaluation. However, it is ill-suited to showing that things don't happen, e.g. that `bf` does not produce any results other than those obtained by branch-first evaluation. More fundamentally, it cannot show that stems are never equal to forks. This would require an analysis of all possible equational calculations, which is altogether harder than merely producing a particular chain of equations. When the equations of tree calculus are only ever applied from left to right then it is easy to show that stems reduce to stems and forks to forks but this direction is not enforced, so that we might conceivably have

$$\Delta\Delta = \Delta\Delta\Delta\Delta\Delta = \dots = \Delta\Delta\Delta?$$

That such a chain of equations cannot exist is completely dependent on the choice of the original evaluation rules of tree calculus. For example, if we were to add the equation $\Delta\Delta\Delta = \Delta\Delta$ then the result would be immediate. Further, from this we could show

$$\begin{aligned}\Delta x &= \Delta\Delta\Delta\Delta x \\ &=? \quad \Delta\Delta\Delta x \\ &= \Delta\end{aligned}$$

which would make every combination equal to Δ .

To show that this sort of derivation is impossible we will replace equational rule $M = N$ with *reduction rules* $M \rightarrow N$ that are oriented from left to right.

Now there are two equality relations on terms. The *structural equality* $M \equiv N$ asserts that M and N have exactly the same structure. The *computational equality* $M = N$ is generated by the reduction rules by declaring that if $M \rightarrow N$ then $M = N$. For example, $\Delta\Delta\Delta\Delta$ is not structurally equal to $\Delta\Delta\Delta(\Delta\Delta)$ but they are computationally equal, since both reduce to Δ .

The reduction rules give less freedom than computational equality because the rules have a direction from left to right, but have more freedom than, say, branch-first evaluation as the reduction rules can still be applied to any sub-term. This proves to be a convenient place to prove results by structural induction, which works with sub-terms.

7.2 Rewriting

A *reduction system* is given by some collection of terms and a binary relation on those terms. If M and N are so related then we may write

$$M > N .$$

The greater-than symbol $>$ is used to suggest that it imposes an order on the terms, in which the *redex* M may be replaced by its *reduct* N but not vice-versa. Closing these rules under all of the term formation rules produces the corresponding *rewriting relation* $r \rightarrow s$. Since this book is never concerned with reduction systems that are not

rewriting systems, we will write $M \rightarrow N$ for $M > N$ without any risk of confusion. Just as we have chains of equations $M_1 = M_2 = \dots = M_k$ we now have chains of reductions $M_1 \rightarrow M_2 \rightarrow \dots \rightarrow M_k$. When such a chain exists we may write $M_1 \rightarrow^k M_k$ for the k -step reduction relation. When the choice of k is not important, we may write $M \rightarrow^* N$. That is, \rightarrow^* is the reflexive, transitive closure of \rightarrow . We may write $M \rightarrow N$ for $M \rightarrow^* N$ when the distinction between a *one-step reduction* and a *multi-step reduction* is not important. As noted above, every rewriting relation gives rise to a corresponding notion of *computational equality* which is obtained by adding the rule $N \rightarrow M$ whenever we have $M \rightarrow N$.

It is not always clear how to orient evaluation equations to form rewriting rules. For example, consider the rules for Roman numerals in Section 2.7. If they are all oriented from left to right, then canonical forms will be tallies, consisting of *I*s only. Alternatively, if the rules are oriented from right to left, then the traditional representations of Roman numerals will become canonical.

The general ideas introduced in the rest of the chapter will be illustrated using two examples derived from Part I, namely the arithmetic expressions of Chapter 2 and the tree calculus of Chapter 3.

Recall that the arithmetic expressions in Section 2.8 are given by the BNF

$$m, n ::= \text{zero} \mid \text{successor } m \mid (m + n) .$$

The rules for addition have a natural order, as

$$\begin{aligned} \text{zero} + n &> n \\ \text{successor } m + n &> m + \text{successor } n . \end{aligned}$$

Note that this order is different from the numerical order in which five is greater than three. The closure of the addition rules above under the term formation rules is achieved by

- (i) If $m_1 \rightarrow m_2$ then $\text{successor } m_1 \rightarrow \text{successor } m_2$.
- (ii) If $m_1 \rightarrow m_2$ then $m_1 + n \rightarrow m_2 + n$.
- (iii) If $n_1 \rightarrow n_2$ then $m_1 + n_1 \rightarrow m_1 + n_2$.

Note that there are two rules for reducing additions; one for reducing on the left of the addition, and one for reducing on its right. This distinction is important when discussing the order in which reductions are made. It is routine to show that multi-step reduction preserves the structure.

Often properties of one-step reduction carry over to multi-step reduction. Here are some examples.

Theorem 19 *If $e_1 \rightarrow^* e_2$ then $\text{successor } e_1 \rightarrow^* \text{successor } e_2$.*

Proof. The proof is by induction on the length of the reduction $e_1 \rightarrow^* e_2$. If the reduction takes no steps then e_2 is e_1 and the results follow by reflexivity. Suppose that

the reduction of e_1 is given by $e_1 \rightarrow e_3 \rightarrow^* e_2$ and the result holds for $e_3 \rightarrow^* e_2$. Then we have

$$\text{successor } e_1 \rightarrow \text{successor } e_3 \rightarrow^* \text{successor } e_2$$

by induction, and the result follows by transitivity of \rightarrow^* . \square

Theorem 20 *If e_1, e_2 and e_3 are arithmetic expressions such that $e_1 \rightarrow^* e_2$ then $e_1 + e_3 \rightarrow^* e_2 + e_3$ and $e_3 + e_1 \rightarrow^* e_3 + e_2$.*

Proof. The proof is by induction on the length of the reduction $e_1 \rightarrow^* e_2$. If the reduction takes no steps then e_2 is e_1 and the results follow by reflexivity.

Suppose that the reduction of e_1 is given by $e_1 \rightarrow e_4 \rightarrow^* e_2$ and the result holds for $e_4 \rightarrow^* e_2$. Then we have

$$e_1 + e_3 \rightarrow e_4 + e_3 \rightarrow^* e_2 + e_3$$

by induction. Similarly, we have $e_3 + e_1 \rightarrow e_3 + e_4 \rightarrow^* e_3 + e_2$. \square

Corollary 21 *If e_1, e_2, e_3 and e_4 are arithmetic expressions such that $e_1 \rightarrow^* e_2$ and $e_3 \rightarrow^* e_4$ then $e_1 + e_3 \rightarrow^* e_2 + e_4$.*

Proof. By two applications of the previous theorem, we have

$$e_1 + e_3 \rightarrow^* e_2 + e_3 \rightarrow^* e_2 + e_4$$

and the result follows by transitivity. Note that we could have used a different reduction sequence, namely

$$e_1 + e_3 \rightarrow^* e_1 + e_4 \rightarrow^* e_2 + e_4 .$$

\square

For tree calculus, the evaluation rules will be oriented from left to right, to get

$$\begin{array}{ll} \Delta\Delta yz & \longrightarrow y \\ \Delta(\Delta x)yz & \longrightarrow yz(xz) \\ \Delta(\Delta wx)yz & \longrightarrow zwx \end{array} \quad \begin{array}{l} (K) \\ (S) \\ (F) \end{array}$$

7.3 Normal Forms

From an abstract point of view, the goal of rewriting is to reduce a term until it can reduce no more. A term r is *irreducible* if it cannot be reduced, i.e. for all terms s there is no one-step reduction $r \rightarrow s$.

In practice, however, the goal is to produce a result or value that is in a particular class of sub-terms. For example, the arithmetic expressions are to produce numbers, represented by the *numerals*, given by the BNF

$$m, n ::= \text{zero} \mid \text{successor } m .$$

Again, the trees are to produce programs, represented by the binary trees.

When these values are exactly the irreducible terms then they may also be called *normal forms*.

As well as the normal forms, it is sometimes convenient to consider terms which reduce to a normal form. If t is a term that reduces to a normal form n by some reduction sequence $t \rightarrow^* n$ then t is *normalisable* or *valuable*. If, further, every reduction sequence starting at t stops at some normal form then t is *strongly normalisable*. If t is merely normalisable we may say it is *weakly normalisable* to emphasise that it is not necessarily strongly normalisable.

For example, suppose that M and N are programs of tree calculus. Then $\text{wait}\{M, N\}$ is also normal but WMN is not. Rather it is strongly normalisable, since all reduction sequences end with $\text{wait}\{M, N\}$. Weak normality is common among applications of the fixpoint combination Y . For example, $Y(KI)\Delta$ is weakly normalisable since

$$Y(KI)\Delta \rightarrow^* KI(Y(KI))\Delta \rightarrow^* \Delta$$

but it is not strongly normalisable since $Y(KI) \rightarrow^* KI(Y(KI))$ yields unbounded reduction sequences.

A reduction relation is *normalising* if each term has a normal form. That is, there is some *some* reduction sequence which reduces it to normal form. A reduction relation is *strongly normalising* if every reduction sequence must terminate. This distinction between "some" and "every" is important in practice, since if some reduction sequences do not terminate then either cleverness or strategy may be required to find the normal form.

Theorem 22 *Every arithmetic expression reduces to a numeral.*

Proof. Let e be an arithmetic expression. The proof is by induction on the structure of e . If e is zero then the result is immediate. If e is successor e_1 and e_1 has the property, then there is a numeral n_1 such that $e_1 \rightarrow^* n_1$. Choose n to be successor n_1 . Then $e \rightarrow^* n$ by Theorem 19.

If e is of the form $e_1 + e_2$ and e_1 and e_2 have the property then there are numerals n_1 and n_2 such that $e_1 \rightarrow^* n_1$ and $e_2 \rightarrow^* n_2$. By Corollary 21, we have $e_1 + e_2 \rightarrow^* n_1 + n_2$. By Theorem 22 there is a numeral n such that $n_1 + n_2 \rightarrow^* n$. Then

$$e = e_1 + e_2 \rightarrow^* n_1 + n_2 \rightarrow^* n$$

as required. \square

Hence, reduction of arithmetic expressions is weakly normalising. Actually, it is strongly normalising, too.

In tree calculus, it is easy to see that the irreducible combinations are exactly the programs. However, it is tricky to characterise the irreducible terms, since variables may block reduction in various ways. For example,

$$x\Delta\Delta\Delta \quad \text{and} \quad \Delta(x\Delta\Delta\Delta)\Delta\Delta$$

are normal forms but

$$\Delta(\Delta(x\Delta\Delta\Delta))\Delta\Delta\Delta$$

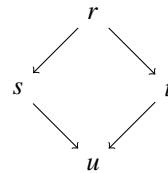
reduces to $\Delta\Delta(x\Delta\Delta\Delta\Delta)\Delta$ and $x\Delta\Delta\Delta\Delta$. Happily, we will not need to characterise the terms that are normal forms, as our focus is on the combinations.

All of the equational reasoning about tree calculus in Part I carries over to the reduction system considered here. Indeed, all of the equational proofs in Coq for Part I have been reworked for the rewriting system, too.

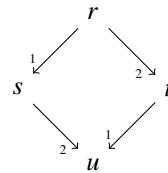
7.4 The Diamond Property

We have seen that every arithmetic expression can be reduced to a numeral. This section will show that the normal form of an arithmetic expression is unique. From this, it follows that numerals are computationally equal if and only if they are structurally equal. There are several ways to do this, but we will adopt an approach which facilitates the later discussion of tree calculus, by introducing the diamond property.

A reduction relation \rightarrow satisfies the *diamond property* if, for each pair of reductions $r \rightarrow s$ and $r \rightarrow t$ there is a term u such that $s \rightarrow u$ and $t \rightarrow u$:



More generally, a *pair* of relations \rightarrow_1 and \rightarrow_2 satisfy the *diamond property* if, for each pair of reductions $r \rightarrow_1 s$ and $r \rightarrow_2 t$ there is a term u such that $s \rightarrow u$ and $t \rightarrow u$:



Clearly, if \rightarrow_1 and \rightarrow_2 satisfy the diamond property, then so do \rightarrow_2 and \rightarrow_1 .

The diamond property extends to reflexive, transitive closures in two steps.

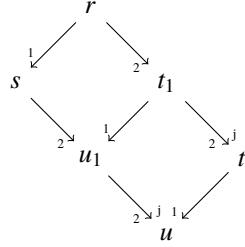
Lemma 23 *Whenever the diamond property holds for a pair of relations \rightarrow_1 and \rightarrow_2 then it holds for \rightarrow_1 and \rightarrow_2^k for any natural number k . Hence it holds for \rightarrow_1 and r_2^* .*

Proof. The proof is by induction on k . Suppose that $r \rightarrow_1 s$ and $r \rightarrow_2^k t$.

If $k = 0$ then $t = r$ and the diamond is completed by s .

If k is the successor of j then there is some t_1 such that $r \rightarrow_2 t_1 \rightarrow_2^j t$. By the diamond property for \rightarrow_1 and \rightarrow_2 there is some u_1 such that $s \rightarrow_2 u_1$ and $t_1 \rightarrow_1 u_1$. By induction, there is a u such that $u_1 \rightarrow_2^j u$ and $t \rightarrow_1 u$. Hence u is the

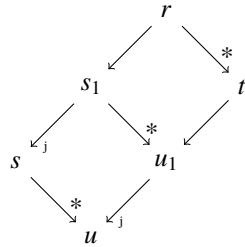
desired term, as shown by the diagram below:



□

Lemma 24 *If \rightarrow has the diamond property then so does \rightarrow^* .*

Proof. It is enough to show that \rightarrow^k and \rightarrow^* have the diamond property for any k . Suppose that $r \rightarrow^k s$ and $r \rightarrow^* t$. If k is 0 then the result is immediate. If k is the successor of some j and we have $r \rightarrow s_1 \rightarrow^j s$ then we can apply Lemma 23 to r, s_1 and t to produce a common reduct u_1 of s_1 and t . Then induction yields a common reduct u of s and u_1 as in the following diagram



□

Now we can show that arithmetic expressions evaluate to at most one numeral.

Theorem 25 *Reduction of arithmetic expressions satisfies the diamond property.*

Proof. Suppose that r, s and t are arithmetic expressions such that $r \rightarrow s$ and $r \rightarrow t$. The proof is by induction on the structure of r . If r is a numeral then it is irreducible, so choose u to be r . If r is an addition $r_1 + r_2$ then there are four possible ways that r can reduce to s .

(Case 1) Suppose s is $s_1 + r_2$ and $r_1 \rightarrow s_1$. Now consider $r \rightarrow t$. Again there are four possibilities.

Suppose that t is $t_1 + r_2$ and $r_1 \rightarrow t_1$. Then by induction there is a term u_1 such that $s_1 \rightarrow u_1$ and $t_1 \rightarrow u_1$ so choose u to be $u_1 + r_2$.

Suppose that t is $r_1 + t_2$ and $r_2 \rightarrow t_2$. Then choose $u = s_1 + t_2$.

Suppose that r_1 is zero. Then r_1 is both irreducible and reduces to s_1 which yields a contradiction.

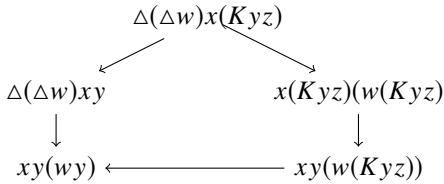


Figure 7.1: Reduction may Duplicate Redexes

Finally, suppose that r_1 is some successor successor r_3 and t is $r_3 + \text{successor } r_2$. Then s_1 must be of the form successor s_3 where $r_3 \rightarrow s_3$. So choose u to be $s_3 + \text{successor } r_2$.

(Case 2) Suppose that s is $r_1 + s_2$ where $r_2 \rightarrow s_2$. Then proceed as in Case 1.

(Case 3) Suppose that r_1 is zero and s is r_2 . Then either t is r_2 , in which case choose u to be r_2 or t is of the form zero + t_2 where $r_2 \rightarrow t_2$, in which case choose u to be t_2 .

(Case 4) Suppose that r_1 is some successor successor r_3 and s is $r_3 + \text{successor } r_2$. Then there are four possibilities for $r \rightarrow t$ which are handled by variations on the arguments above. \square

Theorem 26 *If an arithmetic expression e reduces to two numerals m and n then m and n are structurally equal.*

Proof. Since \rightarrow satisfies the diamond property, it follows that \rightarrow^* satisfies the diamond property, by Lemma 24. Hence, there is an expression u such that $m \rightarrow^* u$ and $n \rightarrow^* u$. However, numerals are irreducible by \rightarrow so it must be that $u \equiv m$ and $u \equiv n$. Hence $m \equiv n$ by transitivity. \square

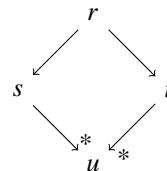
7.5 Confluence of Tree Calculus

Unfortunately, the diamond property fails for the one-step reduction of tree calculus because the stem rule may duplicate redexes, as in Figure 7.1.

That is, it takes *two reduction steps* on the right to complete the diagram, when the diamond property insists on using one only.

In order to show that every combination reduces to at most one program, it is enough to show that reduction is confluent, in the following sense.

A relation \rightarrow is *confluent* if $r \rightarrow s$ and $r \rightarrow t$ implies there is a term u such that $s \rightarrow^* u$ and $t \rightarrow^* u$



$$\begin{array}{c}
 \frac{}{\Delta \rightarrow_s \Delta} \quad \frac{x \rightarrow_s x' \quad y \rightarrow_s y'}{xy \rightarrow_s x'y'} \quad \frac{y \rightarrow_s y'}{\Delta\Delta yz \rightarrow_s y'} \\
 \frac{x \rightarrow_s x' \quad y \rightarrow_s y' \quad z \rightarrow_s z'}{\Delta(\Delta x)yz \rightarrow_s y'z'(x'z')} \quad \frac{w \rightarrow_s w' \quad x \rightarrow_s x' \quad z \rightarrow_s z'}{\Delta(\Delta wx)yz \rightarrow_s z'w'x'}
 \end{array}$$

Figure 7.2: Simultaneous Reduction in Tree Calculus

Theorem 27 If \rightarrow is a confluent reduction relation then each term has at most one normal form.

Proof. Suppose that $r \rightarrow^* s$ and $r \rightarrow^* t$ and that s and t are irreducible. Since \rightarrow is confluent, it follows that \rightarrow^* satisfies the diamond property, so that there is a term u such that $s \rightarrow^* u$ and $t \rightarrow^* u$. Since s and t are irreducible, it follows that s and t are both u . \square

Corollary 28 With respect to a confluent, weakly normalising reduction, every term has a unique normal form.

Corollary 29 Every arithmetic expression has a unique normal form.

Actually, we will not require confluence in this general form. Instead, we will show the diamond property for \rightarrow^* .

Lemma 30 \rightarrow is confluent iff \rightarrow^* satisfies the diamond property.

Proof. Exercise. \square

This may seem like a backward step, in that it is harder to reason about multi-step reductions than single steps. However, we will generalise the reduction relation to support *simultaneous reduction* \rightarrow_s so that both copies of the redex Kxy in Figure 7.1 can be reduced in a single step, as $x(Kyz)(w(Kyz)) \rightarrow_s xy(wy)$. More generally, \rightarrow_s will prove to satisfy the diamond property, so that \rightarrow_s^* does too. Finally, \rightarrow_s^* will prove to be the same relation as \rightarrow^* so that the latter satisfies the diamond property and \rightarrow is confluent.

The *simultaneous reduction* relation is given in Figure 7.2. Now the pentagon in Figure 7.1 can be collapsed to a diamond with respect to simultaneous reduction, as shown in Figure 7.3.

Lemma 31 For all combinations M and N of tree calculus, $M \rightarrow^* N$ if and only if $M \rightarrow_s^* N$.

Theorem 32 (diamond_s_red1) Simultaneous reduction has the diamond property.

Theorem 33 (confluence_tree_calculus) Reduction of tree calculus is confluent.

Proof. Since reduction has the same transitive closure as simultaneous reduction, it is enough to prove confluence of the latter, which follows from Theorem 32. \square

Corollary 34 Every tree has at most one normal form.

Proof. Apply Theorem 27. \square

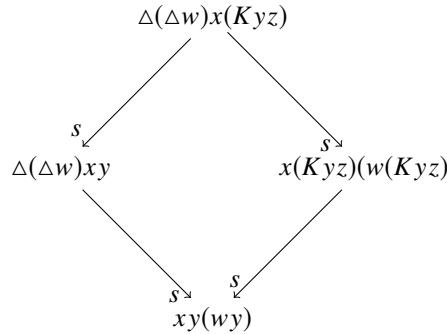


Figure 7.3: Simultaneous Reduction can Handle Duplicate Redexes

7.6 The Halting Problem

Some trees don't have any normal form. For example, the application of $\omega = \Delta(\Delta I)I$ to itself reduces to itself by

$$\omega\omega = (\Delta(\Delta I)I\omega \longrightarrow I\omega(I\omega) \longrightarrow \omega\omega).$$

With a little effort one can prove that no other reduction sequence does any better.

Given that trees without normal forms exist, can the existence of a normal form be decided by a program? This is the adaptation to tree calculus of the Halting Problem posed by Alan Turing; the answer is no.

Define a *halting function* to be a combination h such that, for each combination x , either hx reduces to K and x has a normal form or hx reduces to KI and x does not have a normal form.

Theorem 35 (halting_problem_insoluble) *There is no halting function in tree calculus.*

Proof. Suppose that h is a halting function. Define

$$h' = \lambda^* x. h(xx)(\omega\omega)K.$$

Then $h'h' = h(h'h')(\omega\omega)K$. Now, if $h(h'h') = K$ then $h'h'$ has a normal form but also $h'h' = K(\omega\omega)K = \omega\omega$ which does not have a normal form. Contradiction. On the other hand, if $h(h'h') = KI$ then $h'h'$ does not have a normal form but also $h'h' = KI(\omega\omega)K = K$ which does have a normal form. Contradiction. Since both alternatives lead to contradiction there can be no halting function h . \square

7.7 Standard Reduction

The root-first evaluation of Chapter 6.7 has a useful property that is not shared by the other evaluation strategies of Chapter 6: if a combination M has a normal form N then

root-and-branch evaluation will find it. This follows from a more general result: if N contains k redexes then there is a standard reduction sequence from M to N with at most k violations of the root-and-branch strategy, in a sense that will be made precise. Thus, if k is zero then there are no violations, so that root-and-branch evaluation succeeds. An outline of the proof follows: details can be found in the Coq verification. It adapts the work of Ryo Kashima on lambda-calculus.

