

Linked visualisations via Galois dependencies

ROLY PERERA, The Alan Turing Institute, UK

MINH NGUYEN, University of Bristol, UK

TOMAS PETRICEK, University of Kent, UK

MENG WANG, University of Bristol, UK

We present new language-based dynamic analysis techniques for linking visualisations and other structured outputs to data in a fine-grained way, allowing users to explore how data attributes and visual or other output elements are related by selecting (focusing on) substructures of interest. Our approach builds on bidirectional program slicing techniques based on Galois connections, which provide desirable round-tripping properties. Unlike the prior work, our approach allows selections to be negated, equipping the bidirectional analysis with a De Morgan dual which can be used to link different outputs generated from the same input. This offers a principled language-based foundation for a popular view coordination feature called *brushing and linking* where selections in one chart automatically select corresponding elements in another related chart.

1 INTRODUCTION

Techniques for dynamic dependency analysis have been fruitful, with applications ranging from information-flow security [30] and optimisation [18] to debugging and program comprehension [33, 10]. There are, however, few methods suitable for fine-grained analysis of richly structured outputs, such as data visualisations and multidimensional arrays. Dataflow analyses [27] tend to focus on analysing variables rather than parts of structured values. Where-provenance [7] and related data provenance techniques are fine-grained, but are specific to relational query languages. Taint tracking [21] is also fine-grained, but works forwards from input to output. For many applications, it would be useful to be able to focus on a particular part of a structured output, and have an analysis isolate the input data pertinent only to that substructure.

This is a need that increasingly arises outside of traditional programming. Journalists and data scientists use programs to compute charts and other visual summaries from data, charts which must be interpreted by colleagues, policy makers and lay readers alike. Interpreting a chart correctly means understanding what the components of the visualisation actually *represent*, i.e. the mapping between data and visual elements. But this is a hard task, requiring time and expertise, even with access to the data and source code used to create the visualisation. It is easy for innocent (but devastating) mistakes such as transposing two columns of data to go unnoticed [20]. Since visualisations are simply cases of programs that transform structured inputs (data tables) into structured outputs (charts and other graphics), general-purpose language-based techniques for fine-grained dependency tracking should be able to help with this, by making it possible to reveal these relationships automatically to an interested user.

1.1 Linking structured outputs to structured inputs

First, interpreting a chart would be much easier if the user were able to explore the relationship between the various parts of the chart and the underlying data interactively, discovering the relevant relationships on a need-to-know basis. For example, selecting a particular bar in a bar chart could highlight the relevant data in a table, perhaps showing only the relevant rows, as illustrated

Authors' addresses: Roly Perera, The Alan Turing Institute, London, UK, rperera@turing.ac.uk; Minh Nguyen, min.nguyen@bristol.ac.uk, University of Bristol, Bristol, UK; Tomas Petricek, University of Kent, Canterbury, UK, tpetricek@kent.ac.uk; Meng Wang, meng.wang@bristol.ac.uk, University of Bristol, Bristol, UK.

2021. 2475-1421/2021/1-ART1 \$15.00

<https://doi.org/>

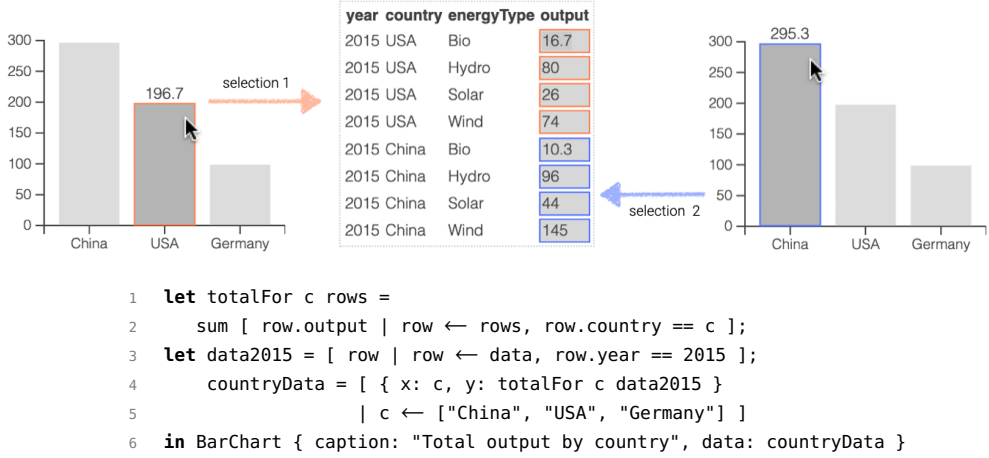


Fig. 1. Fine-grained linking of outputs to inputs, focusing on data for USA (left) and China (right).

in Figure 1. We could certainly do more and say something about the nature of the relationship (summation, in this case), but even just revealing the relevant data puts a reader in a much better position to fact-check or confirm their own understanding of what they are looking at. (The figure shows how selecting the bar for the USA should highlight different data than selecting the bar for China.) Indeed, this is useful enough that visualisation designers sometimes create “data-linked” artefacts like these by hand, such as Nadieh Bremer’s award-winning visualisation of population density growth in Asian cities [6], at the cost of significant programming effort. Libraries such as Altair [31] alleviate some of this work, but require data transformations to be specified using a limited set of combinators provided (and understood) by the library.

What we would like to do is allow data scientists to author analyses and visualisations using an expressive functional language like the one shown in Figure 1, and obtain data linking automatically for the generated artefact, as a baked-in transparency feature. At the core of this is a program analysis problem: we want to be able to focus on a particular visual attribute — say the value of y in the record $\{x: \text{"USA"}, y: 196.7\}$ passed to `BarChart` in the example above — and perform some kind of backwards analysis to determine the relevant inputs, in this case the value of output in four of the records that appear in the data source. Framing this as a program analysis problem not only provides a path to automation, but also invites interesting questions that a hand-crafted solution is unlikely to properly address. For example, does the union of two output selections depend on the union of their respective dependencies? Do dependencies “round-trip”, in that they identify sufficient resources to reconstruct the selected output? Are they minimal? These questions are important to establishing trust, and a language-based approach offers a chance to address them.