A combination x is *ready* if $x\Delta$ is a redex. That is, x is of the form Δty where t is factorable.

An *exact right reduction* $M \rightarrow_n^1 N$ is a reduction $M \rightarrow N$ in which the redex in M is to the right of exactly n other redexes in M . Call n the number of *violations* of root-first evaluation.

A *right reduction* $M \rightarrow_n N$ is given by a sequence of exact right reductions, with some constraints on the violations, that are captured by the following derivation rules

$$\frac{}{M \rightarrow^n M} n \leq \text{redexes}(M) \quad \frac{M \rightarrow_m^1 N \quad N \rightarrow^n P}{M \rightarrow_n P} m \leq n.$$

That is, when right reducing from M to P , the number of possible violations may grow with each step, provided that the final tally is no greater than the number of redexes in P .

If no violations are allowed, so that $M \rightarrow^0 P$, then this is called *leftmost, outermost reduction*.

Head reduction is a restricted form of leftmost, outermost reduction which stops when a factorable form is found. One-step head reduction is given by

$$\frac{\Delta\Delta yz \rightarrow_h y}{\Delta(\Delta x)yz \rightarrow_h yz(xz)} \quad \frac{\Delta(\Delta x)yz \rightarrow_h yz(xz)}{\Delta(\Delta wx)yz \rightarrow_h zwx} \\ \frac{x \rightarrow_h x'}{xz \rightarrow_h x'z} \quad \frac{x \rightarrow_h x'}{\Delta xyz \rightarrow_h \Delta x'yz}$$

A *staged reduction* $M \rightarrow_s P$ begins with a sequence of head reductions $M \rightarrow_h^* N$ where N is a factorable form. Then, if N is some application N_1N_2 it must be that P is some application P_1P_2 with staged reductions from N_1 to P_1 and N_2 to P_2 .

Theorem 36 (standardization) *If $M \rightarrow^* N$ then $M \rightarrow_n N$ for all combinations M and N where n is the number of redexes in N .*

Proof. If $M \rightarrow^* N$ then lemmas show that there is a staged reduction from M to N and so $M \rightarrow_{\text{redexes}(N)} N$. \square

Corollary 37 (leftmost_reduction) *If M is a combination and N is a program then $M \rightarrow^* N$ implies there is a leftmost, outermost reduction from M to N .*

Proof. If N is a program then it has no redexes. Now apply Theorem 36. \square

Theorem 38 (head_reduction_to_factorable_form) *If M is a combination and N is a factorable form and $M \rightarrow^* N$ then there is a factorable form Q such that $M \rightarrow_h^* Q$ and $Q \rightarrow^* N$.*

Proof. As above, there is a staged reduction from M to N . This is given by some head reduction $M \xrightarrow{h}^* Q$ followed by some reduction of the branches of Q to produce N . Since N is a factorable form, it follows that Q is too. \square

7.8 Converse Properties of the Self-Evaluators

Equipped with the machinery of rewriting, we can now prove that the self-evaluators of Chapter 6 do no more than is allowed by the corresponding strategies. All the proofs follows the same pattern. Let us consider branch-first evaluation and `bf`.

The key idea is that if `bf` $M N$ has a factorable form then so does M . That is, `bf` $M N$ needs M since the whole term reduces to something whose next evaluation step must involve M . The proof of the theorem proceeds by induction on the number of simultaneous reduction steps required to reduce `bf` $M N$ and then case analysis on M . For example, if M is of the form $\Delta(\Delta w x) y$ then `bf` $(\text{bf } N w) x$ also reduces to P , and this in fewer steps. Since P is factorable, and `bf` needs its first argument, it follows that `bf` $N w$ reduces to some factorable form, N_1 . Thus $N, w \Rightarrow N_1$ by induction. Further, `bf` $N_1 x$ also reduces to P , so a second use of induction shows that $N_1, x \Rightarrow P$ and the result follows. Although the idea is not so difficult to grasp, the formal proof is quite dense. Here are the definitions required for the key lemma.

A term M is *next* in a term R if R is of the form ΔMNP or if R is of the form $R_1 P$ where M is next in R_1 .

A term M is *active* in a term R if either M is next in R or there is some N such that M is next in N and N is active in R .

A term R *needs* a term M if R reduces to some P in which M is active.

It is routine to prove that `bf` $M N$ needs M and `eager` $M N$ needs M .

Lemma 39 (needy_s_red_factorable) *If R needs M and R reduces to a factorable form P in n simultaneous reduction steps then M reduces to some factorable form Q in n simultaneous steps.*

Proof. The proof is by induction on n . \square

Theorem 40 (bf_to_branch_first_eval) *For all programs M and N and factorable forms P , if `bf` $M N \xrightarrow{*} P$ then there is a Q such that $M, N \Rightarrow Q$ and $P \xrightarrow{*} Q$.*

Proof. The proof is by induction on the number of simultaneous reductions to P and then case analysis of M . If M is a leaf or stem then `bf` $M N \xrightarrow{*} MN$ and $M, N \Rightarrow MN$ so the result follows from confluence of simultaneous reduction. So suppose that M is a fork $\Delta M_0 M_1$. If M_0 is a leaf Q then `bf` $M N \xrightarrow{*} M_1$ and $M, N \Rightarrow M_1$ and the result follows from confluence.

If M_1 is a fork $\Delta M_2 M_3$ then `bf` $M N \xrightarrow{*} \text{bf} (\text{bf } N M_2) M_3$. By confluence, there is some P_1 with simultaneous reductions from `bf` $(\text{bf } N M_2) M_3$ and P . It follows that P_1 is factorable. Further, the first of these two reductions takes the same number of steps as the original reduction to P . Actually, since there is a unique simultaneous reduction of `bf` $(\Delta(\Delta M_3 M_4) M_2) N$ to some P_0 we can rebuild the diamond to show that

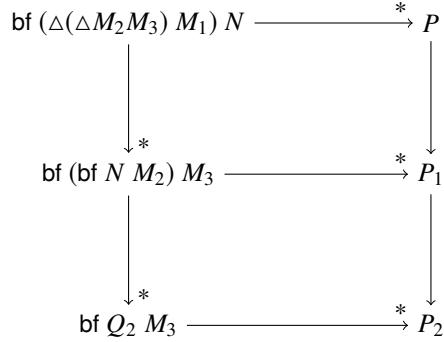


Figure 7.4: Branch-First Evaluation Using the Fork Rule

the simultaneous reduction from $\text{bf } (\text{bf } N M_3) M_4 \xrightarrow{*} P_1$ takes one less step than the original. This is illustrated in the first square of Figure 7.4.

Thus, by Lemma 39 there is a simultaneous reduction of $\text{bf } N M_2$ to some factorable form Q_0 . Hence, by the induction hypothesis, there is a branch-first evaluation $N, M_2 \Rightarrow Q_2$ for some Q_2 . Further, Q_2 is a program, by Theorem 14. So a second application of Lemma 39 yields a factorable form P_2 that completes the second square in Figure 7.4.

Finally, a second application of the induction hypothesis yields an evaluation $Q_2, M_3 \Rightarrow P_3$ for some factorable form P_3 where P_2 reduces to P_3 . Now P_3 is the required factorable form. The reduction from P is by transitivity of reduction: the evaluation is given by

$$\frac{N, M_2 \Rightarrow Q_2 \quad Q_2, M_3 \Rightarrow P_3}{\Delta(\Delta M_2 M_3) M_1, N \Rightarrow P_3}$$

If M_0 is a stem ΔM_2 then $\text{bf } M N \xrightarrow{*}$ eager ($\text{bf } M_2 N$) ($\text{bf } (\text{bf } M_1 N)$) and the argument follows a similar pattern as above, as illustrated in Figure 7.5. \square

Corollary 41 (branch_first_eval_iff_bf) *For all programs M, N and P , the computation $\text{bf } M N$ reduces to P if and only if $M, N \Rightarrow P$.*

Proof. Apply Theorems 15 and 40. \square

Happily, the same techniques apply to root evaluation, root-and-branch evaluation and root-first evaluation.

Theorem 42 (root_eval_iff_root) *For all combinations M and programs P , the computation $\text{root } 'M$ reduces to P if and only if $'M \Rightarrow P$.*

Proof. In one direction, apply Theorem 16. In the other direction, apply the techniques of Theorem 40. \square

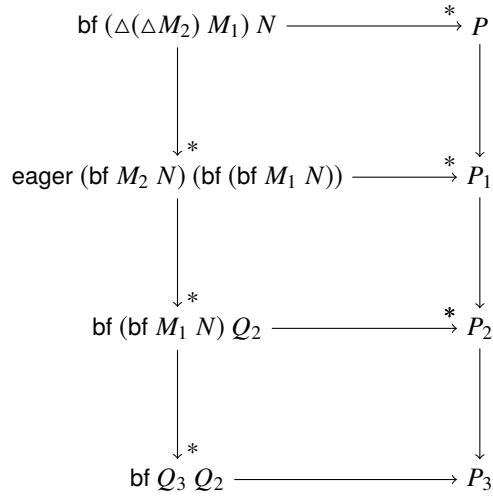


Figure 7.5: Branch-First Evaluation Using the Stem Rule

Theorem 43 (rb_eval_iff_rb) *For all combinations M and programs P , the computation $\text{rb}'M$ reduces to P if and only if $'M \Downarrow P$.*

Proof. In one direction, apply Theorem 17. In the other direction, use induction on the size of P and apply Theorem 42. \square

Theorem 44 (root_first_eval_iff_rf) *For all programs M, N and P , the computation $\text{rf } M N$ reduces to P if and only if $'(MN) \Downarrow P$.*

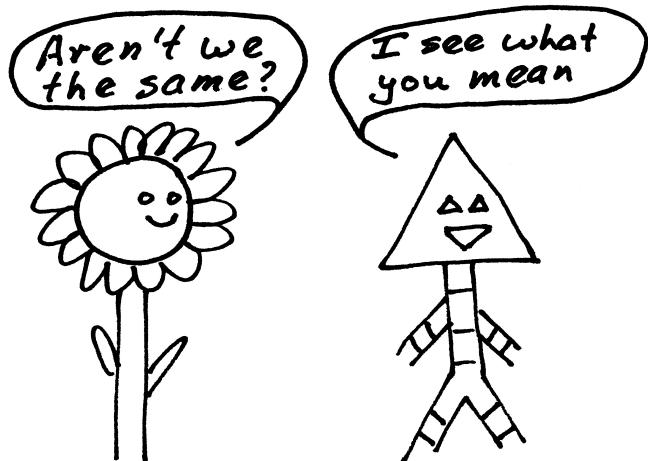
Proof. Apply Theorem 43. \square

Exercises

1. (Hard) Convert the equality of Roman numerals into a rewriting relation such that the normal forms consist of *Is* only, and reduction satisfies the Diamond Property.
2. (Very Hard) Convert the equality of Roman numerals into a rewriting relation such that the normal forms are the usual representations of numbers as Roman numerals only, and reduction satisfies the Diamond Property.
3. Verify all of the lemmas in this chapter, e.g. by running the proofs in Coq.
4. Reprise all of the results from Part I using rewriting instead of equational reasoning.

Chapter 8

Incompleteness of Combinatory Logic



Abstract

Combinatory logic emerged from a critique of predicate calculus by Moses Schönfinkel but was recast as a model of computation by Haskell Curry, who emphasised the correspondence with the λ -calculus of Alonzo Church. Although combinatory logic is both simple and expressive, it has little power of reflection, being unable to even decide equality of its programs. It is easy to embed combinatory logic into tree calculus but there is no meaningful translation in the other direction.

8.1 Logic Without Variables

Combinatory logic emerged from a critique of predicate calculus by Moses Schönfinkel but was recast as a model of computation by Haskell Curry, who emphasised the correspondence with the λ -calculus of Alonzo Church. Although combinatory logic is both simple and expressive, it does not support intensionality, being unable to even decide equality of its programs. It is easy to embed combinatory logic into tree calculus but there is no meaningful translation in the other direction.

Propositional logic is concerned with the validity of arguments such as

$$\frac{p \ \& \ q}{p}$$

where p and q are variables that represent propositions, and the argument asserts that from the conjunction $p \& q$ we can infer the proposition p . Predicate logic is concerned with the validity of arguments such as

$$\frac{\forall x.P(x) \ \& \ Q(x)}{P(x)}$$

where P and Q are variables that represent predicates, so that $P(x)$ is a proposition for any x , and \forall is the universal quantifier that binds x , so that $\forall x.P(x)$ is true if $P(a)$ is true for all values a that x might take.

Unfortunately, the naive use of quantifiers leads to paradoxes, notably Russell's Paradox. One solution is to use *types* to constrain the possible values a that x might take. Schönfinkel's solution was more radical. In the example above, treat P as a function (from x to $P(x)$) and replace the derivation above by

$$\frac{P \wedge Q}{P}$$

where \wedge is a conjunction of functions. Now P and Q are functions (acting on a, b etc.) and \wedge is a function of the functions P and Q that produces another function $P \wedge Q$. That is, Schönfinkel built the first *calculus of functions* in which functions can be used as arguments and returned as results of other functions, so that functions are *first-class citizens*. His goal was a calculus, built from some functions constants such as \wedge , that could capture the expressive power of predicate calculus without binding any variables.

8.2 SK-Calculus

As noted in Chapter 4.1, there are a few options when choosing the operators underpinning combinatory logic. Most common is to use S, K and I , as these appear in the definitions of bracket abstraction and star abstraction, but S and K will suffice. We will continue with all of the terminology of Section 4.1 for combinations and terms built from some class O of operators, and that for rewriting of Chapter 7.2.

The operators S and K of SK -calculus have reduction rules

$$\begin{aligned} Kxy &> x \\ Sxyz &> xz(yz) . \end{aligned}$$

Now I can be defined to be SKK since

$$SKKx \longrightarrow Kx(Kx) \longrightarrow x .$$

Indeed, any combinator of the form SKy is an identity function since $Kx(yx) \longrightarrow x$ for any y .

Theorem 45 (confluence_SK) *Reduction of SK -calculus is confluent.*

Proof. The proof uses the same techniques as that for tree calculus in Theorem 33, by showing that simultaneous reduction satisfies the diamond property. \square

Just as in tree calculus, we can define the programs in a rather simple manner, by the BNF

$$p, q ::= S \mid Sp \mid Spq \mid K \mid Kp .$$

That these are exactly the irreducible combinations, or normal forms, follows from a simple proof by induction on the structure of the combinations.

The reduction rules show that K is used to delete arguments, and S is used to copy them. It turns out that deleting and copying are all that you need to represent any other combinator. In turn, this expressive power will prove to be enough to support all of the extensional functions in Chapter 4. Since the details are not quite the same in SK -calculus as in tree calculus, some redevelopment will be necessary.

8.3 Combinators in SK -Calculus

Every combinator M as in Equation (4.1) can be represented by an SK -combination denoted by

$$M = [x_1][x_2] \dots [x_n]t$$

where the *bracket abstraction* $[x]t$ represents the function that, when applied to x yields t . It is defined as follows:

$$\begin{aligned} [x]x &= I \\ [x]y &= Ky \quad (y \neq x) \\ [x]O &= KO \quad (O \text{ is an operator}) \\ [x]uv &= S([x]u)([x]v) . \end{aligned}$$

The main difference from Chapter 4.2 is that $S([x]u)([x]v)$ has been used instead of $\Delta(\Delta([x]v)([x]u))$. Recall that $\{u/x\}t$ is the result of substituting the term u for the variable x in the term t .

Hence the combinator M in Equation (4.1) is represented by the SK -combination $[x_1][x_2]\dots[x_n]t$.

Since all combinators can now be represented in SK -calculus, we say that SK -calculus is *combinatorially complete*. Note that this does not imply that Δ can be defined in SK -calculus, as it was not introduced as a combinator.

As before, we can optimise bracket abstraction by applying some lightweight analysis. Define $x \in t$ by induction on the structure of t just as in Chapter 4.2. Then define *star abstraction* as follows

$$\begin{aligned}\lambda^*x.t &= Kt && (x \notin t) \\ \lambda^*x.tx &= t && (x \notin t) \\ \lambda^*x.x &= I \\ \lambda^*x.y &= Ky && (y \neq x) \\ \lambda^*x.tu &= S(\lambda^*x.t)(\lambda^*x.u) && (\text{otherwise}).\end{aligned}$$

Once again, the only change is to use of S when abstracting from applications.

The differences between bracket abstraction and star abstraction can be highlighted by considering the combinators that correspond to S and K :

$$\begin{aligned}K_0 &= [x][y]x \\ &= [x]Kx \\ &= S(KK)I \\ S_0 &= [x][y][z]xz(yz) \\ &= [x][y]S([z]xz)([z]yz) \\ &= [x][y]S(S(Kx)I)(S(Ky)I) \\ &= \dots\end{aligned}$$

By contrast, we have

$$\begin{aligned}\lambda^*x.\lambda^*y.x &= \lambda^*x.Kx = K \\ \lambda^*x.\lambda^*y.\lambda^*z.xz(yz) &= \lambda^*x.\lambda^*y.Sxy = \lambda^*x.Sx = S\end{aligned}$$

by which the original operators are recovered.

8.4 Incompleteness of Combinatory Logic

Armed with the techniques of rewriting theory, we can now prove that equality of programs is not definable in combinatory logic. The standard definition of an equality term in SK -calculus is a term eq such that $\text{eq } M N$ is true or false according to whether the programs M and N are equal or not. Further, let us try and capture the meaning of the booleans in terms of reduction rules, such as $\text{true } x\ y \longrightarrow x$. Then preservation of the variables and reduction rules will ensure the preservation of truth, too.

Define an equality term in SK -calculus to a term eq such that, for all variables x and y we have:

- $\text{eq } M\ M\ x\ y \longrightarrow^* x$ for all programs M

- $\text{eq } M \ N \ x \ y \longrightarrow^* y$ for all programs M and N such that $M \not\equiv N$.

Note that we could have written $M \neq N$ above since structural equality and computational equality are the same for programs.

Our goal is to show that there is no equality term in SK -calculus. The basic idea is that SK -calculus cannot distinguish combinators that have the same functional behaviour, such as the combinators SKK and SKS that both represent the identity function. To formalise this requires some new rewriting machinery.

Define an *identity program* to be a program M such that $Mx \longrightarrow^* x$ for some variable x (the choice of x is immaterial).

Define a *separator* for programs M_1 and M_2 to be a program s such that there are distinct variables x and y for which

$$\begin{aligned} s \ M_1 \ x \ y &\longrightarrow x \\ s \ M_2 \ x \ y &\longrightarrow y. \end{aligned}$$

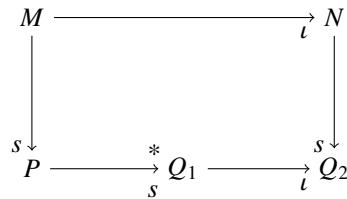
We will show that there are identity programs that cannot be separated in SK -calculus, and so program equality is not definable.

Define the *identity reduction* \longrightarrow_i to be the congruence generated by the rule

$$M \longrightarrow_i I \quad (\text{if } M \text{ is an identity program}).$$

For our purposes, it is enough to show that identity reduction interacts well with the simultaneous reduction relation \longrightarrow_s that was developed in the proof of confluence.

Lemma 46 (pentagon_id_red_s_red1) *For all terms M, N and P and reductions $M \longrightarrow_i N$ and $M \longrightarrow_s P$ there are terms Q_1 and Q_2 such that we have reductions $P \longrightarrow_s^* Q_1 \longrightarrow_i Q_2$ and $N \longrightarrow_s Q_2$ as illustrated in the following pentagon:*

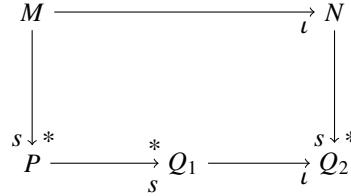


Proof. The pentagon is like the diamond property for \longrightarrow_s and \longrightarrow_i , except that it has been padded with some additional simultaneous reductions. The proof requires some careful case analysis of both sorts of reduction, but is not very illuminating. Details can be found in the Coq verification. \square

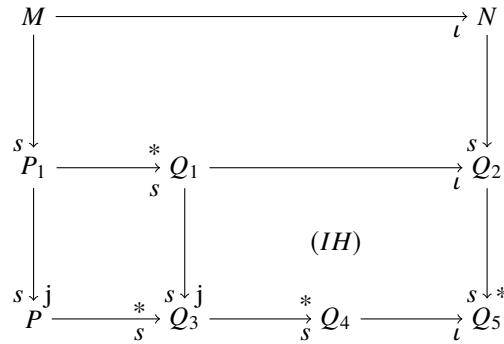
The pentagons can be generalised to handle multiple parallel reductions as follows.

Lemma 47 (pentagon_id_red_s_red) *For all terms M, N and P and for all reductions $M \longrightarrow_i N$ and $M \longrightarrow_s^* P$ there are terms Q_1 and Q_2 such that we have reductions*

$P \xrightarrow{s}^* Q_1 \xrightarrow{\iota} Q_2$ and $N \xrightarrow{s}^* Q_2$ as illustrated in the following pentagon:



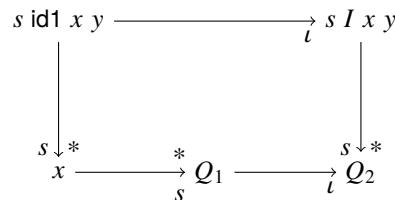
Proof. The proof is by induction on the length k of the parallel reduction sequence $M \xrightarrow{s}^k P$. If k is zero then the result is immediate. If k is $j+1$ then apply the following diagram:



The upper pentagon exists by the previous lemma. The square exists by the diamond property for simultaneous reduction. It follows that the number of reduction steps from Q_1 to Q_3 is the same as the number of reductions from P_1 to P . Hence, the lower pentagon exists by the induction hypothesis. Thus Q_4 and Q_5 are the required terms, since there are simultaneous reductions from $P \xrightarrow{s}^* Q_4$ and $N \xrightarrow{s}^* Q_5$. \square

Theorem 48 (no_separable_identities_in_SK) *Identity programs do not have separators in SK-calculus.*

Proof. Let id1 and id2 be distinct identity programs and assume that they have a separator s using distinct variables x and y . Apply the previous lemma to the identity reduction from id1 to I to get



Now Q_1 must be x since variables are irreducible, and Q_2 must be x since x is not a program. Thus $sIxy \xrightarrow{s}^* x$. By a similar argument using id2 , we have $sIxy \xrightarrow{s}^* y$.

Hence x and y share a reduct, by confluence. Hence they are equal, which contradicts the original assumption. \square

Theorem 49 (equality_of_programs_is_not_definable_in_SK)

There is no equality term in SK-calculus.

Proof. If eq were equality of programs then $\text{eq}(SKK)$ would be a separator for SKK and SKS which contradicts Theorem 48. \square

Since equality of programs is an effectively calculable function, it follows that *SK*-calculus is *not complete* for computation of its own programs. It is worth taking a moment to consider the scope and implications of this theorem. Although it does not limit the ability of *SK*-calculus to compute numerical functions, it does limit its ability to analyse them. More precisely, we have seen that all of the μ -recursive functions can be represented by extensional programs in tree calculus, and this representation can be repeated in *SK*-calculus, but there is little support for their optimisation. For example, suppose that program p can be replaced by program q whose computation takes fewer steps. In order to take advantage of this in a large program f one must search f for copies of p and replace them by q . However, there is no *SK*-combinator that can decide whether some sub-program p' is p or not. By contrast, within tree calculus there are combinations that can decide equality of p and p' , that can search f for occurrences of p and replace them by q . Thus, tree calculus is a far more promising setting for program analysis and optimisation than combinatory logic.

8.5 Meaningful Translation

Before considering translations between combinatory logic and tree calculus, let us consider what it means for a translation $[-]$ between such calculi to be meaningful, in the sense that the meaning of the terms is preserved.

First of all, computational equality must be preserved, so that if $s = t$ then $[s] = [t]$. Here and in all subsequent requirements, the equalities will be computational equalities.

Second, translation should preserve application, so that

$$[t u] = [t][u].$$

Third, values must be preserved. That is, the translation of a program must reduce to a program. Otherwise, successful evaluations in one calculus may be translated to failed computations in the other. For example, suppose that tree calculus is augmented with a fixpoint function operator Z with the rule

$$Z f x = f(Zf)x.$$

and consider a translation back to tree calculus that preserves Δ . If $[Z] = Y$ then the value $Z\Delta$ will be translated by $Y\Delta$ which has no value. Meaningful translation must use some variant of Y_2 .

Finally, variables of the term calculus must be preserved, so that $[x] = x$. This last point is subtle, since the variables are not programs. However, it is necessary to preserve

the meaning of the reduction rules. For example, consider a variant of the arithmetic expressions that includes an error term `error` which blocks any further reduction. Its rules for adding zero become

$$\begin{aligned}\text{zero} + \text{zero} &= \text{zero} \\ \text{zero} + \text{successor } n &= \text{successor } n\end{aligned}$$

but `zero` + `error` does *not* reduce to `error`. Is the embedding of the arithmetic expressions into this larger calculus meaningful? Certainly, it preserves all reductions of combinations but it does not preserve the reduction of terms, since `zero` + `n` does *not* reduce to `n` for any variable `n`.

Another way of looking at it is that the rule

$$\text{zero} + n = n$$

is an assertion about all terms `n` no matter what their nature might be. So, to preserve the meaning requires that this quantification over all possible terms `n` should be preserved. By insisting that variables be preserved, we require that this equation hold in the new calculus, which in turn requires that `zero` + `error` = `error` hold for translation to be meaningful.

We can re-express the four requirements above in the language of rewriting as follows.

Definition 50 A meaningful translation between applicative rewriting systems is a function of their terms that satisfies the following criteria:

1. If $s = t$ then $[s] = [t]$.
2. $[tu] = [t][u]$ for all terms t and u .
3. If v is a value then $[v]$ is valuable.
4. $[x] = x$ for all variables x .

An example of a meaningful translation is the translation of Roman numerals to unary arithmetic in Chapter 2.9. Here is another.

The translation of combinatory logic to tree calculus is deceptively simple, being determined by

$$\begin{aligned}[S] &= S \\ [K] &= K\end{aligned}$$

with variables and applications being preserved.

Theorem 51 (meaningful_translation_from_sk_to_tree) There is some meaningful translation from SK-calculus to tree calculus.

Proof. Variables, closed terms and applications are preserved by definition. That one-step reductions of SK-calculus are mapped to multi-step reductions of tree calculus was established in Chapter 3.3. Further, those proofs of preservation also show that partial applications of (the translations of) `S` and `K` have normal forms if their arguments do. In particular, values are translated to valuable terms. \square

8.6 No Tree Calculus in Combinatory Logic

There is no meaningful translation from tree calculus to *SK*-calculus since tree calculus supports intensionality in ways that *SK*-calculus does not.

Lemma 52 (meaningful_translation_preserves_identity_programs) *If f is a meaningful translation from *SK*-calculus to tree calculus then f preserves identity programs.*

Proof. Routine calculation. \square

Lemma 53 (meaningful_translation_preserves_separators) *If f is any meaningful translation from *SK*-calculus to tree calculus then f preserves separators.*

Proof. Routine calculation. \square

Theorem 54 (no_translation_tree_to_sk) *There is no meaningful translation from tree calculus to *SK*-calculus.*

Proof. Tree calculus supports the program identities $\text{tag}\{K, I\}$ and $\text{tag}\{KI, I\}$ which are separated by `getTag`. If there were such a meaningful translation to *SK*-calculus then it would preserve the program identities by Lemma 52 and also their separator, by Lemma 53. Hence, *SK*-calculus would support separable identity programs but this contradicts Theorem 48 \square

It follows that tree calculus is strictly more expressive than combinatory logic, since there is a meaningful translation from combinators to trees but not in the other direction. In particular, there are functions of programs that are not definable in combinatory logic but are definable in tree calculus, such as the separator of *SKK* and *SKS*.

Exercises

1. Define B from Equation (4.2) as a tree.

2. Define C from Equation (4.2) as a tree.

3. Conjunction is given by

$$\text{conj} = BB(KI) .$$

Check the truth table for conjunction.

4. The logical equivalence iff is given by

$$\text{iff} = BS(B(BI(KI))K) .$$

Check its truth table.

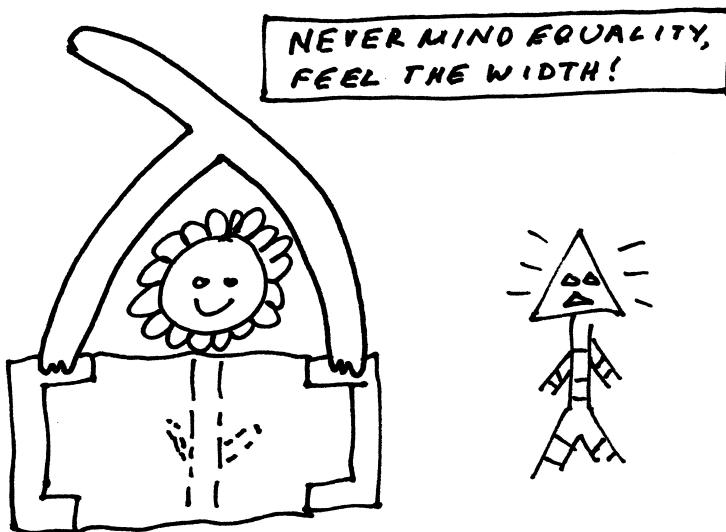
5. Prove that bracket abstraction and star abstraction have the expected properties.

6. Compute the *SK*-combinator of $[f][x][y]fyx$ that corresponds to B in Equation (4.2). Do the same for C in Equation (4.2).

7. Confirm that `getTag0` is a separator for $\text{tag}\{K, I\}$ and $\text{tag}\{KI, I\}$ in tree calculus.
8. Define a separator of SKK and SKS in tree calculus.
9. Verify Lemma 52.
10. Verify Lemma 53
11. Show that the translation from Roman numerals to unary arithmetic in Section 2.9 is meaningful.

Chapter 9

Lambda-Abstraction in VA-Calculus



Abstract

VA-calculus builds variables and abstractions as combinations, whose evaluation rules do not make any use of the side conditions or meta-theory for maintaining scope. Like tree calculus, it supports tagging, so that it supports some intensionality. In particular, tagging supports meaningful translations to and from tree calculus. Even so, VA-calculus is not complete as it cannot define program equality.

9.1 First-Class Abstraction

We have seen the convenience of being able to convert a rule

$$x \mapsto M$$

into an abstraction with the property that applying it to some N results in the substitution of N for x in M . The fundamental idea of λ -calculus is to explore the power of abstraction as the primary means of term formation, as a first-class term construction, and not merely as a syntactic convenience, in the sense that abstraction preserves computational equality as well as syntactic equality. The various rules for operators are replaced by a single rule, the β -rule

$$(\lambda x.M)N = \{N/x\}M . \quad (\beta)$$

which in turn requires rules for performing the substitution of N for x in M . Traditionally, the substitution rules are *not* reduction rules of the λ -calculus, but part of its *meta-theory* which must consider variables and substitutions, as well as terms. When M is some combination of variables then substitution is straightforward but substitution into abstractions is not. This chapter will produce a new variant of λ -calculus called *VA-calculus* built using operators V and A for variables and abstractions, whose evaluation is given by reduction rules that do not require any meta-theory, much as in tree calculus. In particular, it follows that its abstractions are first-class.

Although VA-calculus meets the criterion for being a λ -calculus, in that it supports the abstractions that define the combinators, there cannot be meaningful translations between traditional λ -calculi and VA-calculus because they must take different approaches to substitution into redexes. When substitution is meta-theoretic, it must apply to any term, which implies

$$\{N/x\}(MP) = \{N/x\}M \{N/x\}P$$

even if MP is a redex. However, if substitution is part of the theory proper then such a reduction would overlap with other rules in a way that precludes any meaningful translation. In VA-calculus, substitution rules never apply to redexes, but to partially applied operators only. One might object that this change is too radical, but the precise mechanics of substitution is not part of the general understanding of abstraction, not part of its specification, so VA-calculus qualifies as a λ -calculus.

When considering the difficulties of λ -calculus, substitution into applications is not usually a concern. Rather, it is *scope management* which gets all of the attention, especially as it does not seem to support a combinatory representation. Significant difficulties arise when substituting into an abstraction, as in

$$(\lambda x.\lambda y.M)N = \{N/x\}(\lambda y.M) .$$

This requires a proper understanding of scope. In particular, what is the meaning of $\{N/x\}(\lambda y.M)$ when y is x ? Again, what is the meaning of $\{y/x\}(\lambda y.M)$ when M is y (or contains y)? If this was not immediately obvious to you then you are in good company; the substitution rules are not obvious, are not axiomatic.

The traditional solution to such a name clash in $\{N/x\}(\lambda y.M)$ is to rename y by some variable z that is *fresh*, i.e. not already in use. For example, before substituting N for x in $\lambda x.x$ we should rename the bound name x by some fresh name z that does not appear in N so that we substitute of N for x in $\lambda z.z$ to get $\lambda z.z$. This solution is a little cumbersome: as well as adding equations for the renaming, there is the challenge of finding fresh variables in a systematic way.

Nicolaas Govert de Bruijn observed that the variables can be represented by indices $0, 1, 2, \dots$ now called *de Bruijn indices*, where the index represents (one less than) the number of abstractions required to create a binding. For example, the term K given by $\lambda x.\lambda y.x$ becomes $\lambda(\lambda 1)$ since the x was bound by the second enclosing abstraction. When applied to K itself then β -reduction replaces 0 by K in $\lambda 1$. Then substitution under the λ causes 1 to be replaced by K in the body, to produce λK . Things are more complicated when the argument is a variable. For example, when K is applied to 3 then substitution of 3 for 0 in $\lambda 1$ replaces 1 by 4 to get $\lambda 4$. Of course, all this lifting and lowering of indices uses meta-theory, which we are trying to avoid.

In VA-calculus, substitution does not act on the body of an abstraction. Instead, abstractions carry an *environment* in which variable bindings can be recorded. The abstraction

$$AME$$

has body M and environment E . The reduction rules for A are guided by the intuition that reduction of $AMEN$ applies the “substitution” A_EN to M by analysing the structure of M . For example, if M is the zero index V then $AVEN = N$ or if M is a successor index Vk then $A(Vk)EN = Ek$ where the environment E is here thought of as a substitution. Most interesting is when M is itself an abstraction, in which case we have

$$A(APQ)EN = AP(AQEN).$$

That is, substitution is applied to the environment Q but not the body P of the abstraction APQ . In this manner, the various meta-rules for substitution are expressed as reduction rules for abstraction. Of course, there remains the challenge of creating suitable environments, of creating combinations that represent substitutions.

When using de Bruijn indices, a substitution may be thought of as a list, whose i th entry holds the value of the i th index. The temptation now is to add more operators to represent these lists, but instead we can use A to represent the empty environment, and $A(AE)N$ to add another binding to the environment. Since the empty environment is stable under substitution, we have the rule

$$AAEN = A$$

Similarly, to apply the substitution $A(AE)N$ to some M yields the rule

$$A(AE)NM = AMEN .$$

When this rule is read from right to left, it expresses the β -rule, that applications of abstractions yield substitutions. However, as a reduction rule, it will be read from left to right, so that substitutions may query the body of the closure.

The equality of programs in VA-calculus can be decided by:

- translating the programs to tree calculus;
- deciding equality of the translations using `equal`; and
- translating the result back to VA-calculus.

Figure 9.1: An Invalid Argument

Environments first appeared in the *explicit substitution calculus* of Martine Abadi, Luca Cardelli, Pierre-Louis Curien and Jean-Jacques Lèvy but they were not introduced to solve this problem, so that substitutions have continued to act on the bodies of abstractions, with no reduction in the amount of meta-theory. Also, when a substitution σ acts on environment that is a substitution ρ it is usual to compose them, as functions from variables to terms. In particular, if the environment ρ is empty then their composition is σ . Here, however, the application of a substitution σ to an empty environment yields an empty environment. While composition is more intuitive than application, it usually creates overlapping reduction rules which is unpleasant. This is doubly annoying when σ has no effect on M anyway. In practice, this constraint is often enforced by the use of *closures*. However, checking for closures during computation would be awkward, and is not strictly necessary, so will not be enforced within the calculus itself.

Like SK-calculus, VA-calculus is combinatorially complete without being complete. That is, it supports all combinators and yet does not support equality of programs.

Also, there is a meaningful translation $[-]$ from VA-calculus to tree calculus. For example, it maps A to A_t , which is pictured at the beginning of Part II. It is easy enough to define A_t to capture the desired behaviour when it is applied to three arguments. The challenge is to be able to substitute into its applications to zero, one or two arguments. The solution is to use tags in tree calculus to record the information about the arguments of A_t .

Interestingly, even though VA-calculus is not complete, it does support tagging, since environments are able to store tags, which can be recovered by manipulating scopes in a non-standard manner. In turn, this can be used to define a translation from tree calculus to VA-calculus.

This shows that meaningful translation need not preserve completeness properties. In particular, the translations cannot be used to decide equality of programs. Even though equality of programs is not definable in VA-calculus, it is tempting to adopt the invalid argument in Figure 9.1.

The paradox is resolved as follows. Using translations $[-]$, the equality of programs p and q is specified by the computation of

$$[\text{equal}] [[p]] [[q]].$$

If the double translation from VA-calculus to tree calculus and back were definable in VA-calculus then this specification could be realised in VA-calculus. Conversely, since equality cannot be defined in VA-calculus, it follows that the double translation is *not* definable in VA-calculus. The double translation is performing computations which are beyond the expressive power of VA-calculus.

$$\begin{aligned}
(x@y)z &= (x@y)@z && \text{(suspension)} \\
(\lambda 0y)z &= z && \text{(zero)} \\
(\lambda(Vx)y)z &= yx && \text{(successor)} \\
(\lambda(w@x)y)z &= (\lambda wy)z((\lambda xy)z) && \text{(application)} \\
(\lambda[]y)z &= [] && \text{(empty)} \\
(\lambda[x])yz &= (\lambda zx)y && \text{(substitution)} \\
(\lambda(\lambda wx)y)z &= \lambda w((\lambda xy)z) && \text{(abstraction)}
\end{aligned}$$

Figure 9.2: Informal Rules of VA-Calculus

9.2 VA-Calculus

The term forms of VA-calculus are given by

$$M, N := V \mid A \mid MN .$$

Both V and A are ternary operators. V is used to build indices, with the k th index given by $V^k(V)$. We may write k for $V^k(V)$ when it is clear that k represents an index. A term of the form Vxy is a *suspended application*, which may also be written as $x@y$. They convert to ordinary applications under substitution. Some care is required here since applications of variables must be suspended. For example, to apply the first index VV to some argument x do not use VVx as this is the suspended application of $V@x$. Rather, use the suspended application $V(VV)x$ or $VV@x$. The *abstraction* Axy has *body* x and *environment* y . It may also be written λxy . Note that x is *not* the index, but the body. The operator A can be used as the empty environment, as it is stable under substitution. It may be written as $[]$. The partial application Ax will be used to support substitutions of the form $A(Ax)y$. The term Ax may also be written as $[x]$.

Armed with this notation, the evaluation rules of the calculus may be written informally in Figure 9.2. While this helps to illustrate the multiple roles of the two operators, it is hard to check that the rules cover all cases without overlapping. This is easy to check when the rules are given formally, without any syntactic sugar, in Figure 9.3. It follows that the programs are exactly the binary tree with nodes labelled by V or A , and that reduction is confluent.

Theorem 55 (confluence_va_calculus) VA-calculus is confluent.

$$\begin{aligned}
Vxyz &\longrightarrow V(Vxy)z && \text{(suspension)} \\
AVyz &\longrightarrow z && \text{(zero)} \\
A(Vx)yz &\longrightarrow yx && \text{(successor)} \\
A(Vwx)yz &\longrightarrow Awyz(Axyz) && \text{(application)} \\
AAyz &\longrightarrow A && \text{(empty)} \\
A(Ax)yz &\longrightarrow Azxy && \text{(substitution)} \\
A(Awx)yz &\longrightarrow Aw(Axyz) && \text{(abstraction)}
\end{aligned}$$

Figure 9.3: Reduction Rules of VA-Calculus

Proof. The proof uses simultaneous reduction, just as for *SK*-calculus. None of the reduction rules overlap. \square

The *values* or *programs* of VA-calculus are given by the applications of V or A to zero, one or two programs. Once again, the *valuable* terms are those that reduce to values.

9.3 Combinators

Bracket abstraction and star abstraction are defined by adapting techniques from tree calculus. It may seem strange to do this within a λ -calculus but it will prove convenient when the goal is to substitute into the body of an abstraction.

The identity combination is defined by $\lambda 0[]$ or

$$I = AVA$$

since $Ix = AVAx = x$. Also, we can define *bracket abstraction* by

$$\begin{aligned} [x]_b x &= V \\ [x]_b y &= Vy \quad (y \neq x) \\ [x]_b V &= VV \\ [x]_b A &= VA \\ [x]_b st &= ([x]_b s)@([x]_b t) \end{aligned}$$

$$[x]t = A([x]_b t)I .$$

Theorem 56 (bracket beta) For all terms t and u we have $([x]t)u \longrightarrow \{u/x\}t$.

Proof. We have $([x]t)u = A([x]_b t)Iu$ and now proceed by induction on the structure of t . \square

As in tree calculus, bracket abstraction can be used to support waiting and fixpoint functions in normal form. Again, we can optimise bracket abstraction to produce *star abstraction* defined by

$$\begin{aligned} \lambda_b^* x.x &= V \\ \lambda_b^* x.y &= Vy \quad (y \neq x) \\ \lambda_b^* x.V &= VV \\ \lambda_b^* x.A &= VA \\ \lambda_b^* x.st &= V(s\ t) \quad (\text{if } x \notin s\ t) \\ \lambda_b^* x.st &= (\lambda_b^* x.s)@(\lambda_b^* x.t) \quad (\text{if } x \in s\ t) \\ \\ \lambda^* x.t &= A(\lambda_b^* x.t)I . \end{aligned}$$

Theorem 57 (star_beta) *For all terms t and u we have $(\lambda^* x.t)u \rightarrow \{u/x\}t$.*

Proof. Apply induction on the structure of t . \square

Hence, we may define K and S in the usual manner by

$$\begin{aligned} K &= \lambda^* x. \lambda^* y. x \\ S &= \lambda^* x. \lambda^* y. \lambda^* z. xz(yz) . \end{aligned}$$

Theorem 58 (meaningful_translation_from_sk_to_va) *There is a meaningful translation from SK-calculus to VA-calculus.*

Proof. With S and K defined as above, it remains to check that their partial applications to programs have normal forms. \square

It follows that VA-calculus supports all the extensional functions introduced in Chapter 4. However, it often happens that there are more direct accounts that use non-empty environments to represent substitutions.

For example, *waiting* can be achieved by using an environment to delay application, in

$$\text{wait}\{x, y\} = A(21@0)[y, x, I] .$$

Clearly, this term is a program if x and y are, while application to some z yields xyz .

Conversely, if a λ -calculus does not support waiting then there is no reason to expect a meaningful translation to it from SK-calculus. For example, if the SK-program

$$\text{wait}\{SII, SII\}$$

has the same translation as $(SII)(SII)$ then the translation of the latter must be valuable, which would be a surprise.

9.4 Incompleteness

VA-calculus is incomplete for similar reasons to SK-calculus.