1.2 Linking structured outputs to other structured outputs

Second, authors often present distinct but related aspects of data in separate charts. In this situation a reader should be able to focus on (select) a visual element in one chart or other structured output and automatically see elements of a different chart which were computed using related inputs. For example in Figure 2 below, selecting the bar on the left should automatically highlight all the related visual elements on the right. This is a well-recognised use case called *brushing and linking* [3], which is supported by geospatial applications like GeoDa [2] and charting libraries like Plotly, but tends to be baked into specific views, or require programmer effort and therefore anticipation in

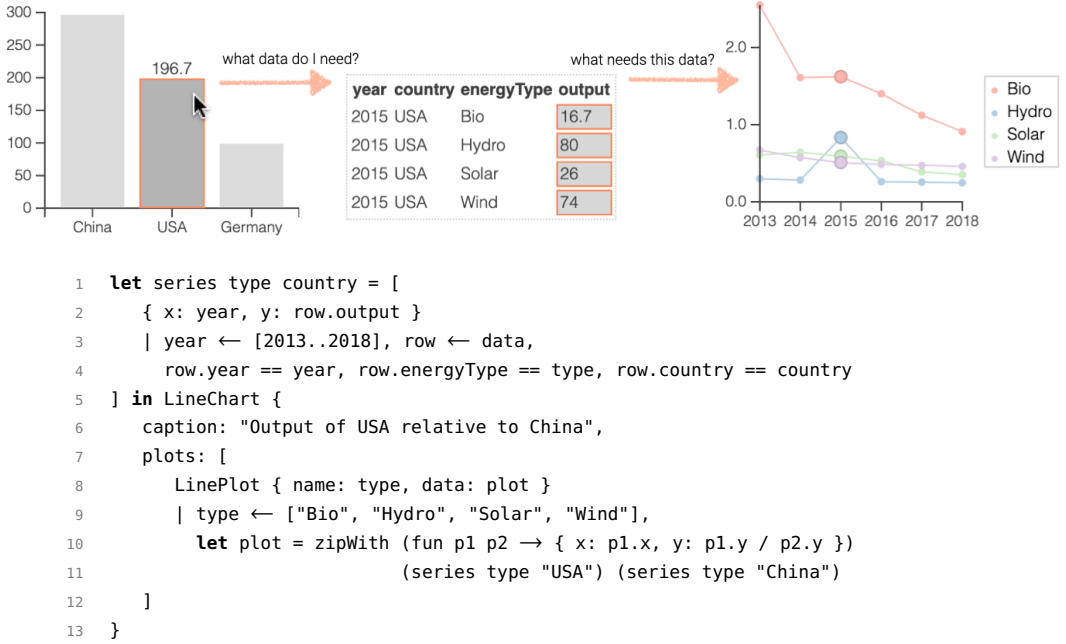


Fig. 2. Linking visualisations via common data dependencies

advance by the chart designer. Moreover these applications and libraries provide no direct access to the common data which explains why elements are related.

Again, we would like to enable a more automated (and ubiquitous) version of brushing and linking, without imposing a burden on the programmer. They should be able to express visualisations and other data transformations using standard functional programming features such as those shown in Figure 2, and have brushing and linking enabled automatically between computed artefacts which depend on common data. At the core of this requirement is a variant of our original program analysis problem: we want to select part of the output and perform a backwards analysis to identify the required inputs, as before, but then also perform a forwards analysis to identify the dependent parts of the other output. In Figure 2, these consist of the value of y for the record passed to `LinePlot` where x has the value 2015, for each `LinePlot` in the list passed to `LineChart`. Moreover, we would also like the brushing and linking feature to be able to provide a concise view of the data that explain why the two selections are linked. Note, however, that the intuition behind the forwards analysis here is not the same as the one we appealed to in the context of round-tripping: there the (hypothetical) question was whether the selected data was *sufficient* to reconstruct the selected output, whereas to identify related items in another view, we must determine those parts for which the selected data is *necessary*. As before, a language-based approach offers the prospect of addressing these sorts of question in a robust way.

1.3 Contributions

To make progress towards these challenges, we present a bidirectional analysis which tracks fine-grained data dependencies between input and output selections, with round-tripping properties characterised by Galois connections. Selections have a complement, which we use to adapt the analysis to compute fine-grained dependencies between two outputs which depend on common

inputs. Recent program slicing techniques [23, 24, 28] allow the user to focus on the output by “erasing” parts deemed to be irrelevant; the erased parts, called *holes*, are propagated backwards by a backwards analysis which identifies parts of the program and input which are no longer needed. Although these approaches also enjoy useful round-tripping properties characterised by Galois connections, they only allow focusing on *prefixes* (the portion of the output or program that remains after the irrelevant parts have been erased), a notion which is not closed under complement. Our specific contributions are as follows:

- a new bidirectional dynamic dependency analysis which operates on selections of arbitrary parts of data values, for a core calculus with lists, records and mutual recursion, and a proof that the analysis is a Galois connection (§ 3);
- a second bidirectional dependency analysis, derived from the first by De Morgan duality, which is also a Galois connection and which can be composed with the first analysis to link outputs to outputs, with an extended example based on matrix convolution (§ 4);
- a richer surface language called Fluid¹, implemented in PureScript, with familiar functional programming features such as piecewise definitions and list comprehensions, and a further Galois connection linking selections between the core and surface languages (§ 5).

Proofs and other supplementary materials can be found at <https://arxiv.org/abs/2109.00445>.

2 CORE LANGUAGE

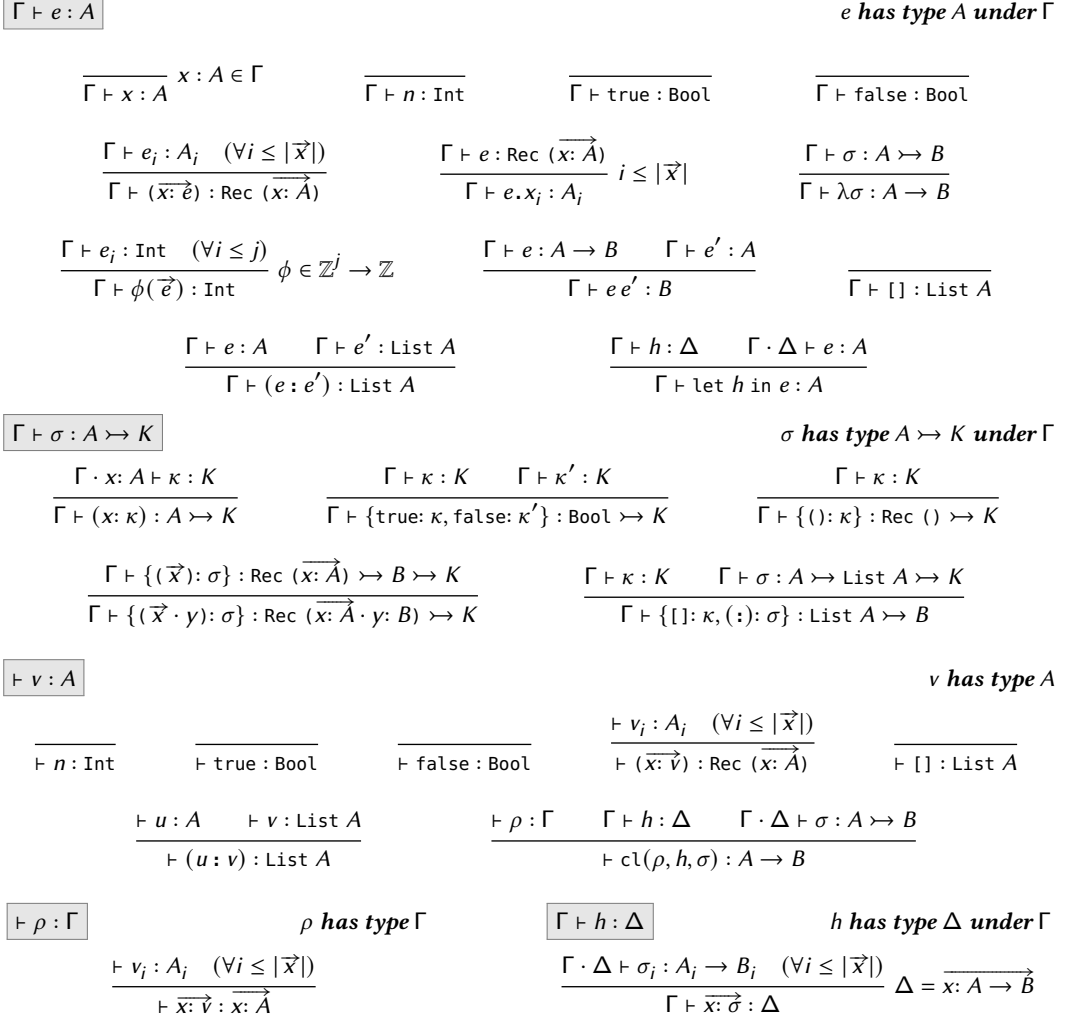
The core calculus which provides the setting for the rest of the paper is a mostly standard call-by-value functional language with datatypes and records. The main unusual feature is the use of *eliminators*, a trie-like construct that provides a uniform syntax and semantics for pattern-matching; this allows us to assume that incomplete or overlapping patterns and other syntactic considerations have been dealt in the surface language. (In § 5 we show how familiar pattern-matching features like case expressions and piecewise function definitions easily desugar into eliminators.) We give a big-step environment-based semantics, which is easier for the backward and forward dependency analyses in § 3, and introduce a compact (term-like) representation of derivation trees in the operational semantics, called *traces*, which we will use to define the analyses over a fixed execution. Mutual recursion requires some care for the backwards analysis, so we also treat that as a core language feature.