Theorem 59 (equality_of_programs_is_not_definable_in_va) *The equality of normal forms is not definable in VA-calculus.*

Proof. The proof is just as for SK-calculus but with the identity programs I and AVI instead of SKK and SKS . That is, one must define the identity-reduction relation, prove the pentagon lemmas, and then show that the existence of an equality term implies that distinct variables are equal. Details are in the formal proofs. \square

9.5 Translation to Tree Calculus

Let us consider how to translate VA-calculus to tree calculus. In order to capture the intensionality of substitution into, say, an abstraction $A z w$ it is necessary to be able to recover the translations of A and z and w from that of their combination, and this without interfering with the functionality of the abstraction. The solution is to tag the functional translation of the abstraction with the information required for substitution. Since this must include the translation of A , we can specify A_t by

$$A_t x y z = \text{getTag } x A_t y z$$

The required tags are determined by the six reduction rules for A so define

$$\begin{aligned} \text{zero_rule} &= \lambda^* a. \lambda^* y. \lambda^* z. z \\ \text{successor_rule} &= \lambda^* x. \lambda^* a. \lambda^* y. \lambda^* z. yx \\ \text{application_rule} &= \lambda^* w. \lambda^* x. \lambda^* a. \lambda^* y. \lambda^* z. yx \\ \text{empty_rule} &= \lambda^* a. \lambda^* y. \lambda^* z. a \\ \text{substitution_rule} &= \lambda^* x. \lambda^* a. \lambda^* y. \lambda^* z. azxy \\ \text{abstraction_rule} &= \lambda^* w. \lambda^* x. \lambda^* a. \lambda^* y. \lambda^* z. aw(axyz) . \end{aligned}$$

Then the operators are translated by

$$\begin{aligned} V_t = Y_{2t} \{ &\text{zero_rule}, \\ &\lambda^* x. \lambda^* a. \text{tag}\{\text{successor_rule } x, \\ &\quad \lambda^* y. \text{tag}\{\text{application_rule } x y, \\ &\quad \quad [z](a(axy)z)\}\}\} \end{aligned}$$

and

$$\begin{aligned} A_t = Y_{2t} \{ &\text{empty_rule}, \\ &\lambda^* x. \lambda^* a. \text{tag}\{\text{substitution_rule } x, \\ &\quad \lambda^* y. \text{tag}\{\text{abstraction_rule } x y, \\ &\quad \quad \text{getTag } x a y\}\}\} . \end{aligned}$$

Note the use of bracket abstraction $[z](a(axy)z)$ in the body of V_t . This is used to avoid premature reduction of xy in the translation of suspended applications. The picture of V_t is given in Figure 9.4. The picture of A_t is given at the beginning of Part II.

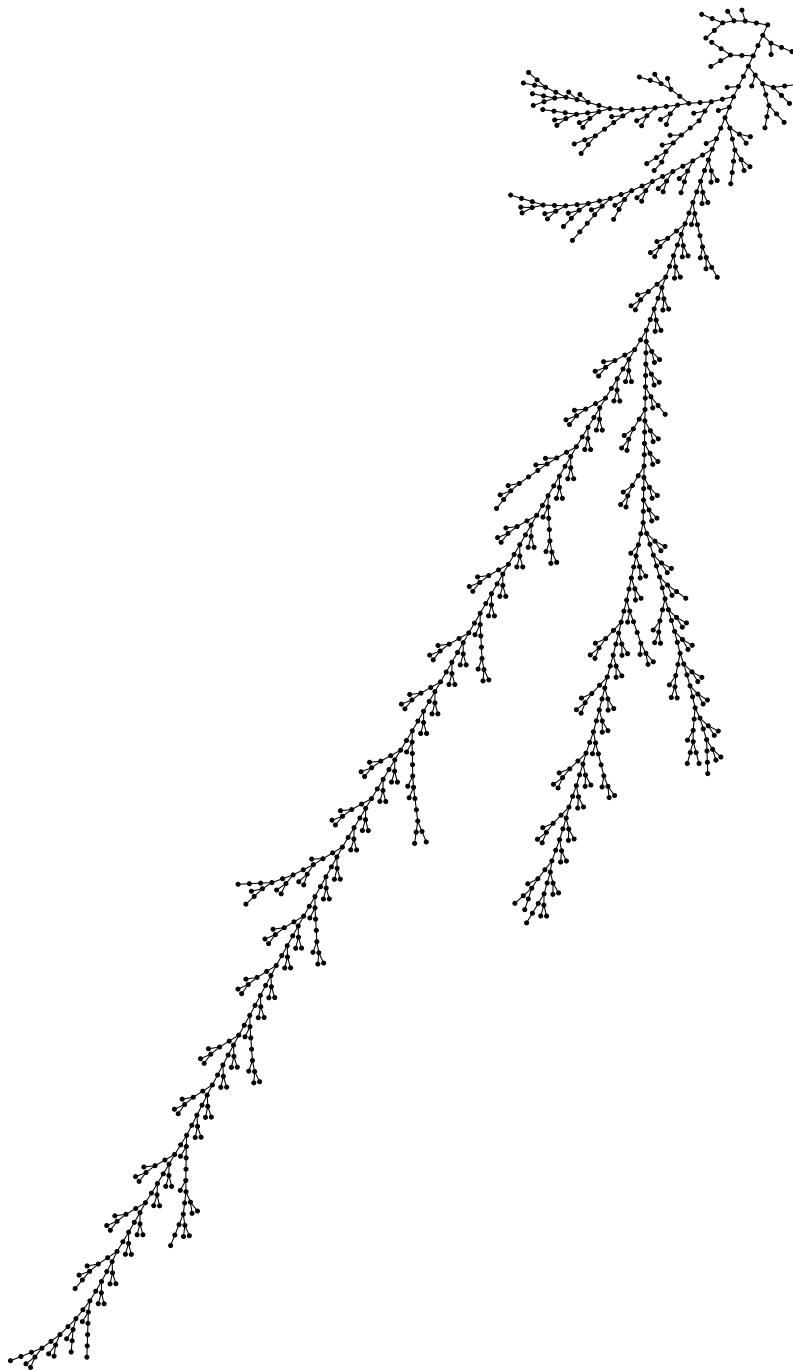
Theorem 60 (meaningful_translation_va_to_tree) *There is a meaningful translation from VA-calculus to tree calculus.*

Proof. The translation is given above. The proof requires many lemmas, which can be found in the Coq implementation. \square

9.6 Tagging

Even though VA-calculus is incomplete, it is still able to support tagging, which is a key aspect of intensional computation. Define

$$\begin{aligned} \text{tag}\{t, f\} &= A(V(V(V21)(V(V^3(f))V))(V^3(t))) (A(A[[K, I], V, I])) \\ \text{getTag} &= A(V(V(V(V(VA)V)(VI))(VKI))(VA))I. \end{aligned}$$

Figure 9.4: The Translation of V to Tree Calculus

so that

$$\begin{aligned}\text{tag}\{t, f\} x &= [K, I]0(fx)t = K(fx)t = fx \\ \text{getTag tag}\{t, f\} &= [K, I](KI)(fx)t = KI(fx)t = t\end{aligned}$$

as required.

Hidden in the blur of symbols is the following idea. By controlling the order in which substitutions are applied, the meaning of $2@1$ in the definition of $\text{tag}\{t, f\}$ can become either K or KI to yield either the application of f or the term t . In turn, this is possible because of the flexibility of the term formation rules: substitutions are first-class terms that can be used to define other substitutions, and V has been applied to f and t without requiring them to be indices.

9.7 Translation from Tree Calculus

As with the translation of VA-calculus to tree calculus, tagging can be used to support a meaningful translation from tree calculus to VA-calculus, as follows. The action of Δ on a leaf is represented by K since $Kyz \rightarrow y$. The actions of Δ on stems and forks requires combinations $\text{stem}\{x\}$ and $\text{fork}\{x, y\}$ such that

$$\begin{aligned}\text{stem}\{x\} y z &\rightarrow yz(xz) \\ \text{fork}\{w, x\} y z &\rightarrow zwx.\end{aligned}$$

These are given by star abstractions in the standard manner. Finally, we require an interpretation \ker of the node, that satisfies

$$\begin{aligned}\ker x y z &= \text{getTag } x y z \\ \text{getTag } \ker &= K \\ \text{getTag } (\ker x) &= \text{stem}\{x\} \\ \text{getTag } (\ker w x) &= \text{fork}\{w, x\}.\end{aligned}$$

It is given by

$$\begin{aligned}\ker &= \text{tag}\{K, \\ &\quad \lambda^* x. \text{tag}\{\text{stem}\{x\}, \\ &\quad \quad \lambda^* y. \text{tag}\{\text{fork}\{x, y\}, \\ &\quad \quad \quad (A(V(V(V(V \text{getTag})(Vx))(Vy))V)I)\}\}\}\end{aligned}$$

and is pictured in Figure 9.5.

Theorem 61 (meaningful_translation_from_tree_to_va) *There is some meaningful translation from tree calculus to VA-calculus.*

Proof. The proof requires many lemmas. See the Coq proofs for details. \square

Corollary 62 *There is no meaningful translation from VA-calculus to combinatory logic.*

Proof. If there were such a translation then it would compose with the translation above to produce a meaningful translation from tree calculus to combinatory logic, which is impossible by Theorem 54. \square

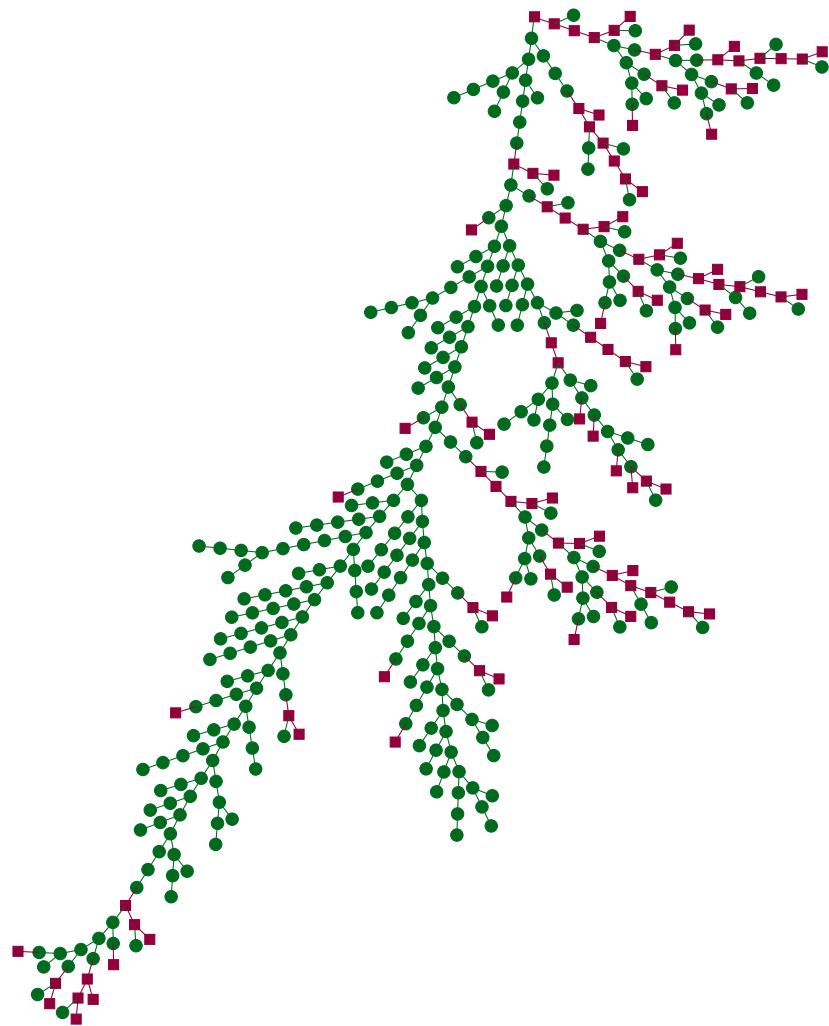


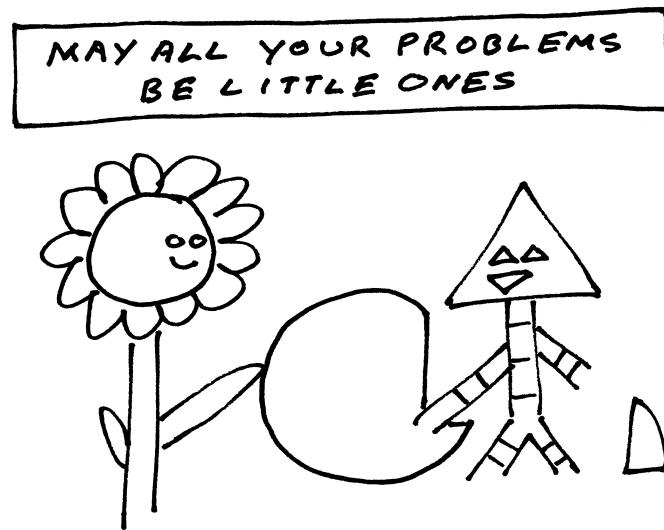
Figure 9.5: The Translation of Δ to VA-Calculus: Circles are V and Squares are A

Exercises

1. (Hard) List the key properties that you think that a lambda-calculus should satisfy, such as being Turing complete, being combinatorially complete, supporting the interpretation of functions as rules, supporting equational reasoning, etc. Try to avoid technical properties that are internal to the calculus. Determine if VA-calculus has these properties. How does VA-calculus compare to your favourite lambda-calculus?
2. (Hard) Is the support for tagging in VA-calculus a bug or a feature? On the one hand, it exploits the dynamic approach to scoping which conflicts with traditional views. On the other hand, it is crucial to the translation from tree calculus.
3. (Hard) VA-calculus sits between combinatory logic and tree calculus: unlike combinatory logic it is intensional, to the extent that is implied by tagging; unlike tree calculus, it does not support reflection as it does not support pattern-matching functions or even program equality. Is this intermediate level of expressiveness interesting in practice?
4. (Hard) Many powerful type systems use abstraction with respect to type variables to support polymorphism. So the VA-calculus can support such types, much as in Chapter 5.5. However, type inference requires the ability to decide type equality, which will not be possible in VA-calculus. How might tree calculus support polymorphic types?

Chapter 10

Divide-and-Conquer in *SF*-Calculus



Abstract

This chapter introduces *SF*-calculus, whose factorisation provides general support for divide-and-conquer algorithms. It is older and a little more complicated than tree calculus but has similar expressive power. There are meaningful translations in both directions.

10.1 Divide and Conquer

Data structures differ from functions in that the emphasis is on storing information rather than using it. This stored information is recovered by queries. More generally, we search data structures, or combine the data to produce bulk properties of the structure. No matter the nature of the structure, these processes can be thought of as applications of a divide-and-conquer strategy, in which the structure is recursively divided into smaller and smaller pieces which are eventually conquered.

Although this is achievable in tree calculus, it is not especially convenient. When evaluating Δxyz , the reduction rules of tree calculus use the two arguments y and z to handle the three possibilities for x , as a leaf, stem or fork, so that y and z must perform double duty: y is used for leaves, z is used for forks, and both are used for stems. More to the point, this approach does not generalise to handle other constructions that use four or more arguments.

This chapter develops a cleaner approach, based on factorisation by an operator F . When evaluating $Fxyz$ there are only two possibilities: if x is an atom (i.e. an operator) then use y else if x is a compound (i.e. a partially applied operator) then use z . That is, F supports divide-and-conquer strategies, in which compounds are divided and atoms are conquered. In a sense, the reduction rules for F correspond to the leaf and fork rules of tree calculus. However, reduction of F does not allow for duplication of any kind, as supported by the stem rule. This can be recovered by admitting the traditional S as a second operator, to get SF -calculus, developed with Thomas Given-Wilson. It turns out that SF -calculus is equivalent to tree calculus in expressive power. In particular, it supports equality of programs and tagging and pattern matching.

10.2 SF-Calculus

The *combinations* of SF -calculus are built using the operators S and F which are the *atoms*. The *compounds* are the applications of S and F to one or two arguments only, which can be thought of as their *partial applications*. Conceptually, there are two reduction rules for F , namely

$$\begin{aligned} Fxyz &\longrightarrow y && (\text{if } x \text{ is an atom}) \\ F(wx)yz &\longrightarrow zwx && (\text{if } wx \text{ is a compound}). \end{aligned}$$

In a more general setting, the atoms may include other operators or constructors for data types. Then the *compounds* are all the applications of atoms that are not redexes. Since S and F are both going to be ternary operators we can *define* the compounds to be combinations of the form Sx, Sxy, Fx and Fxy . Then the *reduction rules* of SF -calculus are formalised in Figure 10.1 by expanding out the side-conditions to the rules. Then the *programs* of SF -calculus are its irreducible terms, which may be thought of as binary trees whose nodes are labelled by S or F .

Theorem 63 *Reduction of SF -calculus is confluent.*

Proof. The proof uses the same techniques as that for tree calculus in Theorem 33, by showing that simultaneous reduction satisfies the diamond property. \square

$$\begin{aligned}
 Sxyz &\longrightarrow xz(yz) \\
 FSyz &\longrightarrow y \\
 FFyz &\longrightarrow y \\
 F(Sx)yz &\longrightarrow zSx \\
 F(Swx)yz &\longrightarrow z(Sw)x \\
 F(Fx)yz &\longrightarrow zFx \\
 F(Fwx)yz &\longrightarrow z(Fw)x .
 \end{aligned}$$

Figure 10.1: Reduction rules of *SF*-calculus

We can define *K* by *FF* since

$$Kyz = FFyz = y$$

as before. That is, *SK*-calculus is embedded within *SF*-calculus, so that all of the machinery of bracket abstraction, star abstraction, fixpoints, arithmetic, etc. carries over with almost no change.

Further, we can test for being a compound by the function

$$\text{isComp} = \lambda^* x. Fx(KI)(K^3 I)$$

and define the component functions *left* and *right* by

$$\begin{aligned}
 \text{left} &= \lambda^* x. FxKK \\
 \text{right} &= \lambda^* x. FxK(KI)
 \end{aligned}$$

since if *wx* is a compound then

$$\begin{aligned}
 \text{left}(wx) &= F(wx)KK = Kw x = w \\
 \text{right}(wx) &= F(wx)K(KI) = KIw x = x .
 \end{aligned}$$

As these examples show, the first argument of *F* is often the last to be known, so that it was tempting to change the order of the arguments of *F*. However, this arrangement makes it easy to define *K*. Again, we could easily replace the operator *S* with *D* to obtain the *DF*-calculus, but let that pass.

As to the converse question, whether *F* can be defined in terms of *S* and *K*, the answer is no, for much the same reasons that one cannot define Δ in terms of *S* and *K*. For example, combinators cannot distinguish *SKK* from *SK(KI)* as both represent the identity function, even though they are easily separated by *right*.

When factorisation is applied recursively, using fixpoints, it supports a general approach to divide-and-conquer algorithms: if the argument is a compound then divide it and recurse; if the argument is an operator then conquer it.

10.3 Intensional Examples

Here are some examples of intensional programs that have been adapted from Chapter 5. A simple example of this is the program *size* which determines the size of a normal

form. It is given by

$$\text{size} = Y_2(\lambda^* x. \lambda^* s. Fx1(\lambda^* y. \lambda^* z. (sy) + (sz))) .$$

Thus, the size of an atom is 1 and the size of a compound MN is the sum of the sizes of M and N .

To define the equality of programs requires a means of deciding if two operators are equal. Define isF by

$$\text{isF} = \lambda^* x. x(KI)(K(KI))K .$$

It maps F to K and S to KI as desired since

$$\begin{aligned}\text{isFF} &= F(KI)(K(KI))K = KKI = K \\ \text{isFS} &= S(KI)(K(KI))K = KIK(K(KI))K = I(KI) = KI .\end{aligned}$$

From this, we can define equality of operators by

$$\text{eqatom} = \lambda^* x. \lambda^* y. \text{iff}(\text{isF } x) (\text{isF } y) .$$

Structural equality of programs can now be given by

$$\begin{aligned}\text{equal} &= Y_3(\lambda^* e. \lambda^* x. \lambda^* y. \\ &\quad Fx(Fy(\text{eqatom } x \ y)(K^3 I)) \\ &\quad (\lambda^* x_1. \lambda^* x_2. Fy(KI)(\lambda^* y_1. \lambda^* y_2. \\ &\quad \text{and}(e \ x_1 \ y_1)(e \ x_2 \ y_2)))) .\end{aligned}$$

Theorem 64 (equal_sf_programs) $\text{equal } p \ p \longrightarrow K$ for all programs p .

Proof. The proof is by induction on the structure of programs. \square

Theorem 65 (unequal_sf_programs) $\text{equal } p \ q \longrightarrow KI$ for all programs p and q that are not equal.

Proof. The proof is by induction on the structure of programs. \square

Tagging is defined by

$$\text{tag}\{t, f\} = S(S(KK)f)t .$$

Application removes the tag since

$$\begin{aligned}\text{tag}\{t, f\} \ x &= S(KK)fx(tx) \\ &= KKx(fx)(tx) \\ &= fx .\end{aligned}$$

Also, the tag can be recovered by defining getTag to be right.

Pattern matching in *SF*-calculus is similar to that of tree calculus. The patterns are built from variables, operators and application. The corresponding *extensions* are defined as follows:

$$\begin{aligned} x \Rightarrow s | r &= \lambda^* x.s \\ S \Rightarrow s | r &= S(S \text{ isS } (Ks))r \\ F \Rightarrow s | r &= S(S \text{ isF } (Ks))r \\ p q \Rightarrow s | r &= S(SFr)(K(p \Rightarrow (q \Rightarrow s | (S(Kr)p)) | (S(Kr)))) \end{aligned}$$

10.4 Translation from Tree Calculus

Since triage and factorisation are so similar, it will be no surprise that there are meaningful translations between tree calculus and *SF*-calculus. However, the translations are not trivial because leaves are not translated to atoms and *F* is not translated to a leaf. Here are some details.

In tree calculus, the atoms are the leaves and the compounds are the stems and forks. So factorisation is there given by

$$F_0 = \lambda^* x. \lambda^* y. \lambda^* z. \text{triage}\{y, z\Delta, \Delta K(Kz)\}x .$$

For example, when applied to some fork $\Delta x_1 x_2$ and to y and z it reduces to

$$\Delta K(Kz)x_1 x_2 = Kz x_1 (\Delta x_1) x_2 = z(\Delta x_1) x_2$$

as desired. Unfortunately, we cannot translate the operator *F* of *SF*-calculus by F_0 since F_0 is a fork so $F_0 F_0 y z$ reduces to an application of z .

Rather, the key to this, and the translations to follow, is to use tags to record the intensional information required for the reduction rules. Here, the translation of Δ applied to x, y and z will reduce to

$$\text{getTag } x \ y \ z .$$

In turn, the required tags are revealed by the following reductions

$$\begin{aligned} Kyz &\longrightarrow y \\ SS(Kx)yz &\longrightarrow Sy(Kxy)z \longrightarrow yz(xz) \\ K(S(SI(Kw))(Kx))yz &\longrightarrow SI(Kw)z(Kxz) \longrightarrow zwx . \end{aligned}$$

Thus, the translation of Δ is given by adding these tags to `getTag` to produce

$$\begin{aligned} \Delta_f = \text{tag}\{K,} \\ \lambda^* x. \text{tag}\{SS(Kx),} \\ \lambda^* y. \text{tag}\{K(S(SIKx)(Ky)),} \\ \text{getTag } x \ y \ z \\ \}\} . \end{aligned}$$

The picture of Δ_f is in Figure 10.2

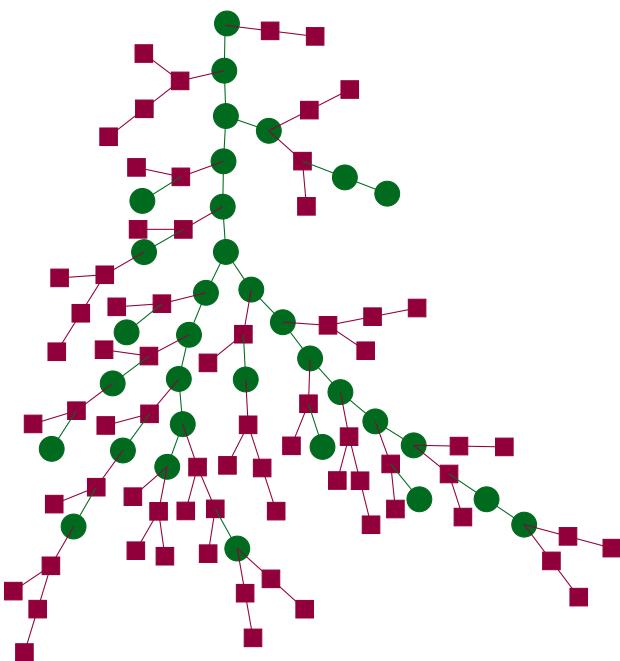


Figure 10.2: The Translation of Δ to SF-Calculus: Circles are S and Squares are F

Theorem 66 (meaningful_translation_from_tree_to_sf) *There is some meaningful translation from tree calculus to SF-calculus.*

Proof. Translate Δ by Δ_f . Details of the proof is routine. \square

10.5 Translation to Tree Calculus

When translating SF-calculus to tree calculus, tags are used once again to keep track of intensional information. The tags that will be used to represent F are revealed by the computations

$$\begin{aligned} Kyz &\longrightarrow y \\ \Delta(\Delta O x)yz &\longrightarrow zOx \\ K \text{ mythree}\{O, w, x\} y z &\longrightarrow z(Ow)x . \end{aligned}$$

where O is an operator, provided that $\text{mythree}\{O, w, x\} z \longrightarrow z(Ow)x$. The obvious solution for mythree is $\lambda^* z. z(Ow)x$ but this is not a program because Ow will reduce. So add some waiting to obtain

$$\text{mythree}\{f, x, y\} = \Delta(\Delta(\Delta(\Delta I)(\Delta(\Delta(Kf))(Kx))))(Ky) .$$

The other point to note is that the translations of the operators O reference themselves in the tags, so that recursion is required. Putting this all together, we can translate F by

$$\begin{aligned} \text{tag}\{K,} \\ Y_2\{\lambda^* f. \lambda^* x. \\ \text{tag}\{\Delta(\Delta \text{tag}\{K, f\} x),} \\ \lambda^* y. \text{tag}\{K(\text{mythree}\{\text{tag}\{K, f\}, x, y\}),} \\ \text{getTag } x \ y \\ \}}\} . \end{aligned}$$

Further, S can be translated by modifying this program to use the combination S of tree calculus in place of getTag . So define

$$\begin{aligned} \text{ternary_op}\{f\} = \text{tag}\{K,} \\ Y_2\{\lambda^* f. \lambda^* x. \\ \text{tag}\{\Delta(\Delta \text{tag}\{K, f\} x),} \\ \lambda^* y. \text{tag}\{K(\text{mythree}\{\text{tag}\{K, f\}, x, y\}),} \\ fxy \\ \}}\} \end{aligned}$$

and then we have

$$\begin{aligned} S_t &= \text{ternary_op}\{S\} \\ F_t &= \text{ternary_op}\{\text{getTag}\} . \end{aligned}$$

For any combination f and terms x, y and z we have

$$\text{ternary_op}\{f\}xyz \longrightarrow fxyz$$

so that

$$\begin{aligned} S_t xyz &\longrightarrow Sxyz \longrightarrow xz(yz) \\ F_t xyz &\longrightarrow \text{getTag } x \ y \ z \end{aligned}$$

as intended.

The pictures of $\text{ternary_op}\{f\}$ is given in Figure 10.3 .

Theorem 67 (meaningful_translation_from_sf_to_tree) *There is some meaningful translation from SF-calculus to tree calculus.*

Proof. The required translation maps S to S_t and F to F_t . The verification is routine. \square

Exercises

1. Reduce the following combinations to normal form:

$$FFK(KI), FSK(KI), F(FF)K(KI), F(FS)K(KI), FKK(KI), FIK(KI) .$$

Take care to distinguish the operators from the compounds!

2. Show that if x is a compound then $(\text{left } x)(\text{right } x)$ reduces to x for any compound x .
3. Define DF -calculus where $Dxyz \longrightarrow yz(xz)$. Can you define a meaningful translation from SF -calculus to DF -calculus? Hint: take care with the atoms.
4. Verify that if M is a normal form then Y_2M reduces to a normal form.
5. Verify the diamond property for simultaneous reduction in SF -calculus.
6. Check the properties of tagging in SF -calculus.

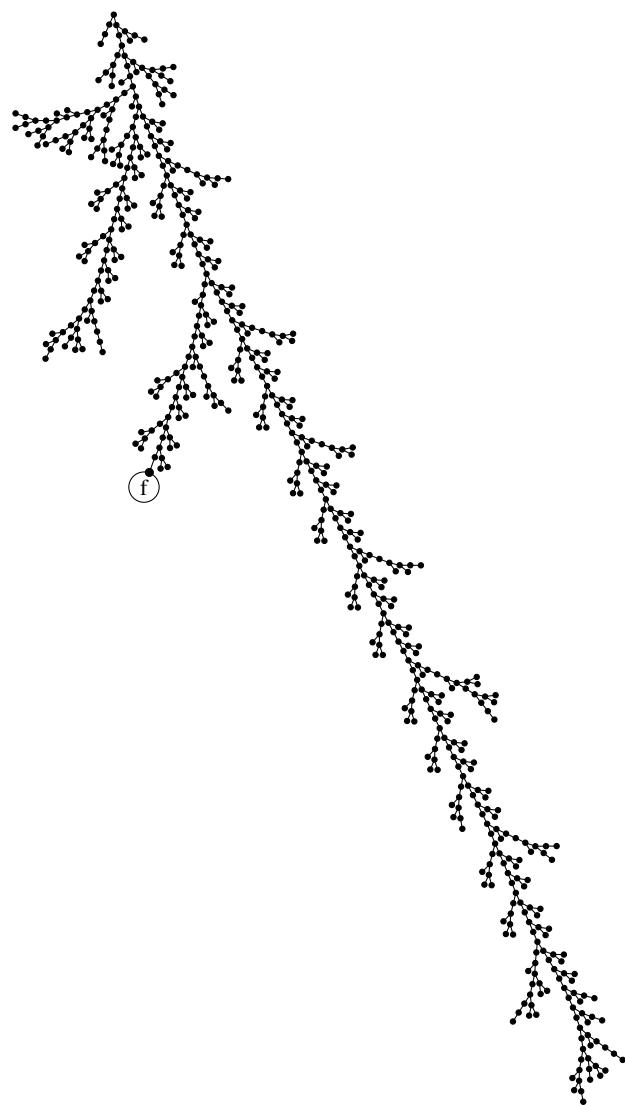
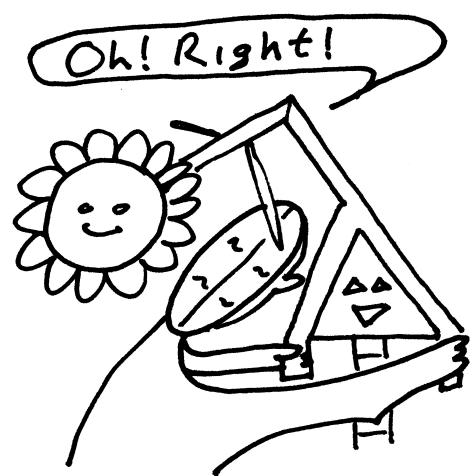


Figure 10.3: `ternary_op{f}` is Used to Translate S and F to Tree Calculus

Chapter 11

Concluding Remarks



11.1 Recapitulation

Tree calculus is a better foundation for computation than any of the traditional models because of its simplicity, naturality, and expressive power, especially its support for program analysis. This may have implications for several disciplines.

Tree calculus is simple because it is built using combinations of a single operator, governed by three simple equations that do not require any side-conditions, or additional machinery of any kind. This compares well with traditional models. For example, Turing machines require a mechanism for converting the transition table into actions upon the tape. Again, lambda-calculus requires machinery for substitution that comes with side-conditions to maintain scoping. Combinatory logic is also equational, but not quite as simple, since it requires more than one operator.

Tree calculus is natural because it computes natural binary trees, whose nodes are unlabelled. More precisely, the functions, values, programs, inputs and outputs are all given by natural binary trees, while the computations are the finitely-branching natural trees. Again, this compares well with traditional models. Turing machines, lambda-abstractions and combinators are artificial. Although the μ -recursive functions compute natural numbers, the functions themselves are artificial. Conversely, if the programs are defined to be the natural numbers then elaborate meta-theory is required to perform the computations. In tree calculus, every binary tree is simultaneously a data structure and a program; it can be the argument of a function or a function that takes an argument.

Tree calculus is powerful because it expresses more functions than the traditional models of computation. As well as the numerical computations expressed as μ -recursive functions, the combinators of combinatory logic, the abstractions of λ -calculus and the analyses of Turing machines, it can decide program equality and self-evaluate. By contrast, traditional models must be augmented with meta-theory to supply the missing expressive power: the μ -recursive functions and the Turing machines maintain a formal separation of functions and arguments so that meta-theory is essential for reflection; combinatory logic and lambda-calculus are incomplete, since they cannot decide equality of values. All of these claims have been captured as theorems and verified using the Coq proof assistant.

The fundamental work is now done, except that there are some outstanding issues which deserve comment. They are raised here, rather than in the introduction or the body of the work, so that the status of the theorems is not mixed up with the verbiage below.

11.2 Program Completeness

In what sense is tree calculus complete? Traditionally, a calculus was deemed complete if it was *Turing-complete*, in that it computed the same class of numerical functions as the Turing machines. The supposition that the Turing-computable numerical functions include all computable numerical functions is quite reasonable, but unproven. This became known as *Turing's Thesis* or the *Church-Turing Thesis*. We have seen that tree calculus is Turing-complete. Further, it supports a rich class of functions that act on programs, including decidable equality, pattern-matching functions, tagging and self-

evaluators, that are not supported by the traditional models. It is not yet clear what the class of computable functions of programs should be, but here are some thoughts.

One option is to replicate the traditional approach but with binary natural trees instead of natural numbers. Now the tree calculus plays the role of the μ -recursive functions so that, by definition, tree calculus is tree-complete. However, all of the traditional models are complete in this sense, too. The problem is that their programs are not binary trees.

A second option is to require that the theory supports an invertible function that converts programs to natural numbers. If the model is also Turing-complete then it supports a rich class of functions of programs. Tree calculus supports such a function but the traditional models of computation do not. The problem is that this privileges the natural numbers over the natural trees, even though the latter are proving to be more important.

The third option combines the previous two, to declare that a theory is *program-complete* if it supports the same class of functions of binary trees as tree calculus, and it supports an invertible function from programs to binary natural trees. Now tree calculus is program-complete by definition but the traditional models of computation are not program-complete. Further, VA-calculus is not program complete, since it cannot define equality of programs, but SF-calculus is program-complete since the translation from programs to natural trees can be defined by pattern-matching.

11.3 Extensional, Intensional and Reflective Programs

One of the themes of this work has been the contrast between extensional, intensional and reflective programs. Although many examples have been given, these terms were never given formal definitions, as they were not required for the theorems. By contrast, it was essential to define the programs precisely in order to prove, for example, that equality of programs is decidable. Here, finally, is an approach.

The fundamental distinction is between extensional and intensional accounts of functions. An extensional description of a function is given by the relationship between inputs and outputs, i.e. a functional relation. In set theory, this becomes the definition of a function, as a set of ordered pairs. An intensional description of a function is an algorithm or construction that produces outputs from inputs, in the style of, say, the *intuitionism* of L.E.J. Brouwer. This is not to be confused with an intentional description, which includes the intent of the programmer, as captured by, say, the type of the program. For example, all tautologies are intensionally equal to the term for truth, even though they have different intentions. Conversely, two programmers who intend to produce the identity function may produce programs that differ intensionally, as we saw in combinatory logic.

On this understanding, an extensional program p should gain knowledge of an extensional function f only by applying f to arguments. Arguably, p should be an abstraction of some applications of variables, i.e. a combinator, in which case, a model is *extensionally complete* if and only if it is combinatorially complete. All of the models of computation in this book are extensionally complete in this sense.

Intensional completeness is perhaps the ability to perform arbitrary program analyses, as supported by pattern-matching functions. That is, the allowable patterns are given by terms with the property that if all variables in the pattern are replaced by programs then the result is a program. Then there is little or no choice for the matching algorithm. Of course, this must be formalised in a way that avoids circularity, e.g. by defining the programs without reference to pattern-matching. In this sense, tree calculus and *SF*-calculus are *intensionally complete* but the traditional models and *VA*-calculus are not.

Finally, the reflective programs can be viewed as intensional programs whose intent is to be self-applied, as when the size program is applied to itself, or a self-evaluator is self-evaluated. Other examples have arisen during translation, as when A_t is applied to itself.

11.4 Translations to Tree Calculus

Since all of the traditional models of computation have meaningful translations to tree calculus, that preserve computations, it is natural to ask if every model of computation has such a translation. This is a separate question to completeness since, for example, *VA*-calculus supports a meaningful translation from tree calculus but is not complete. The answer appears to be a qualified yes.

If the model is combinational, and its evaluation rules do not overlap then it seems that there is a translation to tree calculus, in which tagging is used to keep track of the internal structure, just as was done for *VA*-calculus and *SF*-calculus. However, if the rules do overlap then the situation does not look promising, since it seems unlikely that tree calculus can support the ambiguity of overlapping rules. In particular, it seems unlikely that one can define parallel-or in tree calculus, whose overlapping rules evaluate both or true x and or x true to true. This intuition is reinforced by the theorem that reduction to normal form can be achieved by leftmost, outermost reduction. The prospect of a translation is even more remote if the model is not confluent, or not combinational.

11.5 Copying is Meta-Theoretic

In criticising traditional models of computation, I have stressed their reliance on meta-theory for, say, substitution, or quotation. Can the same criticism be levelled at tree calculus? A notable feature of the Turing model is that it can be physically implemented in a way that each computation step takes the same fixed amount of time. In particular, to read or write a symbol from a finite alphabet is well controlled. By contrast, evaluation rules for combinations commonly take one step to copy, delete or permute combinations. Since combinations may be arbitrarily large, a single computation step may take an arbitrarily long time. Perhaps a better calculus will keep all of the benefits of tree calculus while supporting the physicality of the Turing model.

The obvious approach would be to use pointers, or let-expressions, to allow sharing, so that trees are generalised to graphs. This would be a dramatic increase in the

complexity of the theory since, for example, equality of graphs is much harder to prove than that of trees. So this is best left to another time.

11.6 Implications for Programming Languages

The implementation of programming languages requires every program to play two roles, as both a function from inputs to outputs and as a data structure that can be manipulated by a compiler. Within compilers, the current emphasis is on the data structures. Programs are parsed into syntax trees which are then analysed and optimised before being converted to machine code. So, in practice, trees already play a central role even though they have been almost ignored in the basic theory. One reason for this is that each analysis stage requires its own syntax but this is a burden that may now be avoided. For example, if types are represented as tags, they won't interfere with functionality, and so can be left in place for as long as convenient, even as typed assembly language.

Program optimisations change the program's algorithm while preserving its functional behaviour. A pre-requisite for supporting optimisation in general is the ability to interpret programs. So self-interpretation serves as a convenient proxy for the ability of a language to optimise its own programs and support a self-compiler.

A side effect of the current approach is that compilation tends to be staged as: analysis, optimisation, execution. It is difficult to perform analyses after optimisation, or to optimise during execution. Of course, these things are possible, but they are tricky, so best left to a few experts among those constructing compilers.

These remarks apply even more to those who write applications. Currently, they are unable to influence the compilation process, e.g. to add a new optimisation, without engaging a compilation expert, or becoming one themselves. This is the motivation for domain-specific languages, where users may access these design decisions. A language based on tree calculus may reduce or eliminate the barriers between developers and compilers. It would also be a broad-spectrum language, supporting both high-level abstractions and manipulation of bits and leaves. Such a language should support a wide variety of programming styles, including functional, imperative and object-oriented styles, and a wide variety of type systems, too.

The ability to represent types as tags, means that types and terms can belong to the same syntactic class. As well as being a conceptually simpler means of supporting simple types, or parametric polymorphism, there is potential to develop dependently-typed calculi in which dependent products and sums may be analysed as freely as any program of tree calculus.

11.7 Implications for Logic

The Curry-Howard Isomorphism from the 1960s recognises the deep correspondence between propositions and their proofs on the one hand, and type and their terms on the other. However, this is no accident, since the very first function calculus, by Moses Schönfinkel, was built to give an account of predicate logic. Now that we have a more

expressive calculus of functions, perhaps we should look for a more expressive logic which will support proofs that correspond to intensional and reflective programs.

11.8 Implications for Mathematics

Various formalisms have been offered as a foundation for mathematics. The oldest uses numbers, the most popular is set theory, even logic and category theory have been proposed. Might trees be even better?

According to Pythagoras,

All is number

This idea is both simple and appealing but is not especially useful, as illustrated in Figure 1.1. However, it still has quite a hold on us. It re-appeared in Leopold Kronecker's dictum

God created the integers, all else is the work of man.

In computing theory tradition, the natural numbers are the common ground upon which comparisons of systems have been made. However, we have seen that the numbers are ill-suited to represent numerical functions.

In the set theory of Georg Cantor, functions are given by their extensions, as sets of ordered pairs. However, even the simplest numerical functions are then given by infinite sets, while the set of properties of natural numbers is even larger! In the standard terminology, the natural numbers are countably infinite, while the predicates on them are uncountably infinite. This is somewhat troubling when the predicates that can be described in any language are merely countable. Even worse is that naive set theory leads to paradoxes, most famously Russell's Paradox, which arises by considering the set of all sets. These infinities re-appear in Dana Scott's semantics of lambda-calculus, in which the denotation of any recursive function is the limit of its unfoldings.

The development of predicate calculus was introduced by Gottlieb Frege, and formalised by Bertrand Russell and Alfred North Whitehead. While it was the most significant development in logic since Aristotle, it led to Russell's Paradox mentioned above. Russell avoided the paradoxes by introducing a *type hierarchy*, in which higher-order functions act on lower-order functions, but this leads to restrictions on expressive power that re-appear in typed programming.

Category theory was developed by Samuel Eilenberg and Saunders Mac Lane as a means of formalising some relationships between geometry (topology) and algebra, and especially to define the *natural transformations*. Here, *arrows* represent extensional functions in a completely abstract manner, with no attempt to describe their internal structure, e.g. as sets or lambda-abstractions. That is, unlike lambda-calculus, the exploitation of extensionality avoids any use of algorithms, representations or intensionality. Surprisingly many properties of functions, data structures and logic can be developed in this way. William Lawvere considered category theory as a possible foundation for mathematics, but as soon as sets of objects and arrows appear, so do the set-theoretic difficulties.

By contrast, tree calculus is able to describe all of the predicates and functions of interest as finite trees. There is no need to consider the infinite, with its paradoxes, as it is constructive. Schönfinkel's goal has been met by going beyond the expressive power of predicate calculus. Hence, we can improve on Pythagoras by declaring

All is tree

Let us see how far this can take us.

11.9 Implications for Traditional Computing Theory

That tree calculus is a better foundation for computing than any previous theory of computation conflicts with the traditional understanding, first articulated by Alonzo Church, that all of the traditional models of computation were equivalent and unlikely to be surpassed. After all, tree calculus can be implemented as a Turing machine, so does not appear to be adding anything. Taken at face value, the tension between this view and the theorems presented in this work appears to be a paradox. However, the paradox is resolved by observing that Church's notion of equivalence is very weak. First, it is limited to computation of numbers without any concern for computation of programs. Second, the translations between models are not meaningful, as they do not preserve the process of computation but merely preserve numerical results. Third, the double translations, from one model to another and then back again, are not usually definable in the original model. When comparisons use the meaningful translations introduced here then the differences between models become much clearer. For example, there is no meaningful translation from VA-calculus to combinatory logic, or from combinatory logic to tree calculus.