2.1 Syntax and typing

Although our implementation is untyped, types help describe the structure of the core language. Figure 3 introduces the types A, B which include `Bool`, `Int` and function types $A \rightarrow B$, but also lists `List A` and records $\text{Rec } (\vec{x} : \vec{A})$ which exemplify the two kinds of structured data which are of interest: recursive datatypes with varying structure, and tabular data with a fixed shape. As usual the notation $x : A$ denotes the binding of x to A (understood formally as a pair); $\vec{x} : \vec{A}$ denotes the sequence of bindings that results from zipping same-length sequences \vec{x} and \vec{A} . In a record type $\text{Rec } (\vec{x} : \vec{A})$ the field names in \vec{x} are required to be unique.

The terms e of the language are defined in Figure 3. These include Boolean constants `true` and `false`, integers n , variables x , and applications $e e'$. Primitives are not first-class; the expression $\phi(\vec{e})$ is the fully saturated application of ϕ to a sequence of arguments. (First-class and infix primitives are provided by desugarings in § 5). We also provide list constructors `nil []` and `cons $e : e'$` , record construction $(\vec{x} : \vec{e})$ and record projection $e.x$. The final two term forms, anonymous functions $\lambda\sigma$ and recursive let-bindings `let h in e` where h is of the form $\vec{x} : \vec{\sigma}$, are explained below

¹See <https://github.com/explorable-viz/fluid/>. To generate the figures in this paper, check out tag v0.4.2 and follow the instructions in `artifact-evaluation.md`.



2.2.1 Eliminators as functions. We can now revisit the term forms $\lambda\sigma$ and $\text{let } h \text{ in } e$. If σ is an eliminator of type $A \multimap B$, then $\lambda\sigma$ is an anonymous function of type $A \rightarrow B$. If h is of the form $\overline{x}:\vec{\sigma}$, then $\text{let } h \text{ in } e$ introduces a sequence of mutually recursive functions which are in scope in e . The typing rule for $\text{let } h \text{ in } e$ uses an auxiliary typing judgement $\Gamma \vdash h : \Delta$ which assigns to every x in Δ the function type $A \rightarrow B$ if the σ to which x is bound in h has the eliminator type $A \multimap B$.

2.2.2 Values. Values v , u , and environments ρ are also defined in Figure 3, and are standard for call-by-value. (Environments are more convenient than substitution for tracking variable usage.) To support mutual recursion, the closure form $\text{cl}(\rho, h, \sigma)$ captures the (possibly empty) sequence h of functions with which the function was mutually defined, in addition to the ambient environment ρ . For the typing judgements $\vdash \rho : \Gamma$ and $\vdash v : A$ for environments and values (Figure 4), only the closure case is worth noting, which delegates to the typing rules for recursive definitions and eliminators.

2.2.3 Evaluation. Figure 6 gives the operational semantics of the core language. In § 3 we will define forward and backward analyses over a single execution; in anticipation of that use case, we treat the operational semantics as an inductive data type, following the “proved transitions” approach of Boudol and Castellani for reversible CCS [5]. The inhabitants of this data type are derivation trees explaining how a result was computed, and the analyses will be defined by structural recursion over these trees. Expressed in terms of inference rules, these trees can become quite cumbersome, so we introduce an equivalent but more term-like syntax for them, called a *trace* (Figure 5), similar to the approach taken by Perera et al. for π -calculus [24].

Trace				
$T, U ::=$	x	variable	$\phi(\vec{U}_n)$	primitive application
	$\text{true} \mid \text{false}$	Boolean	$\text{let } h \text{ in } T$	recursive let
	n	integer	Match	
	$(\overline{x}:\vec{T})$	record	$w ::=$	x variable
	$T_{\overline{x}:\vec{v}}.y$	record projection		$\text{true} \mid \text{false}$ Boolean
	$[] \mid T : U$	list		$(\overline{x}:\vec{w})$ record
	$\lambda\sigma$	anonymous function		$[] \mid w : w'$ list
	$T U \blacktriangleright w : T'$	application		

Fig. 5. Syntax of traces and matches

The judgement $T :: \rho, e \Rightarrow v$ defined at the top of Figure 6 states that term e under environment ρ evaluates to value v , and that T is a proof term that witness that fact. (In the figure, the traces appear in grey, to reinforce the idea that they are not part of the definition of \Rightarrow but rather a notation for its inhabitants.) The rules for Booleans, integers and lists are standard and have unsurprising trace forms. For variables, we give an explicit inductive definition of the environment lookup relation \in at the bottom of the figure, again so that later we can perform analysis over a proof that an environment contains a binding. The lambda rule is standard except that we specify ε for the sequence of definitions being simultaneously defined, since a lambda is not recursive. For record construction, the trace form contains a subtrace T_i for each field, and for record projection, which also uses the lookup relation \in , the trace form $T_{\overline{x}:\vec{v}}.y$ records both the record $\overline{x}:\vec{v}$ and the field name y that was selected.