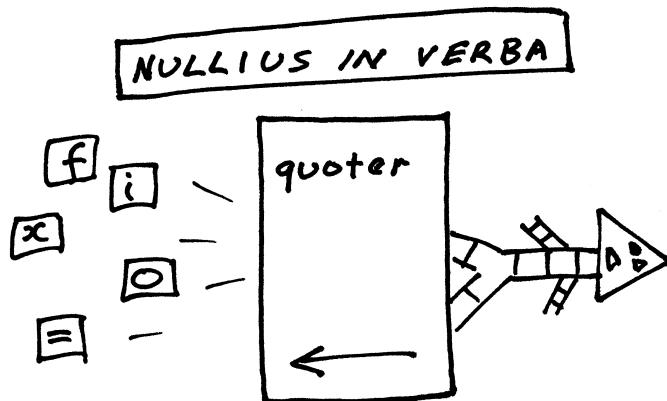
Jose Vergara and I developed the critique above in a much refereed yet unpublished paper covering eighty years of computing theory. For those who are troubled by the paradox above, it is included, unchanged, in the appendix.

As the remarks above show, the simplicity, naturality and expressive power of trees calculus surpasses that of any previous theory of computation, and suggests many avenues for further exploration, in both theory and practice.

Appendix A

Confusion in the Church-Turing Thesis

Barry Jay and Jose Vergara



Abstract

The Church-Turing Thesis confuses numerical computations with symbolic computations. In particular, any model of computability in which equality is not definable, such as the λ -models underpinning higher-order programming languages, is *not* equivalent to the Turing model. However, a modern combinatory calculus, the *SF*-calculus, can query its closed normal forms, and so yields a model of computability that is equivalent to the Turing model. This ability to support queries as well as the usual functions has profound implications for programming language design.

A.1 Introduction

The foundations of computing contain an error, the belief that the λ -calculus gives a complete account of computation, an error born of the confusion of numeric and symbolic computation in the Church-Turing Thesis (CTT). The confusion is resolved by: defining equivalence of models of computability that act on different symbolic domains; proving that the λ -calculus model of computability is *not* equivalent to the Turing model of computability; and proving that there are rewriting systems whose model of computability *is* equivalent to the Turing model. To the extent that rewriting systems like λ -calculus underpin the design of programming languages, this opens up significant new possibilities.

So far, responses to our claim of error have ranged from mild approval, through discouragement, to mockery and hostility. If you are feeling negative, then it is up to us to persuade you to engage in the painstaking effort of reconsidering some fundamental beliefs. If your reaction is mild approval then it is up to us to outline the implications, especially the impact on programming language design.

To be clear, the error is not in the work of Alan Turing, or Turing's Thesis (TT), and has no impact on the study of computing hardware or algorithmic complexity. Further, there is no logical error in the refereed papers of Alonzo Church, or in Church's Thesis (CT). Rather, the source of the conflict lies within the standard account of the expressive power of λ -calculus; as shown in our recent paper *Conflicting accounts of λ -definability* [42].

Corollary 68 *The equality function of closed normal forms known as Church's δ is λ -definable.*

can be sourced back to a statement in Kleene's book [46, p. 320] "*The equivalence of the computable to the λ -definable functions ... was proved by Turing 1937*". This without any restriction on the domain of the computable functions; implies that Church's δ is λ -definable.

However, in Barendregt's book [1, p. 520, Corollary 20.3.3] we have its negation

Corollary 69 *Church's δ is not λ -definable.*

This error is not of recent origin, but has been present from the beginning. Looking at the CTT historically does not help us to solve the problem; according to Soare [69, page 11], the Church-Turing Thesis is first mentioned by name in Steven Cole Kleene's book *Introduction to Metamathematics* [46]. However, close examination shows that the thesis is never actually stated there, neither in the section of that name, nor in the preceding text. Since Kleene took some pains to define Turing's Thesis and Church's Thesis, this is a little surprising. In the decades since, one can find many statements of the thesis, e.g. [25, 68, 77, 22, 31, 5], but with a variety of distinct meanings, including TT, CT, or some claim of equivalence, but no clear winner. Hence, we must conclude that

There is no canonical statement of the Church-Turing Thesis.

Further, there is no way to add a definition that makes Kleene's book whole, since it already contains an error due to the confusion of numeric and symbolic computation, similar to that identified above.

Although Kleene's book cannot supply a definition of CTT, there is one last source to consider. According to Solomon Feferman [22], the Church-Turing Thesis was born (but not named) in Alonzo Church's 1937 review of Alan Turing's paper on computability [75] which declared

As a matter of fact, there is involved here the equivalence of three different notions: computability by a Turing machine, general recursiveness in the sense of Herbrand-Gödel-Kleene, and λ -definability in the sense of Kleene and the present reviewer.

However: computability by a Turing machine concerns words in some alphabet; general recursion concerns natural numbers; and λ -definability concerns λ -abstractions. So it is not clear whether computation is limited to numbers, or also includes symbolic computations.

If computation is limited to the domain of natural numbers then the statement is true, since all three notions describe the same set of numerical functions, but this version of CTT has no relevance to the design of higher-order languages that now dominate programming language design. As both Turing and Church understood, greater relevance requires consideration of arbitrary domains. However, it is not then clear, at least from the review, what "equivalence of notions" is to mean. The natural interpretation is that a notion is a predicate, so that equivalence of notions is equality of sets. This works for numeric computations but cannot work in general since, as noted above, there are Turing computable functions of λ -terms that are not λ -definable.

If a notion is not a predicate then there is no obvious alternative interpretation. However, by replacing "notion" with "model of computability" we are led to consider equivalence of models of computability. When these models use different domains, then the equivalence must employ simulations or encodings of one domain in another.

Certainly, Church used Gödelisation to simulate symbolic computation by numeric computation [15]. Conversely, λ -calculus can simulate numeric computation through Church encodings. However, the mere existence of simulations in both directions makes for a poor notion of equivalence.

To see this, consider the question of computational completeness, or adequacy. In the common understanding, as exemplified by the version by Barendregt et al given earlier, the Church-Turing Thesis includes the claim that anything that can be computed can be computed by a Turing machine, or using general recursion, or in λ -calculus, that these three models are computationally complete. In this context, it is to be expected that the simulations themselves are computable, since otherwise one could import, say, the solution to the Halting Problem as part of the simulation. While this expectation is natural, it cannot be enforced, since the simulations are meta-functions, outside of any particular model. That is probably why the CTT is called a thesis and not a theorem.

However, this is to give up too easily. When there are simulations in both directions then we can compose them, e.g. to obtain simulations of numbers by numbers, and λ -terms by λ -terms. That is, a natural number is mapped to the Gödel number of its Church encoding, while a λ -term is mapped to the Church encoding of its Gödel

number. If the models are computationally complete, then both these re-codings should be computable in their respective models. Now the re-coding of natural numbers is general recursive, but the re-coding of λ -terms is *not* λ -definable, even for closed normal forms. Further, the re-coding of λ -terms in closed normal form is effectively calculable, so it follows that

The λ -calculus and Turing machines yield models of computability that
are **not** equivalent; λ -calculus is **not** computationally complete.

This, in outline, identifies the confusion, shows how it leads to error, and clarifies how to compare models of computability. What then, are the consequences of abandoning CTT? Since we still have CT and TT, perhaps it doesn't matter too much. Our response comes in two parts. First, we will show that there are new calculi that improve upon λ -calculus. Second, these new calculi suggest radical approaches to the design of programming languages. Both propositions will be incredible to adherents of CTT.

The great strength and weakness of λ -calculus is that there is only one way to obtain information, namely to evaluate functions; it is *extensional*. It follows that there is no uniform mechanism for performing analysis of functions or other structures, for writing queries; it is not *intensional*. Rather, there are ad hoc techniques for analysing (the representations of) natural numbers, etc. Hence, when general, symbolic computation is confused with numerical computation, the λ -calculus appears to be more powerful than it is.

However, once the general importance of intensionality is recognised, we can discover or build calculi that support it, as well as extensionality. Our latest published example is *SF*-calculus [38], a combinatory calculus which is more expressive than traditional combinatory logic [32] in that it can Gödelise its own closed normal forms, and so perform arbitrary analyses of closed normal forms. Thus,

The *SF*-calculus and Turing machines yield models of computability that
are equivalent.

Now let us consider the implications for programming language design. The supposed pre-eminence of λ -calculus makes it impossible to comprehend why there are so many programming styles in use, or even why there are so many λ -calculi. It compares badly with the Turing model, which works so well for representing hardware. Further, λ -calculus has constrained the thinking of those who explore the possibilities for programming language design. Freed from these constraints, we offer an alternative view of the possibilities.

The existence of a calculus with generic support for both extensionality and intensionality opens up many possibilities for programming language design. It can be used to support generic queries of data structures, so that a program that searches for an item may be applied to a list or binary tree or any other sort of structure, as has been implemented in the programming language **bondi**[9]. More generally, *SF*-calculus is able to analyse programs, so that there is no need to separate program analysis from program execution. Further, if the equivalence of *SF*-calculus and Turing machines is strong enough, if all programming is built from extensional and intensional functions, then there may be a single programming language that is universal for software, in the sense that the Turing model is universal for hardware.

- (1) “... every function which would naturally be regarded as computable is computable under his [Turing’s] definition, i.e. by one of his machines ...” (page 376)
- (2) “The equivalence of the computable to the λ -definable functions . . . was proved by Turing 1937.” (page 320)
- (3) Every function which would naturally be regarded as computable is λ -definable. (equivalence)
- (4) The equality of λ -terms in closed normal form is λ -definable. (specialise)

Figure A.1: A faulty argument, with premises quoted from Kleene [46]

The structure of the paper is as follows. Section A.1 is the introduction. Section A.2 exposes an error in Kleene’s book. Sections A.3–A.5 consider Church’s Thesis, Turing’s Thesis and the Church-Turing Thesis as described in Kleene’s book, but drawing on other sources as appropriate. Sections A.6 and A.7 consider models of computability and their comparison. Section A.8 considers the impact of λ -calculus on programming language design. Section A.10 surveys some models of intensional computation. Section A.11 draws conclusions.

A.2 Confusion Leads to Error

Figure A.1 contains a valid argument with a false conclusion. [A closed normal form is an irreducible term that has no free variables.] Hence, one of the premises must be false, but these are direct quotations from Kleene’s book. Hence, Kleene’s book contains a false statement.

Now let us consider this in more detail. The premises (1) and (2) are direct quotes from Kleene’s book. Then (3) follows from (1) by Leibnitz equality, by substituting equals for equals, in this case by substituting “ λ -definable” for “computable” as per the equivalence in (2). Finally, (4) follows from (3) by specialisation, since equality of λ -terms in closed normal form is naturally regarded as computable. The conclusion asserts the λ -definability of equality of closed λ -terms in normal form. Since the conclusion is false [4, page 519] there must be a fault in the argument.

Two objections have been raised about this. First, it has been suggested that we are supporting a false conclusion. However, to assert that the argument is valid is not to support the truth of the conclusion. Rather, that the conclusion is false is essential to our purpose, of showing that one of the premises must be false. Second, it has been suggested that we must have taken the quotes out of context. However, it is hard to construct a context that would invalidate the argument, unless “every function which would naturally be regarded as computable” can be construed to exclude some functions which would naturally be regarded as computable, such as equality of λ -terms in closed normal form.

The source of the problem is that (1) only makes sense if the computable functions are taken to include symbolic functions, while in (2), Turing’s proof was for *numerical* functions only, and does not include the function in (4). The source of the error is the

confusion of symbolic and numeric computations. To clarify the issues, let us consider the theses of Church and Turing, etc. and their integration within the Church-Turing Thesis.

A.3 Church's Thesis

The background to Church's Thesis is Church's paper of 1936 *An unsolvable problem of number theory* [15] which, alongside Alan Turing's work discussed later, showed that some numerical problems do not have a computable solution, so that Hilbert's decision problem does not have a general solution. It is clear that Church's primary focus was on numerical functions, as all his formal definitions were expressed in numerical terms. For example, he writes, “A function F of one positive integer is said to be λ -definable if . . .” [15, page 349]. Again, in Theorem XVII he writes “Every λ -definable function of positive integers is recursive”.

That said, he does broaden the discussion when considering *effective calculability*. On the one hand, an *effectively calculable* function of positive integers is defined to be “a recursive functions of positive integers” or “a λ -definable function of positive integers”. On the other hand, he writes:

. . . [in] some particular system of symbolic logic . . . it is necessary that each rule of procedure be an effectively calculable operation, . . . Suppose we interpret this to mean that, in terms of a system of Gödel representations for the expressions of the logic, each rule of procedure must be a recursive operation, . . .

That is to say, he supposes that a symbolic function is recursive if it can be simulated by a numerical function that is recursive, where simulation is defined using Gödel numbering. Later, this supposition hardens to become a premise when he writes, in a footnote:

. . . in view of the Gödel representation and the ideas associated with it, symbolic logic can now be regarded, mathematically, as a branch of elementary number theory.

By this definition, the symbolic function that decides equality of closed normal λ -terms is effectively calculable, is recursive, since it is simulated by the equality of natural numbers. However, there is no sense in which equality of closed normal λ -terms is λ -definable. Although there is no logical error here, Church does not show any awareness of the tension between his accounts of λ -definability and effective calculability, which makes it easy for Kleene to fall into error.

In contrast to Church's narrow constructions, Kleene's definitions have broad scope. For example, he gives three definitions of λ -definability in his paper *λ -definability and recursiveness* [45], according to whether the domain of definition is the non-negative integers, the λ -terms themselves, or other mathematical entities for which a λ -representation has been given. It is the interplay between these definitions that is the primary cause of confusion.

Kleene introduces Church's Thesis as Thesis I in his 1952 book (Section 60, page 300) as follows:

(CT) Every effectively calculable function (effectively decidable predicate) is general recursive.

The crucial question is to determine the domain of the effectively calculable functions.

At the point where Church's Thesis is stated, in Section 60, the general recursive functions are all numerical, so it would seem that the effectively calculable functions must also be numerical. However, he does not include the phrase "of positive integers" in his statement of the thesis, in the careful manner of Church. We are required to add this rider in order to make sense of the thesis.

Later, in Section 62, Church's Thesis, Kleene presents seven pages of arguments for the thesis, which he groups under four headings A–D. In "(B) Equivalence of diverse formulations" he asserts that the set of λ -definable functions and the set of Turing computable functions are "co-extensive" with the set of general recursive functions. Again, this only makes sense if the functions are presumed to be numerical functions. This paragraph is also the source of statement (2) from Figure A.1.

If we were to stop at this point, then the explanation of the faulty argument would be quite simple: statement (2) should be read in a context where all functions are numerical, or be replaced by "The equivalence of the computable to the λ -definable **numerical** functions was proved by Turing 1937". The restriction to numerical functions propagates to statement (3) which cannot then be specialised to equality of λ -terms.

However, in "(D) Symbolic logics and symbolic algorithms", Kleene reprises Church's definition of symbolic functions that are recursive, so that, by the end of Section 62, we have two definitions of effectively calculable functions and of recursive functions. So there are two versions of Church's Thesis, one for numerical functions (NCT) and one for symbolic functions (SCT). Unpacking the definition of a general recursive symbolic function to make the role of simulations explicit, the two versions become:

(NCT) Every effectively calculable numerical function (effectively decidable numerical predicate) is general recursive.

(SCT) Every effectively calculable function (effectively decidable predicate) can be simulated by a function which is general recursive.

Now consider the faulty argument in the context of SCT. We can infer that every λ -definable function is general recursive, but *not* that every general recursive function is λ -definable: these sets are not co-extensive; there is no equivalence in which (2) holds for symbolic computation.

Summarising, we see that Church was careful to separate theorems about numerical functions from the discussion of computation in symbolic logic. By contrast, Kleene presents a single statement of Church's Thesis with evidence that confuses the numerical with the symbolic. In turn, this confuses two different questions: whether two sets of numerical functions are the same; and whether there is an encoding that allows functions in a symbolic logic to be simulated by recursive functions on numbers. These

confusions can be defused by isolating two versions of Church's Thesis which, from the viewpoint of Post [62], qualify as scientific laws. Then the faulty argument can be evaluated as follows. If (2) refers to numerical computation only then the specialisation that produces (4) is invalid. If (2) refers to symbolic computation then it is false, on the assumption that the equivalence is a Leibnitz equality, that supports substitution of equals for equals. We shall explore alternative notions of equivalence in Sections A.6 and A.7.

A.4 Turing's Thesis

Turing's paper of 1936 *On Computable Numbers, with an application to the Entscheidungsproblem* was, like Church's paper, concerned with numerical computation and Hilbert's decision problem. Like Church, Turing was careful to limit his definitions, e.g. of *computable functions*, to numerical functions while showing awareness of a broader scope. For example, in the first paragraph he writes:

Although the subject of this paper is ostensibly the computable *numbers*, it is almost equally easy to define and investigate computable functions of an integral variable, or a real or computable variable, computable predicates, and so forth.

Similarly, the use of an unspecified alphabet of symbols on the tape of a Turing machine encourages us to consider computation over arbitrary symbolic domains.

Once again, Kleene confuses these two meanings in his third piece of evidence for Church's Thesis, headed (C): Turing's concept of a computing machine [46, page 320] where Kleene writes "Turing's computable functions (1936-7) are those which can be computed by a machine of a kind which is designed, according to his analysis, to reproduce all the sorts of operations which a human computer could perform, working according to preassigned instructions." As we have seen above "Turing's computable functions" are, by definition, numerical, while a human computer faces no such restriction.

In more compressed form, this confusion re-appears in Kleene's statement of Turing's Thesis. It is given in Section 70. Turing's thesis, as a sub-ordinate clause of the opening statement which, when elevated to an independent thesis, becomes:

(TT) Every function which would naturally be regarded as computable is computable under Turing's definition, i.e. by one of his machines.

Now the phrase "every function which would naturally be regarded as computable" surely includes all computations in formal systems such as λ -calculus or combinatory logic [32]. For example, it would be a distortion to assume that "naturally" here refers to the natural numbers. On the other hand, "Turing's definition" is certainly numerical. As with Church's Thesis, the solution is to create a numerical thesis (NTT) and a symbolic one (STT) as follows:

(NTT) Every numerical function which would naturally be regarded as computable is computable under Turing's definition.

(STT) Every function which would naturally be regarded as computable can be simulated by a function which is computable under Turing's definition.

From the viewpoint of Post [62], both versions of the thesis will qualify as scientific laws.

A.5 The Church-Turing Thesis

As noted in the introduction, the Church-Turing Thesis is mentioned in Kleene's book without being defined. Among the statements in Kleene's book, the closest candidate is the opening of Section 70:

Turing's thesis that every function which would naturally be regarded as computable is computable under his definition, i.e. by one of his machines, is equivalent to Church's thesis by Theorem XXX.

On first reading, it is rather difficult to determine the nature of this declaration, as it contains the statement of Turing's thesis (TT) plus the statement of a theorem,

Turing's Thesis is equivalent to Church's Thesis.

with its proof "by Theorem XXX". If this is the Church-Turing Thesis, then it is a theorem asserting logical equivalence of two theses. Further, the proof by Theorem XXX (on the same page) is numerical, which implies that the thesis is also numerical. One must look elsewhere for a general, symbolic result, e.g. in Kleene's evidence for Church's Thesis mentioned earlier. In any event, the quote above cannot carry the burden that CTT is expected to bear.

The subsequent literature of the last fifty years has thrown up many versions of the theses, e.g. [25, 68, 22, 5]. Perhaps the best way to make sense of this variety is to view the Church-Turing Thesis as the class of all statements that are logically equivalent to any version of Church's Thesis or of Turing's Thesis. In this manner, all of the theses and proofs of logical equivalence are gathered under a single heading. This broad interpretation may explain why Kleene did not give a statement of it. In any event, this broad interpretation seems most appropriate when considering the impact of the Church-Turing Thesis on programming language design in Section A.8. Before leaving the founders of the theory of computation, let us consider how they might have compared models of computability, to decide relative expressive power, and completeness.

John Longley's Book *Higher order computability* [50] presents an impressive review of the last 50 years of the literature and looks at the question of what is the meaning of computability for domains beyond natural numbers.

Whereas the work of Church and Turing from the 1930s provided a definitive concept of computability for natural numbers and similar discrete data, our present inquiry begins by asking what 'computability' might mean for data of more complex kinds. In particular, can one develop a good theory of computability in settings where 'computable operations' may themselves be passed as inputs to other computable operations? What does it mean to

‘compute with’ (for example) a function whose arguments are themselves functions?

However it leaves for future work the development of models that are applicable for the design and development of programming languages for higher order computation.

In addition; The paper *The representational foundations of computation* [63] Explores the role of representations (encodings) for computability theory for a philosophical point of view. The paper points out the importance of being aware of the gap between the domain of a model of computation and a desired higher order domain. However the paper does not identify any problem with the literature in particular the CTT.

A.6 Models of Computability

We adopt the simplest definition of model of computability in which the discussion of simulation makes sense. This was introduced by Boker and Dershowitz as a *model of computation* [8], but since the focus is upon functions for which a computation is possible, rather than the actual mechanics of computation, it seems more accurate to call them models of computability. Note, too, that the domain of the computable functions is not actually required to be symbolic in any way, or even to be enumerable, however natural this may be. This makes the notion too weak for many purposes, but here it emphasises the generality of our results.

A *model of computability* (D, \mathcal{F}) is given by a domain D which is a set of *values* that provides arguments and results of computations, and a collection \mathcal{F} of partial functions from powers of D to D . Here are some examples.

The partial recursive functions on natural numbers form a model with domain given by the natural numbers and functions given by the partial recursive functions. Call this the *recursive model of computability*.

Recall that an injective function is a total function that does not identify distinct arguments. For any finite alphabet Σ , and any domain D equipped with an injective function from D to the words of Σ , the *Turing model of computability on D* has domain D and partial functions given by those which can be computed by a Turing machine with alphabet Σ . When the choice of domain is understood from the context, or unimportant, then it may be called the *Turing model of computability*.