The rule for (mutually) recursive functions $\text{let } h \text{ in } e$, where h is a sequence $\overline{x}:\vec{\sigma}$ of function definitions, makes use of the auxiliary relation $\rho, h \Rightarrow \rho'$ at the bottom of Figure 6 which turns h into an environment ρ' binding each function name x_i to a closure $\text{cl}(\rho, h, \sigma_i)$ capturing ρ and a

$T :: \rho, e \Rightarrow v$		<i>T witnesses that e evaluates to v in ρ</i>	
$\Rightarrow\text{-var}$	$\Rightarrow\text{-lambda}$	$\Rightarrow\text{-true}$	$\Rightarrow\text{-false}$
$\frac{x: v \in \rho}{x :: \rho, x \Rightarrow v}$	$\frac{\lambda \sigma :: \rho, \lambda \sigma \Rightarrow \text{cl}(\rho, \varepsilon, \sigma)}{\lambda \sigma :: \rho, \lambda \sigma \Rightarrow \text{cl}(\rho, \varepsilon, \sigma)}$	$\frac{}{\text{true} :: \rho, \text{true} \Rightarrow \text{true}}$	$\frac{}{\text{false} :: \rho, \text{false} \Rightarrow \text{false}}$
$\Rightarrow\text{-int}$	$\Rightarrow\text{-record}$	$\Rightarrow\text{-project}$	
$\frac{}{n :: \rho, n \Rightarrow n}$	$\frac{T_i :: \rho, e_i \Rightarrow v_i \quad (\forall i \leq \vec{x})}{(x: \vec{T}) :: \rho, (x: \vec{e}) \Rightarrow (x: \vec{v})}$	$\frac{T :: \rho, e \Rightarrow (x: \vec{v}) \quad y: v' \in x: \vec{v}}{T_{x: \vec{v}} \cdot y :: \rho, e.y \Rightarrow v'}$	
$\Rightarrow\text{-nil}$	$\Rightarrow\text{-cons}$	$\Rightarrow\text{-apply-prim}$	
$\frac{}{[] :: \rho, [] \Rightarrow []}$	$\frac{T :: \rho, e \Rightarrow v \quad U :: \rho, e' \Rightarrow v'}{T : U :: \rho, e : e' \Rightarrow v : v'}$	$\frac{U :: \rho, e_i \Rightarrow n_i \quad (\forall i \leq j) \quad \phi \in \mathbb{Z}^j \rightarrow \mathbb{Z}}{\phi(\vec{U}_n) :: \rho, \phi(\vec{e}) \Rightarrow \phi(\vec{n})}$	
$\Rightarrow\text{-let-rec}$	$\Rightarrow\text{-apply}$		
$\frac{\rho, h \rightarrow \rho' \quad T :: \rho \cdot \rho', e \Rightarrow v}{\text{let } h \text{ in } T :: \rho, \text{let } h \text{ in } e \Rightarrow v}$	$\frac{T :: \rho, e \Rightarrow \text{cl}(\rho_1, h, \sigma') \quad \rho_1, h \rightarrow \rho_2 \quad U :: \rho, e' \Rightarrow v \quad w :: v, \sigma' \rightsquigarrow \rho_3, e'' \quad T' :: \rho_1 \cdot \rho_2 \cdot \rho_3, e'' \Rightarrow v'}{T U \blacktriangleright w: T' :: \rho, e e' \Rightarrow v'}$		
$w :: v, \sigma \rightsquigarrow \rho, \kappa$		<i>w witnesses that σ matches v and yields ρ and κ</i>	
$\rightsquigarrow\text{-true}$		$\rightsquigarrow\text{-false}$	
$\frac{}{\text{true} :: \text{true}, \{\text{true}: \kappa, \text{false}: \kappa'\} \rightsquigarrow \varepsilon, \kappa}$		$\frac{}{\text{false} :: \text{false}, \{\text{true}: \kappa, \text{false}: \kappa'\} \rightsquigarrow \varepsilon, \kappa'}$	
$\rightsquigarrow\text{-var}$		$\rightsquigarrow\text{-nil}$	
$\frac{}{x :: v, x: \kappa \rightsquigarrow x: v, \kappa}$		$\frac{}{[] :: [], \{[]: \kappa, (:): \sigma\} \rightsquigarrow \varepsilon, \kappa}$	
		$\rightsquigarrow\text{-unit}$	
		$\frac{}{() :: (), \{(): \kappa\} \rightsquigarrow \varepsilon, \kappa}$	
$\rightsquigarrow\text{-cons}$		$\rightsquigarrow\text{-record}$	
$\frac{w :: v, \sigma \rightsquigarrow \rho, \tau \quad w' :: v', \tau \rightsquigarrow \rho', \kappa'}{(w : w') :: v : v', \{[]: \kappa, (:): \sigma\} \rightsquigarrow \rho \cdot \rho', \kappa'}$		$\frac{(x: \vec{w}) :: x: \vec{v}, \sigma \rightsquigarrow \rho, \sigma' \quad w' :: u, \sigma' \rightsquigarrow \rho', \kappa}{(x: \vec{w} \cdot y: w') :: (x: \vec{v} \cdot y: u), \{(\vec{x} \cdot y): \sigma\} \rightsquigarrow \rho \cdot \rho', \kappa}$	
$x: v \in \rho$		$\rho, h \rightarrow \rho'$	
$\in\text{-head}$		$\rightarrow\text{-rec-defs}$	
$\frac{}{x: v \in (\rho \cdot x: v)}$		$\frac{v_i = \text{cl}(\rho, x: \vec{\sigma}, \sigma_i) \quad (\forall i \in \vec{x})}{\rho, x: \vec{\sigma} \rightarrow x: \vec{v}}$	
$x: v \text{ is contained by } \rho$		<i>h generates ρ' in ρ</i>	
$\in\text{-tail}$			
$\frac{x: v \in \rho}{x: v \in (\rho \cdot y: u)} \quad x \neq y$			

Fig. 6. Operational semantics

copy of h . For primitive applications, the trace records the values of the arguments which were passed to the operation ϕ . The rule for application $e e'$ is slightly non-standard, because it must deal with both mutual recursion and pattern-matching. First we unpack the recursive definitions h from the closure $\text{cl}(\rho_1, h, \sigma)$ computed by e , and again use the auxiliary relation \rightarrow to promote this into an environment ρ_2 of closures. We then use the relation \rightsquigarrow explained below to match v against the eliminator σ , obtaining the branch e'' of the function to be executed and parameter bindings ρ_3 . In addition to subtraces T and U for the function and argument, the application trace $T U \blacktriangleright w: T'$ also has subtraces w for the pattern-match and T' for the selected branch.