Any *applicative rewriting system* [73] has a *normal model* whose values are the closed terms in normal form, and whose partial functions are those representable by closed terms in normal form. Further, any subset D of values determines a model, where the partial functions are now defined on a value in D only if the result is also in D .

For example, classical combinatory logic [32] has terms built from applications of the operators S and K and variables. As well as its normal model, there is a *numerical model* whose domain is restricted to be the Church numerals. Also, one can use Polish notation to encode combinators as words using the alphabet $\Sigma = \{A, S, K\}$, where A is for application. For example, $S(KK)$ is mapped to $ASAKK$. This yields a Turing model of computability for SK -normal forms.

Similarly, λ -calculus has *normal models* and *numerical models*, once the terms and reduction rules have been specified. First, a λ -term is unchanged by renaming of bound variables, i.e. is an α -equivalence class in the syntax [4]. Since this equivalence is a distraction, we will work with λ -terms using deBruijn notation [20] so that, for example, $\lambda x.x$ becomes $\lambda 0$ and $\lambda x.\lambda y.xy$ becomes $\lambda\lambda 10$. Second, there are various choices of reduction rules possible, with each choice producing a normal λ -model. Define a λ -model of higher-order programming to be any normal λ -model in which equality of closed normal forms is not definable. This excludes any model whose domain is numerical, and would seem to include any models that are relevant to higher-order programming. Certainly the λ -models of higher-order programming include those given by β -reduction alone, or $\beta\eta$ -reduction.

A.7 Comparison of Models

Now consider what it means for one model of computability to be more expressive than another. The simple interpretation requires that their computable functions have the same domain and then compares sets of computable functions by subset inclusion, as is done by John Mitchell [54] and Neil Jones [43]. The choice of domain is important here. For example, if the domain consists of natural numbers then the λ -model and the recursive model are indeed equivalent. However, this restriction is unreasonable for modeling higher-order programming since functions must be among the values. Now it is an easy matter to see that any normal λ -model of computability has fewer computable functions than the Turing model. For example, no λ -term can decide equality of values but this function is in the Turing model.

The complex interpretation of relative expressive power allows the domains to vary, but now the comparison of computability over domains D_1 and D_2 must be mediated by simulations, which are given by encodings. Note that it makes no sense to consider a simulation in one direction, and compare sets of functions in the other direction, since, as observed by Boker and Dershowitz [8], this can lead to paradoxes. Rather, there should be encodings of each domain in the other that are, in some sense, inverse to each other. Various choices are possible here but two requirements seem to be essential. First, the encodings should be injective functions, since distinct values should not be identified. Second, the encodings should be *passive*, in the sense that they are not adding expressive power beyond that of their target model. In particular, they should be effectively calculable. This requirement can only be verified informally, on a case by case basis, since functions from D_1 to D_2 are not in the scope of either model.

A *simulation* of a model of computability (D_1, \mathcal{F}_1) in another such (D_2, \mathcal{F}_2) is given by an injective *encoding* $\rho : D_1 \rightarrow D_2$ such that every function f_1 in \mathcal{F}_1 can be *simulated* by a function f_2 in \mathcal{F}_2 in the following sense: for all $x_1, \dots, x_n \in D_1$ such that $f_1(x_1, \dots, x_n)$ is defined, then $f_2(\rho(x_1), \dots, \rho(x_n))$ is defined and

$$\rho(f_1(x_1, \dots, x_n)) = f_2(\rho(x_1), \dots, \rho(x_n)).$$

For example, Gödelisation provides a simulation of the normal SK-model into the recursive model. More generally, Church and Kleene both use this notion of simulation to define recursive symbolic functions. Gödelisation seems to be passive.

Further, the encoding of the natural numbers using Church numerals provides a simulation of the recursive model into the normal SK-model or any normal λ -model.

Other related notions of simulation can be found in the literature, e.g. [43, 8]. For example, Richard Montague [57, page 430] considers, and Hartley Rogers [65, page 28] adopts, a slightly different approach, in which the encoding of numbers is achieved by reversing a bijective Gödelisation. However, this inverse encoding may not be a simulation. For example, the equality of numbers cannot be simulated by a λ -term over the domain of closed λ -terms in normal form. Rogers, like Kleene, ensures a simulation by *defining* the computable functions in the symbolic domain to be all simulations of partial recursive functions, but this says nothing about λ -definability.

Given the formal definition of simulations, it may appear that the strongest possible notion of equivalence is that each model simulates the other. However, if the encodings are passive then so are the *recodings* from D_1 to D_1 and from D_2 to D_2 obtained by encoding twice. Since these are in the scope of the two models, we can require that recodings be computable.

Let (D_1, \mathcal{F}_1) and (D_2, \mathcal{F}_2) be two models of computability with simulations $\rho_2 : D_1 \rightarrow D_2$ and $\rho_1 : D_2 \rightarrow D_1$. Then (D_2, \mathcal{F}_2) is *at least as expressive* as (D_1, \mathcal{F}_1) if the recoding $\rho_2 \circ \rho_1 : D_2 \rightarrow D_2$ is computable in \mathcal{F}_2 . If, in addition, (D_1, \mathcal{F}_1) is more expressive than (D_2, \mathcal{F}_2) then the two models are *weakly equivalent*. If the recodings both have computable inverses then the weak equivalence is an *equivalence*. Note that these notions of equivalence is indeed an equivalence relation on models of computability.

It is interesting to compare this definition with those for equivalence of *partial combinatory algebras* by Cockett and Hofstra [18] and John Longley [49]. They would not require the encodings to be injective, but Longley would require that the recodings be invertible. Adding the latter requirement is perfectly reasonable but is immaterial in the current setting.

It is easy to prove that the recursive model is at least as expressive as any λ -model. Our focus will be on the converse.

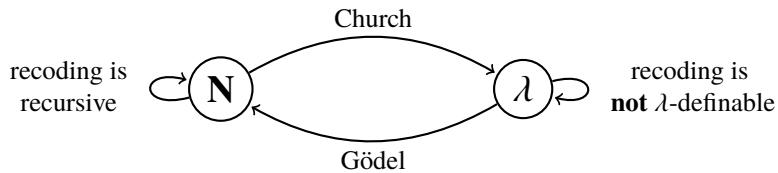
Theorem 70 *Any model of computability that is at least as expressive as the recursive function model can define equality of values.*

Proof. The recursive model is presumed to use 0 and 1 for booleans. In the other model, identify the booleans with the encodings of 0 and 1 so that the equality function is given by recoding its arguments and then applying the simulation of the equality of numbers. \square

Corollary 71 *The normal model of computability for SK-calculus is not weakly equivalent to the recursive function model.*

Proof. If the normal SK-model could define equality then it could distinguish the values SKK and SKS but the standard translation from combinatory logic to λ -calculus identifies them (both reduce to the identity), and so they cannot be distinguished by any SK-combinator. \square

Corollary 72 *No λ -model of higher-order programming is weakly equivalent to the recursive model of computability.*

Figure A.2: Recoding is not λ -definable

Proof. Since normal λ -models do not define equality, the result is immediate. Note that Longley has proved the analogous result for his definition of equivalence [49]. \square

Now we can express some classical results in the new terminology.

Theorem 73 *Turing's Numerical Thesis is logically equivalent to Church's Numerical Thesis.*

Proof. Apply Kleene's 30th theorem, i.e. Theorem XXX [46, page 376]. \square

Theorem 74 *The recursive model of computability is weakly equivalent to any Turing model of computability.*

Proof. The traditional simulations yield encodings that are computable. \square

Corollary 75 *Turing's Symbolic Thesis is logically equivalent to Church's Symbolic Thesis.*

Proof. Any simulation into a Turing model yields a simulation into the recursive model by composing with the simulation given by weak equivalence. The converse is similar. \square

Corollary 76 *No λ -model of higher-order programming is weakly equivalent to the Turing model.*

Proof. Since weak equivalence is transitive, the result follows from Corollaries 72 and 74. \square

Now any reasonable notion of equivalence must imply weak equivalence. So it follows that the Turing model of computability is strictly more expressive than any λ -model of computability suitable for modelling higher-order programming. The relationships are summarised in Figure A.2, where N denotes the natural numbers, λ denotes the λ -terms in closed normal form, and the arrows denote the various codings and re-codings.

A.8 Programming Language Design

Here are three examples, from the last fifty years, of how confusion in the Church-Turing Thesis has limited the design space for programming languages.

Peter Landin's seminal paper of 1966 *The Next 700 Programming Languages* [48] proposes a powerful model of programming language development in which λ -calculus is the universal intermediate language. That is: create a source language with various additional features such as types, or let-declarations; transform this into λ -calculus; then implement an evaluation strategy for λ -calculus (e.g. lazy or eager) as a Turing machine. His main comment about the suitability of λ -calculus for this role is:

A possible first step in the research program is 1700 doctoral theses called
"A Correspondence between x and Church's λ -notation."

So, Landin sees a central role for λ -calculus, with a research program that would occupy a generation of computer scientists. This research program influenced many languages designs, including Algol68 [64], Scheme [71], and ML [53]. This approach continues to hold sway.

Matthias Felleisen, in his paper of 1990 on the expressive power of programming languages [23], comments:

Comparing the set of computable functions that a language can represent
is useless because the languages in question are usually universal; other
measures do not exist.

After this appeal to the Church-Turing Thesis, the paper goes on to consider various λ -calculi, in the belief that nothing has been left out. Although Felleisen makes some useful distinctions, his paper excludes the possibility of going beyond λ -calculus, which limits its scope.

Indeed the development of Embedded Domain Specific Languages (DSL) using Higher Order Syntax [19]; can be complex due to the need to reflect and reify the terms of a language between two levels. A deep level where lambdas are encoded as lambdas in the host language, and a shallow level where lambdas and other terms are encoded as data structures. An understanding of programming language implementation beyond the CTT has the potential to simplify the development of embedded DSL's by using a uniform approach to functions and data structures.

Robert Harper's book *Practical Foundations for Programming Languages* [30] contains a version of Church's Thesis (given as Church's Law) which is careful to limit its scope to natural numbers. It also emphasises that equality is not λ -definable (given as Scott's Theorem). However, while the title proclaims the subject matter to be the foundation for programming languages in general, it is solely focused on the λ -calculus, with no allowance made for other possibilities. If there remains any doubt about the author's views about foundations, consider the following slogan, attributed to him by Dana Scott in the year of the book's publication. In a talk during the Turing Centenary celebrations of 2012, he asserts [67]:

λ conquers all!

There is no explicit justification given for this focus, which we can only assume is based upon Landin’s research program, and the Church-Turing Thesis.

Here are some examples of calculi and languages that don’t easily fit into Landin’s program, since they may exceed the expressive power of λ -calculus. Candidates include: first-order languages such as SQL [14]; languages without an underlying calculus, such as Lisp [52] with its operators `car` and `cdr`; and the intensional programming language Rum [72]. Richer examples include the *self-calculus* of Abadi and Cardelli for object-orientation [1], the *pattern calculus* for structure polymorphism [35], the *pure pattern calculus* for generic queries of data structures [39, 36]. Again, the **bondi** programming language [9] uses pure pattern calculus to support both generic forms of the usual database queries, and a pattern-matching account of object-orientation, including method specialisation, sub-typing, etc.

The richer calculi above have *not* been shown equivalent to λ -calculus. Rather, all evidence points the other way, since, for example, no-one has defined generic queries in pure λ -calculus.

Summarising, while the ability to simulate λ -calculus as a Turing machine has been enormously fruitful as an aid to implementation, the larger claims of the Church-Turing Thesis have been suggesting unnecessary limits on programming language design for almost fifty years.

A.9 Extensional and Intensional Computation

An account of functions is intensional if it allows access to the underlying algorithm of a function and not just its input-output relation. In this sense, the Turing computable functions, being mere relations, are extensional, while the Turing model of computation, in which we have access to the machine that performs the computation, is intensional. In this sense, the lambda-calculus is also intensional [76]. However, this ability to access the algorithm, to analyse a lambda-abstraction, cannot be represented within the lambda-calculus itself. According to this stronger standard, the lambda-calculus is *extensional* in the sense of Church [16, page 2]. By contrast, an *intensional* theory of computation is able to examine the algorithms, to analyse or query them, just as we query data structures for pairs, lists or binary trees. In this sense, the Turing model of functions is intensional, since functions are represented as symbols on a tape, all of which are readable.

Even within λ -calculus, intensionality proves to be quite desirable, as a means of recording type information or naming constructors, etc. Since such information, such intensions, cannot be properly represented within the pure λ -calculus, the community is continually developing new variants of λ -calculus in which the additional information is provided to the program analyst. Somehow, the existence of hundreds, if not thousands of different calculi does not seem to disturb our foundational beliefs to any extent. So resigned are we to the situation that the reference work *Term Rewriting Systems* defines the λ -calculus to be a family of calculi!

Intensionality is so central to partial evaluation and decompilation that various efforts have been made to extend λ -calculus with Gödelisation. This has been done for simply typed λ -calculus [6], a combinatory calculus [29], and untyped λ -

calculus augmented with some labels [56]. However, some of the attractive features of pure λ -calculus, such as being typable, confluent, and a rewriting system have been compromised.

Others have tried to develop a direct understanding of intensionality. It has been the subject of much research by philosophers [57, 21, 13, 44], logicians [24, 74, 11], type theorists [51, 58, 10] and computer scientists [33, 3, 12], so before proceeding, let us determine what it will mean for us. In the concrete setting of λ -calculus, when should two λ -terms be considered intensionally equal? Should this be limited to closed normal forms or are arbitrary terms to be included? In part, the answer depends upon whether your semantics is denotational or operational.

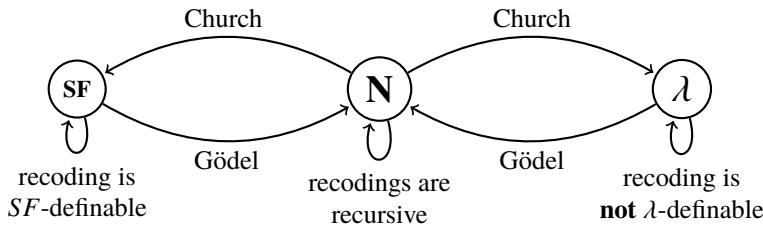
Denotational semantics constructs the meaning of a program from that of its fragments, whose contexts may supply values to free variables, or determine whether or not the evaluation of the fragment terminates. Examples may be found in *domain theory* [47, 66], *abstract and algebraic data types* [28, 55], the *effective topos* [34], and *partial combinatory algebras* [7, 17, 49]. This suggests that all terms are included but equality of arbitrary lambda terms is not computable [4, page 519].

By contrast, other semantics do not account for terms without normal form or for open terms, and so avoid the need to assign them values. For example, *axiomatic recursion theory* [65, 43] uses Kleene equality [46, page 327]. Again, *operational semantics* in the style of Gordon Plotkin's *structured operational semantics* [60] can limit its values to be closed terms that are, in some sense, normal, e.g. are irreducible, or in head-normal form [4], etc. Thereby, various problems caused by non-termination, such as the difficulty of defining the parallel-or function [2], do not arise. In particular, it is easy to represent equality.

Thus, we see the challenge is to extend standard calculi, such as λ -calculus, with the general ability to query internal structure, and this while retaining many of the attractive features of λ -calculus, such as being a rewriting system, especially one that is confluent or typable. The resulting calculi are expressed, not as a variant of λ -calculus, but as an improvement upon it. As a result, we are routinely criticised for violating the Church-Turing Thesis, which motivated the production of this paper. Pure pattern calculus is an example of this approach; two more follow below.

A.10 SF-Calculus

Like λ -calculus, the traditional combinatory logic, the SK -calculus, is extensional. The SF -calculus [38] supports intensionality by replacing the operator K of SK -calculus with a *factorisation operator* F that is able to query the internal structure of terms in closed normal form. Such terms are either operators, also called atoms, or irreducible applications, called compounds. Being irreducible, it is perfectly safe to split such a compound PQ into its components P and Q , which reveals their internal structure. It cannot be stressed too much that not every application is a compound. For example, $FFFF$ is not a compound as it reduces to the atom F . In this manner, closed normal forms can be completely analysed into their constituent operators, whose equality can be tested by extensional means. Details of the calculus can be found in the original paper [38].

Figure A.3: Recoding is *SF*-definable but not λ -definable

The approach supports definable equality of closed normal forms, typed self-interpreters [40] (see also [61]), and Guy Steele’s approach [70] to growing a language [41]. It can also be extended to a concurrent setting [27, 26]. Since the factorisation operator F cannot be defined in terms of S and K , the *SF*-calculus is strictly more expressive than *SK*-calculus. Indeed, *SF*-calculus is complete in the following sense.

Theorem 77 *The normal model of computability for *SF*-calculus is equivalent to the recursive model.*

Proof. That Gödelisation and Church encoding are both simulations follows from the work of Church [15] and Kleene [45], so that it is enough to show that both recodings are computable. It is easy to see that the recoding of numbers to numbers is recursive. In the other direction, the recoding of *SF*-combinators can be described by a pattern-matching function that acts on the combinators in normal form. Such pattern-matching functions are represented by *SF*-combinators because *SF*-calculus is *structure complete* [38]. Note that this proof does not apply for *SK*-calculus as this is merely *combinatorially complete* [73]. Also, since the recodings are computably invertible, this is an equivalence. \square

The situation can be summarised in Figure A.3, where *SF* denotes the closed normal forms of *SF*-calculus.

This makes *SF*-calculus a superior foundation for higher-order programming languages compared to, say, *SK*-calculus. The main thing that is missing is native support for λ -abstraction, which is the subject of current research.

Ongoing work has developed λ *SF*-calculus [37], which combines λ -calculus and *SF*-calculus, so that one may query λ -abstractions. Indeed, Gödelisation of arbitrary closed normal forms, including λ -abstractions, is definable within the calculus.

This calculus, or something like it, may well provide a universal calculus in which to support all of the variants of λ -calculus, the different programming styles, and all the machinery necessary for their implementation. However, we are getting ahead of ourselves. For now, the point is that such a calculus may well exist, but we will never look for it if we believe that λ -calculus gives a complete account of computation.

A.11 Conclusions

The Church-Turing Thesis has been a confusion since it was first named, but not defined, by Kleene in 1952. The numerical results of Church and Turing support numerical versions of their eponymous theses, in which sets of numerical functions are co-extensive. Further, there are separate theses for symbolic computation, involving simulations of one model of computability in another. Kleene confused these two settings, with a little encouragement from Church.

Once the role of simulations is made explicit, it is easier to see that mutual simulation yields an equivalence of models only if both re-codings are computable, each in its respective model. This requirement exposes the limitations of λ -calculus, since Gödelisation is not λ -definable, even for closed λ -terms in normal form. λ -definability is not, in this sense, equivalent to Turing computability.

These limitations are, in some sense, well known within the λ -calculus community, in that λ -calculus cannot define equality, even of closed normal forms. Indeed, those working with categorical models of computability, or analysing programs defined as λ -terms, are acutely aware of these limitations. However, the community as a whole manages to work around these limitations without giving up on the Church-Turing Thesis. Overall, this is easier to do when the thesis has no canonical form, and is labeled a thesis, and not a theorem. Students who ask awkward questions may be told “Beware the Turing tarpit!” [59] or “Don’t look under the lambda!” or “You’re asking the wrong question!” which closes off discussion without clarifying anything.

The limitations of λ -calculus are essential to its nature, since λ -terms cannot directly query the internal structure of their arguments; the expressive power of λ -calculus is extensional. This support for abstraction barriers is often, and rightly, held up as one of its key strengths. Equally, its ability to query the internal structure of natural numbers, e.g. to compute predecessors, is also celebrated. However, its inability to query internal structure in a general way goes unremarked.

Rather, intensional computation requires a fresh outlook. The simplest illustration of this is the SF -calculus whose factorisation operator F is able to uniformly decompose normal forms to their constituent operators. Since SF -calculus also has all of the expressive power of SK -calculus, its normal model of computability is equivalent to the Turing model or the recursive function model. Current work on λSF -calculus suggests we may be able to define a calculus that, unlike λ -calculus, is universal, in the same way that there is a universal Turing machine.

The implications of this for programming language design are profound. The **bondi** programming language has already shown how the usual database queries can be made polymorphic, and that object-orientation can be defined in terms of pattern-matching. Now the factorisation operator paves the way for program analysis to be conducted in the source language, so that growing a language can become easier than ever.

In short, confusion in the Church-Turing Thesis has obscured the fundamental limitations of λ -calculus as a foundation for programming languages. It is time to wind up Landin’s research program, and pursue the development of intensional calculi and programming languages.

Acknowledgments We thank Jacques Carette, Robin Cockett, John Crossley, Nachum

Dershowitz, Thomas Given-Wilson, Neil Jones, Jens Palsberg, Reuben Rowe, Peter Selinger, and Eric Torreborre for comments on drafts of this paper.

Bibliography

- [1] Martin Abadi and Luca Cardelli. *A Theory of Objects*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition, 1996.
- [2] Samson Abramsky, Radha Jagadeesan, and Pasquale Malacaria. Full abstraction for PCF. *Information and Computation*, 163(2):409–470, December 2000.
- [3] Sergei Artemov and Eduardo Bonelli. The intensional lambda calculus. In Sergei N. Artemov and Anil Nerode, editors, *Logical Foundations of Computer Science*, volume 4514 of *Lecture Notes in Computer Science*, pages 12–25. Springer Berlin Heidelberg, 2007.
- [4] Henk P. Barendregt. *The Lambda Calculus — Its Syntax and Semantics*, volume 103 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, Amsterdam, revised edition, 1984.
- [5] Henk P. Barendregt and Giulio Manzonetto. Turing’s contributions to lambda calculus. In B. Cooper and J. van Leeuwen, editors, *Alan Turing — His Work and Impact*, pages 139–143. Elsevier, Boston, 2013.
- [6] Ulrich Berger and Helmut Schwichtenberg. An inverse of the evaluation functional for typed λ -calculus. In Gilles Kahn, editor, *IEEE Symposium on Logic in Computer Science (LICS)*, pages 203–211. IEEE Computer Society Press, July 1991.
- [7] Inge Bethke. *Notes on Partial Combinatory Algebras*. PhD thesis, Universiteit van Amsterdam, 1988.
- [8] Udi Boker and Nachum Dershowitz. Comparing computational power. *Logic Journal of the IGPL*, 4(5):633–647, 2006.
- [9] bondi programming language. bondi.it.uts.edu.au/, 2014.
- [10] Ana Bove, Peter Dybjer, and Ulf Norell. A brief overview of Agda — a functional language with dependent types. In *Proceedings of the 22nd International Conference on Theorem Proving in Higher Order Logics*, TPHOLs ’09, pages 73–78, Berlin, Heidelberg, 2009. Springer-Verlag.
- [11] L.E. Jan Brouwer. *Brouwer’s Cambridge Lectures on Intuitionism*. Cambridge University Press, Cambridge ; New York, 1981. Edited by Dirk van Dalen.

- [12] Jacques Carette and Aaron Stump. Towards typing for small-step direct reflection. In *Proceedings of the ACM SIGPLAN 2012 Workshop on Partial Evaluation and Program Manipulation*, PEPM ’12, pages 93–96, New York, NY, USA, 2012. ACM.
- [13] Carlo Cellucci. Proof theory and theory of meaning. In MariaLuisa Dalla Chiara, editor, *Italian Studies in the Philosophy of Science*, volume 47 of *Boston Studies in the Philosophy of Science*, pages 13–29. Springer Netherlands, Amsterdam, 1981.
- [14] Donald D. Chamberlin and Raymond F. Boyce. Sequel: A structured English query language. In *Proceedings of the 1974 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control*, SIGFIDET ’74, pages 249–264, New York, NY, USA, 1974. ACM.
- [15] Alonzo Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2):pp. 345–363, 1936.
- [16] Alonzo Church. *The calculi of lambda conversion*. Number 6 in Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, USA, 1985.
- [17] J. Robin B. Cockett and Pieter J. W. Hofstra. Introduction to Turing categories. *Annals of Pure and Applied Logic*, 156(2-3):183–209, 2008.
- [18] J. Robin B. Cockett and Pieter J. W. Hofstra. Categorical simulations. *Journal of Pure and Applied Algebra*, 214(10):1835–1853, 2010.
- [19] Olivier Danvy. Type-directed partial evaluation. In *Proceedings of the 23rd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL ’96, pages 242–257, New York, NY, USA, 1996. ACM.
- [20] Nicolas de Bruijn. A survey of the project AUTOMATH. In Jonathon P. Seldin and John R. Hindley, editors, *To H.B. Curry: Essays in Combinatory Logic, Lambda Calculus and Formalism*, pages 579–606. Academic Press, Amsterdam, 1980.
- [21] Solomon Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49(1):35–92, 1960.
- [22] Solomon Feferman. Turing’s thesis. *Notices of the American Mathematical Society*, 53:1200–1206, 2006.
- [23] Matthias Felleisen. On the expressive power of programming languages. In Neil Jones, editor, *ESOP ’90*, volume 432 of *Lecture Notes in Computer Science*, pages 134–151. Springer Berlin, Heidelberg, 1990.
- [24] Gottlob Frege. On sense and reference. In Peter Geach and Max Black, editors, *Translations from the Philosophical Writings of Gottlob Frege*, pages 56–78. Basil Blackwell, Oxford, 1960. Originally published in 1892 as “Über Sinn und Bedeutung.” Translated by Max Black.

- [25] Robin Gandy. Church's thesis and principles for mechanisms. In K. Jon Barwise, H. J. Keisler, and K. Kunen, editors, *The Kleene Symposium*, pages 123–148. North-Holland, Amsterdam, 1980.
- [26] Thomas Given-Wilson. An intensional concurrent faithful encoding of Turing machines. In Ivan Lanese, Alberto Lluch Lafuente, Ana Sokolova, and Hugo Torres Vieira, editors, Proceedings 7th Interaction and Concurrency Experience, Berlin, Germany, 6th June 2014, volume 132 of *Electronic Proceedings in Theoretical Computer Science*, pages 19–35. Open Publishing Association, 2014.
- [27] Thomas Given-Wilson, Daniele Gorla, and Barry Jay. A Concurrent Pattern Calculus. *Logical Methods in Computer Science*, 10(3), 2014.
- [28] J. A. Goguen, J. W. Thatcher, E. G. Wagner, and J. B. Wright. Initial algebra semantics and continuous algebras. *Journal of the ACM*, 24(1):68–95, January 1977.
- [29] Mayer Goldberg. Gödelization in the lambda calculus. *Information Processing Letters*, 75(1):13–16, 2000.
- [30] Robert Harper. *Practical Foundations for Programming Languages*. Cambridge University Press, New York, NY, USA, 2012.
- [31] James L. Hein. *Discrete Structures, Logic, and Computability*. Jones and Bartlett Publishers, Inc., USA, 3rd edition, 2010.
- [32] J Roger Hindley and Jonathan P Seldin. *Introduction to Combinators and [lambda]-calculus*, volume 1. CUP Archive, 1986.
- [33] Jerry R. Hobbs and Stanley J. Rosenschein. Making computational sense of Montague's intensional logic. *Artificial Intelligence*, 9:287–306, 1978.
- [34] J. Martin Hyland. The effective topos. In A.S. Troelstra and D. van Dalen, editors, *The L. E. J. Brouwer Centenary Symposium Proceedings of the Conference held in Noordwijkerhout*, volume 110 of *Studies in Logic and the Foundations of Mathematics*, pages 165–216. Elsevier, Amsterdam, 1982.
- [35] Barry Jay. The pattern calculus. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 26(6):911–937, November 2004.
- [36] Barry Jay. *Pattern Calculus: Computing with Functions and Structures*. Springer, 2009.
- [37] Barry Jay. Programs as data structures in lambda-SF-calculus. Proceedings of MFPS XXXII, May 2016. To appear.
- [38] Barry Jay and Thomas Given-Wilson. A combinatory account of internal structure. *Journal of Symbolic Logic*, 76(3):807–826, 2011.
- [39] Barry Jay and Delia Kesner. First-class patterns. *Journal of Functional Programming*, 19(2):191–225, 2009.

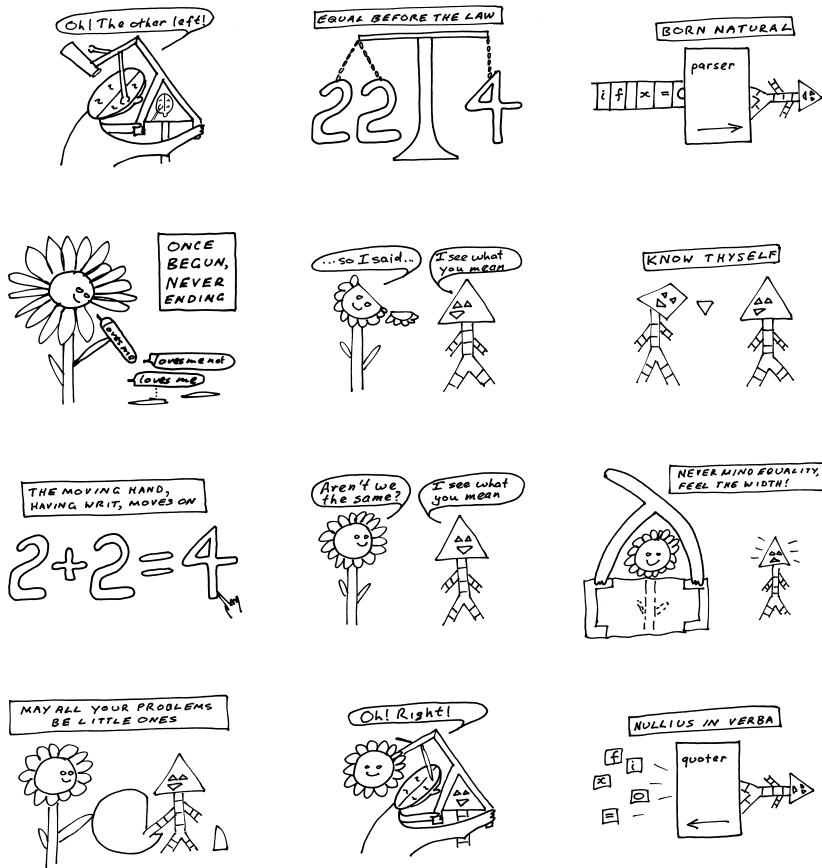
- [40] Barry Jay and Jens Palsberg. Typed self-interpretation by pattern matching. In *Proceedings of the 2011 ACM Sigplan International Conference on Functional Programming*, pages 247–58, 2011.
- [41] Barry Jay and Jose Vergara. Growing a language in pattern calculus. In *Theoretical Aspects of Software Engineering (TASE), 2013 International Symposium on*, pages 233–240. IEEE, 2013.
- [42] Barry Jay and Jose Vergara. Conflicting accounts of λ -definability. *Journal of Logical and Algebraic Methods in Programming*, 87:1 – 3, 2017.
- [43] Neil D. Jones. *Computability and Complexity: From a Programming Perspective*. MIT Press, Cambridge, MA, USA, 1997.
- [44] Jerrold J. Katz. The end of Milleanism: Multiple bearers, improper names, and compositional meaning. *The Journal of Philosophy*, 98(3):137–166, 2000.
- [45] Stephen C. Kleene. λ -definability and recursiveness. *Duke Mathematical Journal*, 2:340–353, 1936.
- [46] Steven C. Kleene. *Introduction to Metamathematics*. Bibliotheca Mathematica. North-Holland Publishing Company, Amsterdam, 1952. Co-publisher: Wolters-Noordhoff; 8th revised ed.1980.
- [47] Georg Kreisel. Some reasons for generalizing recursion theory. In *Logic Colloquium '69*, pages 139–198, Amsterdam, 1969. North-Holland Publishing Co.
- [48] Peter J. Landin. The Next 700 Programming Languages. *Communications of the ACM*, 9:157–166, 1966.
- [49] John Longley. Computability structures, simulations and realizability. *Mathematical Structures in Computer Science*, 24(02):e240201, 2014.
- [50] John Longley and Dag Normann. *Higher-Order Computability*. Theory and Applications of Computability. Springer, 2015.
- [51] Per Martin-Löf. *Intuitionistic Type Theory*. Bibliopolis, Napoli, 1984.
- [52] John McCarthy. History of Lisp. *SIGPLAN Notices*, 13(8):217–223, August 1978.
- [53] Robin Milner, Mads Tofte, and David MacQueen. *The Definition of Standard ML*. MIT Press, Cambridge, MA, USA, 1997.
- [54] John C. Mitchell. On abstraction and the expressive power of programming languages. *Science of Computer Programming*, 21(2):141–163, 1993.
- [55] John C. Mitchell and Gordon D. Plotkin. Abstract types have existential type. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 10(3):470–502, July 1988.

- [56] Torben Æ. Mogensen. Gödelization in the untyped lambda-calculus. In Olivier Danvy, editor, *Proceedings of the 1999 ACM SIGPLAN Workshop on Partial Evaluation and Semantics-Based Program Manipulation, San Antonio, Texas, USA, January 22-23, 1999. Technical report BRICS-NS-99-1*, pages 19–24. University of Aarhus, 1999.
- [57] Richard Montague. Towards a general theory of computability. *Synthese*, 12(4):429–438, 1960.
- [58] Reinhard Muskens. Intensional models for the theory of types. *Journal of Symbolic Logic*, 72(1):98–118, 2007.
- [59] Alan J. Perlis. Special feature: Epigrams on programming. *SIGPLAN Notices*, 17(9):7–13, September 1982.
- [60] Gordon D. Plotkin. The origins of structural operational semantics. *Journal of Logic and Algebraic Programming*, 60:3–15, 2004.
- [61] Andrew Polonsky. Axiomatizing the quote. In Marc Bezem, editor, *Computer Science Logic (CSL'11) — 25th International Workshop/20th Annual Conference of the EACSL*, volume 12 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 458–469, Dagstuhl, Germany, 2011. Schloss Dagstuhl—Leibniz-Zentrum für Informatik.
- [62] Emil L. Post. Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics*, 65(2):pp. 197–215, 1943.
- [63] Michael Rescorla. The representational foundations of computation. *Philosophia Mathematica*, pages 338–366, 2015.
- [64] John C. Reynolds. The essence of ALGOL. In Peter W. O’Hearn and Robert D. Tennent, editors, *ALGOL-like Languages, Volume 1*, pages 67–88. Birkhauser Boston Inc., Cambridge, MA, USA, 1997.
- [65] Hartley Jr. Rogers. *Theory of Recursive Functions and Effective Computability*. MIT Press, Cambridge, MA, USA, 1987.
- [66] Dana S. Scott. Data types as lattices. *SIAM Journal on Computing*, 5:522–587, 1976.
- [67] Dana S Scott. λ -calculus: Then & now. In *ACM Turing Centenary Celebration*, page 9. ACM, 2012.
- [68] Robert I. Soare. Computability and recursion. *The Bulletin of Symbolic Logic*, 2(3):pp. 284–321, 1996.
- [69] Robert I. Soare. Chapter 1 The history and concept of computability. In Edward R. Griffor, editor, *Handbook of Computability Theory*, volume 140 of *Studies in Logic and the Foundations of Mathematics*, pages 3–36. Elsevier, 1999.

- [70] G.L. Steele. Growing a language. *Higher-Order and Symbolic Computation*, 12(3):221–236, 1999.
- [71] G. J. Sussman and G. L. Steele. SCHEME: An interpreter for extended lambda calculus. AI Memo 349, MIT Artificial Intelligence Laboratory, December 1975.
- [72] Carolyn Talcott. *The essence of Rum — A Theory of Intensional and Extensional Aspects of Lisp Computation*. PhD thesis, Stanford University, August 1985.
- [73] Terese. *Term Rewriting Systems*, volume 55 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge U.K, 2003.
- [74] Pavel Tichý. Intension in terms of Turing machines. *Studia Logica*, 24(1):7–21, 1969.
- [75] Alan M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42(2):230–265, 1936.
- [76] G. Graham White. Pluralism ignored: The Church-Turing thesis and philosophical practice. In Arnold Beckmann, Erzsébet Csuhaj-Varjú, and Klaus Meer, editors, *Language, Life, Limits: 10th Conference on Computability in Europe, CiE 2014, Budapest, Hungary, June 23-27, 2014. Proceedings*, pages 373–382, Cham, 2014. Springer International Publishing.
- [77] Song Y. Yan. *An Introduction to Formal Languages and Machine Computation*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2000.

Appendix B

Shasta and Fir



Glossary

μ -recursion is a class of general recursive functions that act on natural numbers. Since the functions are not numbers μ -recursive functions cannot be applied to other such without elaborate encoding and decoding.

abstraction with respect to some variable x converts a term t into a function of x . Although the most common realisation of this is in lambda calculus, the term form for λ -abstraction does not appear in any of the BNFs in this work. Either the constructions $[x]t$ or $\lambda^*x.t$ are used to specify combinations, or the operator A is used to build abstractions in VAR -calculus..

atoms are the indivisible combinations, here synonymous with operators.

Backus-Naur Form (BNF) is a notation for describing inductively-defined classes of terms or expressions, such as trees, numbers, combinations, programs or types.

binary trees are trees in which each node has zero, one or two branches. In tree calculus these are the values.

breadth-first evaluation (similar to lazy evaluation) requires that evaluation at a node proceed as soon as possible, so that its branches are evaluated only if necessary.

calculus (plural “calculi”) is any system for doing calculations.

canonical forms are a well-specified sub-class of expressions, e.g. given by a BNF, that are the intended results of computation. When these are the irreducible terms of a rewriting system then they are also known as the normal forms.

combinations are expressions built from operators by application, such as combinators.

combinators are functions obtained by abstracting with respect to all variables in expressions built from variables by application. This class of functions can be built as combinations of a small set of operators, such as S and K .

compilers convert a program text or description into executable instructions.

compounds are combinations in which the leading operator does not have enough arguments to be evaluated, e.g. SMN where S requires three arguments.

computation combines a program with its inputs to calculate the output. In the calculi of this work, all combinations are computations.

confluence is a property of rewriting systems which ensures that each expression has at most one normal form.

dependent types blur the distinction between types and terms. For example, a term may be of array type which depends on, which is built using, a term that represents the length of the array.

depth-first evaluation (similar to eager evaluation), requires that all branches of a node be evaluated before any evaluation of the node itself.

equational reasoning is driven by the ability to replace equals by equals in any context, a property which fails in some models of computation.

evaluation strategies constrain the application of rewriting rules, usually to ensure that evaluation becomes deterministic. Examples include depth-first evaluation and breadth-first evaluation.

extensional programs do not query the internal structure of their arguments. Extensional functions can be identified with the combinators.

extensions are pattern-matching functions that extend a default function with a new case.

factorable forms are trees that are either a leaf, stem or fork.

factorisation is used to support divide-and-conquer strategies, where the compounds are divided and the atoms are conquered.

forks are nodes that have exactly two branches.

Gödel numbering is used to convert any symbolic expression into a natural number.

general recursion describes a large class of recursive functions; see also μ -recursion.

inputs see **values**.

intensional programs may query the internal structure of their arguments, e.g. to determine if they are atoms or compounds, or to perform triage.

interpreters act on a program paired with its inputs to produce some output. The program may be a term in a calculus or some representation of it obtained by quotation. Often the output is the same as that obtained by evaluating the program directly but if the program has been quoted then the output may be no more than the result of unquoting, to recover the original term.

intuitionism expects all mathematics to be constructive so that, for example, a proof of existence must include the ability to construct a witness that has the desired property.

irreducible terms cannot be reduced, as no sub-term instantiates the left-hand side of a rewriting rule.

judgments are the premises and conclusions of formal systems for, say, performing evaluation according to some strategy.

kernel see **leaves**.

leaves are nodes that have no branches.

meaningful translations preserve the process of evaluation as well as the results.

normal forms are canonical forms that are irreducible.

operators are indivisible constants used to build up combinations.

output see **values**.

pattern calculus is a family of calculi that base computation on pattern matching. This supports generic queries of data structures but not of pattern-matching functions and so cannot support reflection.

programs have been variously identified with their text or syntax, the executable code produced by a compiler, a function from input to outputs, or a term in some calculus. In the technical chapters of this work, a program is identified with a value in some calculus, e.g. a binary tree.

quotation is used to represent functions and computations as data structures.

recursion is the ability of a function or program to call itself during evaluation with inputs that have been freshly calculated; see also general recursion.

reflective programs are intensional programs that are designed to query the internal structure of other programs, including themselves.

rewriting systems perform computation by applying their given rewriting rules to terms, in the hope of simplifying them to a canonical form.

self-evaluators are a particular kind of self-interpreter that may exist when programs and their inputs are irreducible. They evaluate the program according to some strategy, such as depth-first or breadth-first evaluation, so that evaluation is deterministic, even though evaluation in the calculus itself is non-deterministic.

self-interpreters are interpreters in which the language or calculus of the interpreter is the same as that of the program.

serialisation converts computations into natural numbers (or strings of bits) that can be communicated to another computer.

stems are nodes that have exactly one branch.

syntax trees commonly arise after parsing of an input string to determine the intended structure.

tagging of a function with a tag allows the tag to change the intensional behaviour of the term without changing the extensional behaviour of the function.

terms are expressions built from variables and operators by applications, and so include the combinations. The variables serve as placeholders for unknown terms.

triage performs three-way case analysis on binary trees.

Turing complete systems are able to compute the same class of functions as the Turing machines, as discussed in the appendix.

Turing machines compute by using a transition table to determine the next action as a function of the current state and the current symbol on the tape. Since the transition table is not a tape, Turing machines cannot be applied to each other without elaborate encoding and decoding.

types are formal systems that are used to classify programs.

values are the permissible results of computations. In rewriting systems, they are usually the irreducible terms or normal forms. In the calculi of this work, the values are identified with the programs, their inputs and their outputs. In tree calculus, the values are the binary trees; trees with three or more branches are computations.

variables are used in two closely related ways: as placeholders for unknown terms, say in evaluation rules; or to support abstraction. In the terms of tree calculus, the variables are drawn from some unspecified, infinite class, the only obligation being that their equality should be decidable. In *VAR*-calculus, the variables are represented as indices built using the operator *V*.

Index

- addition, 48
- Backus-Naur Form, 14
- binary arithmetic, 12
- binary tree, 27
- body, 61
- booleans, 32
- bracket abstraction, 41, 93
- Church, 4, 92
- closures, 104
- combinatorially complete, 42, 94
- combinatory algebra, 7
- combinatory logic, 6, 28
- compiler, 4, 127
- computational equality, 78
- confluence, 84
- Curry, 92
- data structure, 4, 127
- default function, 61
- derivation, 14
- extensional function, 40
- extensions, 61
- factorable, 64
- factorisation, 28
- factorisation calculus, 6
- fork, 27
- functional programming, 92
- Gödel, 4
- identities, 95
- interpreter, 4
- irreducible, 80
- judgment, 14
- kernel, 27
- Kleene, 7
- lambda-calculus, 6
- leaf, 27
- logic, 92
- natural numbers, 27
- natural tree, 4
- one-step reduction, 79
- pattern, 61
- pentagon, 95
- program, 27
- quotation, 4, 24
- reduction rules, 78
- reflection, 4, 24
- Roman numerals, 12
- Russell's Paradox, 92
- Schönfinkel, 92
- self-evaluator, 5
- self-interpreter, 5
- separator, 95
- serialisation, 4
- simultaneous reduction, 85
- star abstraction, 43, 94
- stem, 27
- structural equality, 78
- substitution, 41
- syntax tree, 4, 127
- tagging, 5, 6
- tallies, 12
- ternary arithmetic, 12

triage, 60
Turing, 4
Turing complete, 6
types, 92

VA-calculus, 6
value, 27
variables, 40