2.2.4 Pattern matching. The judgement $w :: v, \sigma \rightsquigarrow \rho, \kappa$ also defined in Figure 6 states that eliminator σ can match v and produce environment ρ and continuation κ , with ρ containing the variable bindings that arose during the match. *Matches* w are a compact notation for proof terms for the \rightsquigarrow relation, analogous to traces for the \Rightarrow relation, and again appear in grey in the figure.

Variable eliminators $x: \kappa$ match any value, returning the singleton environment $x: v$ and continuation κ . Boolean eliminators match any Boolean value, returning the appropriate branch and empty environment ε . List eliminators $\{[]: \kappa, (:): \sigma\}$ match any list. The nil case is analogous to the handling of Booleans; the cons case depends on the fact that the nested eliminator σ for the cons branch has the curried type $A \multimap \text{List } A \multimap K$. First, we recursively match the head v of type A using σ , obtaining bindings ρ and eliminator $\tau: \text{List } A \multimap K$ as the continuation. Then the tail v' is matched using τ to yield additional bindings ρ' and final continuation κ' of type K . As a simple example, which omits the proof terms w , consider the following pattern-match:

$$\rightsquigarrow\text{-cons} \frac{\rightsquigarrow\text{-var} \frac{}{5, x: xs: e_2 \rightsquigarrow x: 5, xs: e_2} \quad \rightsquigarrow\text{-var} \frac{}{6 : [], xs: e_2 \rightsquigarrow xs: (6 : []), e_2}}{5 : 6 : [], \{[]: e_1, (:): x: xs: e_2\} \rightsquigarrow (x: 5) \cdot (xs: 6), e_2}$$

Here the eliminator $\{[]: e_1, (:): x: xs: e_2\}$ is used to match $5 : 6 : []$. The $[]$ case is disregarded; the $(:)$ case is used to retrieve a variable eliminator $x: xs: e_2$, which is used to match the head 5. This produces the binding $x: 5$ and a further variable eliminator $xs: e_2$ as the continuation, which is used to match the tail. This produces the additional binding $xs: (6 : [])$ and the expression e_2 as the continuation. To see how this might generalise to a nested pattern, consider how one could replace the inner variable eliminator $xs: e_2$ by another list eliminator.

Record matching is similar: the empty record case resembles the nil case, and the non-empty case relies on the nested eliminator having curried type $\text{Rec } (\overline{x: \vec{A}}) \multimap B \multimap K$. The initial part $\overline{x: \vec{v}}$ of the record is matched using σ , returning another eliminator σ' of type $B \multimap K$. Then the last field $y: u$ is matched using σ' to yield final continuation κ of type K .

3 A BIDIRECTIONAL DYNAMIC DEPENDENCY ANALYSIS

We now extend the core language from § 2 with a bidirectional mechanism for tracking data dependencies. § 3.1 establishes a way of selecting (parts of) values, such as the height of a bar in a bar chart. § 3.2 defines a forward analysis function \nearrow_T which specifies how selections on programs and environments (collectively: *input selections*) become selections on outputs; selections represent *availability*, with the computed output selection indicating the data available to the downstream computation. § 3.3 defines a backward dependency function \searrow_T specifying how output selections are mapped back to inputs; then selections represent *demands*, with the computed input selection identifying the data needed from the upstream computation. Both functions are monotonic. This will become important in § 3.4, where we show that \searrow_T and \nearrow_T form a *Galois connection*, establishing the round-tripping properties alluded to in § 1.1.

3.1 Lattices of selections

Our approach to representing selections is shown in Figure 7. The basic idea is to parameterise the type Val of values by a (bounded) lattice \mathcal{A} of *selection states* α . We add selection states to Booleans, integers, records and lists; while it would present no complications to equip closures with selection states too, for present purposes we are only interested in dependencies between first-order data, so closures are not (directly) selectable. Closures do however have selectable parts, and moreover capture the current *argument availability*, explained in § 3.2.2 below, which is also a selection state α . We parameterise the type Term of terms similarly, allowing us to trace data dependencies back

to expressions that appear in the source code, but only add selection states to the term constructors corresponding to selectable values. We return to this in § 5.

<i>Terms selections</i>		<i>Value selections</i>		
$e \in \text{Term } \mathcal{A} ::=$...	$u, v \in \text{Val } \mathcal{A} ::=$	\square	hole
	\square		$\text{true}_\alpha \mid \text{false}_\alpha$	Boolean
	$\text{true}_\alpha \mid \text{false}_\alpha$		n_α	integer
	n_α		$(\vec{x} : \vec{v})_\alpha$	record
	$(\vec{x} : \vec{e})_\alpha$		$[]_\alpha \mid u :_\alpha v$	list
	$[]_\alpha \mid e :_\alpha e'$		$\text{cl}(\rho, h, \alpha, \sigma)$	closure
$\alpha, \beta \in \mathcal{A}$				selection state

Fig. 7. Selection states, term selections and value selections

The top and bottom elements \top and \perp of \mathcal{A} represent fully selected and fully unselected; the meet and join operations \sqcap and \sqcup , which have \top and \perp as their respective units, are used to combine selection information. In Figure 1, the data field of `BarChart` expects a list of records with fields `x` and `y`, mapping strings representing categorical data to floats determining the height of the corresponding bar; the record computed for China is $(x: \text{"China"} \cdot y: 295.3)$. The two-point lattice $2 \stackrel{\text{def}}{=} \langle \{\text{tt}, \text{ff}\}, \text{tt}, \text{ff}, \wedge, \vee \rangle$ can be used to represent the selection of the field `y` within this record as $(x: \text{"China"}_{\text{ff}} \cdot y: 295.3_{\text{tt}})_{\text{ff}}$, indicating that the number 295.3 is selected, but that neither the string "China", nor the record itself, is selected. Because lattices are closed under component-wise products, we sometimes write $(\alpha, \beta) \sqsubseteq (\alpha', \beta')$ to mean that $\alpha \sqsubseteq \alpha'$ and $\beta \sqsubseteq \beta'$. This also suggests more interesting lattices of selections, such as vectors of Booleans to represent multiple selections simultaneously, which might be visualised using different colours (as in Figure 1).

3.1.1 Selections of a value. The analyses which follow will be defined with respect to a fixed computation, and so we will often need to talk about the selections of a given value. To make this notion precise, consider that the raw (selection-free) syntax described in § 2 can be recovered from a term selection via an erasure operation $[\cdot] : \text{Val } \mathcal{A} \rightarrow \text{Val } 1$ which forgets the selection information, where 1 is the trivial one-point lattice. We refer to $[\nu]$ as the *shape* of ν . Allowing \mathbf{u}, \mathbf{v} from now on to range over raw values, and reserving u, v for value selections, we can then define:

Definition 3.1 (Selections of \mathbf{v}). Define $\text{Sel}_{\mathbf{v}} \mathcal{A}$ to be the set of all values $v \in \text{Val } \mathcal{A}$ with shape \mathbf{v} , i.e. that erase to \mathbf{v} .

Since its elements have a fixed shape, the pointwise comparison of any $v, v' \in \text{Sel}_{\mathbf{v}} \mathcal{A}$ using the partial order \sqsubseteq of \mathcal{A} is well defined, as is the pointwise application (zip) of a binary operation [12]. It should therefore be clear that if \mathcal{A} is a lattice, then $\text{Sel}_{\mathbf{v}} \mathcal{A}$ is also a lattice, with $\top_{\mathbf{v}}, \perp_{\mathbf{v}}, \sqcap_{\mathbf{v}}$, and $\sqcup_{\mathbf{v}}$ defined pointwise. For example, if u and u' have the same shape and v and v' have the same shape, the join of the lists $(u :_\alpha v)$ and $(u' :_{\alpha'} v')$ is defined and equal to $(u \sqcup u') :_{\alpha \sqcup \alpha'} (v \sqcup v')$. Similarly, the top element of $\text{Sel}_{\mathbf{v}} \mathcal{A}$ is the selection of \mathbf{v} which has \top at every selection position. (We omit the \mathbf{v} indices from these lattice operations if it is clear which lattice is being referred to.) The notion of the “selections” of \mathbf{v} extends to the other syntactic forms.

3.1.2 Environment selections and hole equivalence. The notion of the “selections” of \mathbf{v} also extends (pointwise) to environments, so that $\text{Sel}_{\rho} \mathcal{A}$ means the set of environment selections ρ' of shape ρ , where the variables in ρ' are bound to selections of the corresponding variables in ρ . One challenge arises from the pointwise use of \sqcup to combine environment selections. Since environments contain other environments recursively, via closures, a naive implementation of environment join is a very

$v \sqsubseteq v'$					
$\square \sqsubseteq v$	$\frac{\alpha \sqsubseteq \alpha'}{n_\alpha \sqsubseteq n_{\alpha'}}$	$\frac{}{n_\perp \sqsubseteq \square}$	$\frac{\alpha \sqsubseteq \alpha'}{\text{true}_\alpha \sqsubseteq \text{true}_{\alpha'}}$	$\frac{}{\text{true}_\perp \sqsubseteq \square}$	$\frac{\alpha \sqsubseteq \alpha'}{\text{false}_\alpha \sqsubseteq \text{false}_{\alpha'}}$
$\frac{}{\text{false}_\perp \sqsubseteq \square}$	$\frac{\alpha \sqsubseteq \alpha' \quad v_i \sqsubseteq u_i \quad (\forall i \in \vec{x})}{(\vec{x}:\vec{v})_\alpha \sqsubseteq (\vec{x}:\vec{u})_{\alpha'}}$		$\frac{v_i \sqsubseteq \square \quad (\forall i \in \vec{x})}{(\vec{x}:\vec{v})_\perp \sqsubseteq \square}$		$\frac{\alpha \sqsubseteq \alpha'}{[]_\alpha \sqsubseteq []_{\alpha'}}$
$\frac{}{[]_\perp \sqsubseteq \square}$	$\frac{\alpha \sqsubseteq \alpha' \quad v_i \sqsubseteq u_i \quad (\forall i \in \vec{x})}{(\vec{x}:\vec{v})_\alpha \sqsubseteq (\vec{x}:\vec{u})_{\alpha'}}$		$\frac{v_i \sqsubseteq \square \quad (\forall i \in \vec{x})}{(\vec{x}:\vec{v})_\perp \sqsubseteq \square}$		$\frac{\alpha \sqsubseteq \alpha'}{[]_\alpha \sqsubseteq []_{\alpha'}}$
$\frac{(\alpha, v, v') \sqsubseteq (\alpha', v, v')}{v :_\alpha v' \sqsubseteq u :_{\alpha'} u'}$	$\frac{(v, v') \sqsubseteq (\square, \square)}{v :_\perp v' \sqsubseteq \square}$	$\frac{(\rho, h, \alpha, \sigma) \sqsubseteq (\rho', h', \alpha', \sigma')}{\text{cl}(\rho, h, \alpha, \sigma) \sqsubseteq \text{cl}(\rho', h', \alpha', \sigma')}$	$\frac{(\rho, h, \sigma) \sqsubseteq (\square_\rho, \square, \square)}{\text{cl}(\rho, h, \perp, \sigma) \sqsubseteq \square}$		

Fig. 8. Preorder on value selections

expensive operation. One solution is to employ an efficient representation of values which are fully unselected, which is often the case during the backward analysis.

We therefore augment the set of value selections $\text{Val } \mathcal{A}$ with a distinguished element *hole*, written \square , which is an alternative notation for \perp_v for any v , i.e. the selection of shape v which has \perp at every selection position, and generalise this idea to terms and eliminators. The equivalence of \square to any such bottom element is established explicitly by the preorder order defined (for values) in Figure 8: the first rule always allows \square on the left-hand side of \sqsubseteq , and other rules allow \square on the right-hand side of \sqsubseteq as long as all the selections that appear on the left-hand side are \perp . (The rules for terms e and eliminators σ are analogous and are omitted.) If we write \doteq for the equivalence relation induced by \sqsubseteq on values selections, which we call *hole-equivalence*, it should be clear that $\square \sqsubseteq v \doteq v$ and $\square \sqsubseteq \square \doteq \square$. This means the join of two selections v, v' of v can be implemented efficiently, whenever one selection is \square , by simply discarding \square and returning the other selection without further processing.

Definition 3.2 (Hole equivalence). Define \doteq as the intersection of \sqsubseteq and \supseteq .

Because \square is equivalent to \perp_v for any v , all such bottom elements are hole-equivalent. For example, the value selection $\square :_\top \square$ is hole-equivalent to $5_\perp :_\top \square$, but also to $6_\perp :_\top []_\perp$, and so the last two selections, even though they have different shapes, are hole-equivalent by transitivity. In practice we only use the hole ordering to compare selections with the same shape.

3.2 Forward data dependency

We now define the core bidirectional data dependency analyses for a fixed computation $T :: \rho, e \Rightarrow v$, where T is a trace. In practice one would obtain T by first evaluating e in ρ , and then run multiple forward or backward analyses over T with appropriate lattices. We start with the forward dependency function \mathcal{J}_T which “replays” evaluation, turning input availability into output availability, with T guiding the analysis whenever holes in the input selection would mean the analysis would otherwise get stuck. \mathcal{J}_T uses the auxiliary function \mathcal{J}_w for forward-analysing a pattern-match; we explain this first, as it introduces the key idea of a selection as identifying the data available to a downstream computation.

3.2.1 Forward match. Figure 9 defines a family of *forward-match* functions \mathcal{J}_w of type $\text{Sel}_{v,\sigma} \mathcal{A} \rightarrow (\text{Sel}_{\rho,\kappa} \mathcal{A}) \times \mathcal{A}$ for any $w :: v, \sigma \rightsquigarrow \rho, \kappa$. (The definition is presented in a relational style for readability, but should be understood as a total function defined by structural recursion on w , which appears in grey to emphasise the connection to Figure 6.) Forward match replays the match

witnessed by w , turning availability $(v, \sigma) \in \text{Sel}_{v, \sigma} \mathcal{A}$ on the matched value and eliminator into availability $(\rho, \kappa) \in \text{Sel}_{\rho, \kappa} \mathcal{A}$ on the variable bindings and continuation yielded by the match.

$\mathcal{F}_w^\mathcal{A}$ also returns the *meet* of the selection states associated with the part of \mathbf{v} which was matched by σ . We call this the *argument availability*, since it represents the availability of the matched part of a function argument. In the variable case, the empty part of \mathbf{v} was matched and so the argument availability in this context is simply \top , the unit for \sqcap . In the Boolean case, the argument availability is simply the α on true_α or false_α ; the empty list and empty record cases are similar. In the cons case, we return the meet of the α on the cons node itself with the availabilities β and β' computed for v and v' . Non-empty records are similar, but to process the initial part of the record, we supply the neutral selection state \top on the subrecord in order to use the definition recursively. (Note that these subrecords exist only as intermediate artefacts of the interpreter.)

One might hope to be able to dispense with the match witness w and simply define $\mathcal{F}^\mathcal{A}$ by case analysis on v and σ . However, it is then unclear how to proceed in the event that either v or σ is a hole. In particular, it is not clear how to obtain the ρ associated with the original pattern-match in order to produce an environment selection $\rho' \in \text{Sel}_\rho \mathcal{A}$. If $\mathcal{F}^\mathcal{A}$ is defined with respect to a known w , this can be achieved via additional rules $\mathcal{F}^\mathcal{A}\text{-hole-}v$ and $\mathcal{F}^\mathcal{A}\text{-hole-}\sigma$ that define the behaviour at hole to be the same as the behaviour at any \doteq -equivalent value in $\text{Sel}_v \mathcal{A}$ or $\text{Sel}_\sigma \mathcal{A}$.

Operationally, these hole rules can be interpreted as “expanding” the holes in v or σ , in a shape-preserving way, until another rule of the definition applies. Recall the pattern-matching example from § 2.2.4. This pattern-match has the witness $x : xs$, recording that the list $5 : 6 : []$ was matched to the depth of a single cons. Suppose we wish to forward-analyse over the pattern-match using \sqcap to represent the selection on the matched list. The information in the match witness allows us to expand \sqcap to $\sqcap : \perp \sqcap$ and then use the $\mathcal{F}^\mathcal{A}\text{-cons}$ rule to derive the following forward-match:

$$\mathcal{F}^\mathcal{A}\text{-hole-}v \frac{\mathcal{F}^\mathcal{A}\text{-cons} \frac{\mathcal{F}^\mathcal{A}\text{-var} \frac{}{\sqcap, x : xs : e_2 \quad \mathcal{F}_x^\mathcal{A} \quad x : \sqcap, xs : e_2, \top} \quad \mathcal{F}^\mathcal{A}\text{-var} \frac{}{\sqcap, xs : e_2 \quad \mathcal{F}_{xs}^\mathcal{A} \quad xs : \sqcap, e_2, \top}}{\sqcap : \perp \sqcap, \{[] : e_1, (:) : x : xs : e_2\} \quad \mathcal{F}_{x : xs}^\mathcal{A} \quad (x : \sqcap) \cdot (xs : \sqcap), e_2, \perp}}{\sqcap, \{[] : e_1, (:) : x : xs : e_2\} \quad \mathcal{F}_{x : xs}^\mathcal{A} \quad (x : \sqcap) \cdot (xs : \sqcap), e_2, \perp}$$

Lemma 3.3 below implies that an implementation is free to replace any term by a hole-equivalent one of the same shape, with the result of $\mathcal{F}_w^\mathcal{A}$ being unique up to \doteq . This justifies the strategy of expanding holes just enough for a non-hole rule to apply; there will be exactly one such rule, corresponding to the execution path originally taken, and although there may be multiple possible expansions, they will produce hole-equivalent results. This also explains why it is reasonable to think of $\mathcal{F}_w^\mathcal{A}$ not just as a relation, but as a function.

LEMMA 3.3 (MONOTONICITY OF $\mathcal{F}_w^\mathcal{A}$). *Suppose $w :: \mathbf{v}, \sigma \rightsquigarrow \rho, \kappa$, with $v, \sigma \mathcal{F}_w^\mathcal{A} \rho, \kappa, \alpha$ and $v', \sigma' \mathcal{F}_w^\mathcal{A} \rho', \kappa', \alpha'$. If $(v, \sigma) \sqsubseteq (v', \sigma')$ then $(\rho, \kappa, \alpha) \sqsubseteq (\rho', \kappa', \alpha')$.*

The forward-match function $\mathcal{F}_w^\mathcal{A}$ is a key component of the forward evaluation function $\mathcal{F}_T^\mathcal{A}$ defined in § 3.2.2 below. When forward-analysing a function call, the argument is forward-matched using $\mathcal{F}_w^\mathcal{A}$, and the resulting argument availability α used to upper-bound the availability of any partial values constructed by that function, establishing a forward link from resources consumed and resources produced. Since the dynamic context of a function call extends over multiple evaluation steps, $\mathcal{F}_T^\mathcal{A}$ is threaded with an additional input α which tracks the active argument availability; at the outermost level, before there are any active function calls, this has the value \top .

3.2.2 Forward evaluation. Figure 10 defines a family of *forward-evaluation* functions $\mathcal{F}_T^\mathcal{A}$ of type $(\text{Sel}_{\rho, \mathbf{e}} \mathcal{A}) \times \mathcal{A} \rightarrow \text{Sel}_v \mathcal{A}$ for any $T :: \rho, \mathbf{e} \Rightarrow \mathbf{v}$. (Like forward match, forward evaluation is

$v, \sigma \vdash_w \rho, \kappa, \alpha$			v and σ forward-match along w to ρ and κ , with argument availability α		
$\vdash^{\text{hole-}v}$			$\vdash^{\text{hole-}\sigma}$		
$\square \doteq v$			$\square \doteq \sigma$		
$v, \sigma \vdash_w \rho, \kappa, \alpha$			$v, \sigma \vdash_w \rho, \kappa, \alpha$		
$\square, \sigma \vdash_w \rho, \kappa, \alpha$			$v, \square \vdash_w \rho, \kappa, \alpha$		
\vdash^{var}			$v, x: \kappa \vdash_x x: v, \kappa, \top$		
\vdash^{true}			\vdash^{false}		
$\text{true}_\alpha, \{\text{true}: \kappa, \text{false}: \kappa'\} \vdash_{\text{true}} \varepsilon, \kappa, \alpha$			$\text{false}_\alpha, \{\text{true}: \kappa, \text{false}: \kappa'\} \vdash_{\text{false}} \varepsilon, \kappa', \alpha$		
\vdash^{unit}			\vdash^{record}		
$()_\alpha, \{(): \kappa\} \vdash_{()} \varepsilon, \kappa, \alpha$			$(\vec{x}: \vec{v})_\top, \{(\vec{x}): \sigma\} \vdash_{(\vec{x}: \vec{w})} \rho, \sigma', \beta \quad u, \sigma' \vdash_w \rho', \kappa, \beta'$		
\vdash^{nil}			\vdash^{cons}		
$[]_\alpha, \{[]: \kappa, (:): \sigma'\} \vdash_{[]} \varepsilon, \kappa, \alpha$			$v, \sigma \vdash_w \rho, \tau, \beta \quad v', \tau \vdash_{w'} \rho', \kappa', \beta'$		
			$v :_\alpha v', \{[]: \kappa, (:): \sigma\} \vdash_{w: w'} \rho \cdot \rho', \kappa', \alpha \sqcap \beta \sqcap \beta'$		

Fig. 9. Forward match

presented in a relational style, but should be read as a total function defined by structural recursion on T .) Forward evaluation replays T , taking a selection $(\rho, e) \in \text{Sel}_{\rho, e} \mathcal{A}$ identifying the available parts of the environment and program, and an $\alpha \in \mathcal{A}$ representing the argument availability for the dynamically innermost function call, and returning a selection $v \in \text{Sel}_v \mathcal{A}$ identifying the outputs that can be produced using only the available resources. The rules resemble those for the evaluation relation \Rightarrow . The general pattern is that each rule takes the active argument availability α , combines it (using \sqcap) with any availability supplied on the expression form consumed at that step, and uses the result as the availability of any partial values constructed at that step. The argument availability α is passed down unchanged to any subcomputations, except in the case of function application.

Function application. In the application case, the rule must determine a new argument availability for the function body, because the function context is changing. First, we unpacks the β stored in the closure, representing the argument availability which was active when the closure was constructed. Then we determine an additional selection state β' , representing the availability of the matched part of the current argument, by forward-matching v with the eliminator σ from the closure. These are combined using \sqcap to represent the conjoined availability of all arguments that were pattern-matched in order to execute the function body, and the result $\beta \sqcap \beta'$ used to forward-evaluate the function body. The auxiliary function $\mathcal{A}_{\rho, h}: (\text{Sel}_{\rho, h} \mathcal{A}) \times \mathcal{A} \rightarrow \text{Sel}_{\rho'} \mathcal{A}$ for any $\rho, h \rightarrow \rho'$ is given at the bottom of Figure 10 and resembles \rightarrow , but captures the active argument availability into each closure.

Primitive application. Since primitive operations are opaque, their input-output dependencies cannot be derived from their execution, but must be supplied by the primitive operation itself. More specifically, every primitive $\phi \in \text{Int}^i \rightarrow \text{Int}$ is required to provide a forward-dependency function $\phi_{\vec{n}*}: \mathcal{A}^i \rightarrow \mathcal{A}$ for every $\vec{n} \in \text{Int}^i$ which specifies how to turn an input selection $\vec{\alpha} \in \mathcal{A}^i$ for \vec{n} into an output selection α' on $\phi(\vec{n})$. There is one such function per possible input \vec{n} so that the dynamic dependencies for that specific call can depend on the values passed to the operation. For example, in our implementation, the dependency function for multiplication establishes (for non-zero n) that both $n * 0$ and $0 * n$ depend only on 0. However, primitives are free to implement

$\rho, e, \alpha \not\vdash_T v$			ρ and e , with argument availability α , forward-evaluate along T to v		
$\not\vdash$ -hole			$\not\vdash$ -var		
$\square \doteq e \quad \rho, e, \alpha \not\vdash_T v$			$x: v \in \rho$		
$\rho, \square, \alpha \not\vdash_T v$			$\rho, x, \alpha \not\vdash_x v$		
$\not\vdash$ -true			$\not\vdash$ -lambda		
$\rho, \text{true}_{\alpha'}, \alpha \not\vdash_{\text{true}} \text{true}_{\alpha \sqcap \alpha'}$			$\rho, \lambda \sigma, \alpha \not\vdash_{\lambda \sigma'} \text{cl}(\rho, \varepsilon, \alpha, \sigma)$		
$\not\vdash$ -false			$\not\vdash$ -int		
$\rho, \text{false}_{\alpha'}, \alpha \not\vdash_{\text{false}} \text{false}_{\alpha \sqcap \alpha'}$			$\rho, n_{\alpha'}, \alpha \not\vdash_n n_{\alpha \sqcap \alpha'}$		
$\not\vdash$ -record			$\not\vdash$ -project		
$\rho, e_i, \alpha \not\vdash_{T_i} v_i \quad (\forall i \leq \vec{x})$			$\rho, e, \alpha \not\vdash_T \doteq (\vec{x}: \vec{u})_{\beta} \quad y: v' \in \vec{x}: \vec{u}$		
$\rho, (\vec{x}: \vec{e})_{\alpha'}, \alpha \not\vdash_{(\vec{x}: \vec{T})} (\vec{x}: \vec{v})_{\alpha \sqcap \alpha'}$			$\rho, e, y, \alpha \not\vdash_{T_{\vec{x}: \vec{v}} \cdot y} v'$		
$\not\vdash$ -cons			$\not\vdash$ -nil		
$\rho, e, \alpha \not\vdash_T v \quad \rho, e', \alpha \not\vdash_U v'$			$\rho, []_{\alpha'}, \alpha \not\vdash_{[]} []_{\alpha \sqcap \alpha'}$		
$\rho, e:_{\alpha'} e', \alpha \not\vdash_{T:U} v:_{\alpha \sqcap \alpha'} v'$			$\not\vdash$ -apply-prim		
$\not\vdash$ -apply			$\rho, e_i, \alpha \not\vdash_{U_i} \doteq n_i \beta_i \quad (\forall i \leq \vec{n})$		
$\rho, e, \alpha \not\vdash_T \doteq \text{cl}(\rho_1, h, \beta, \sigma) \quad \rho_1, h, \beta \not\vdash \rho_2 \quad \rho, e', \alpha \not\vdash_U v$			$\rho, \phi(\vec{e}), \alpha \not\vdash_{\phi(\vec{U}_n)} \phi(\vec{n})_{\alpha'} \quad \phi_{\vec{n}*}(\vec{\beta}) = \alpha'$		
$v, \sigma \not\vdash_w \rho_3, e'', \beta' \quad \rho_1 \cdot \rho_2 \cdot \rho_3, e'', \beta \sqcap \beta' \not\vdash_{T'} v'$			$\not\vdash$ -let-rec		
$\rho, e e', \alpha \not\vdash_{T \triangleright w: T'} v'$			$\rho, h', \alpha \not\vdash \rho' \quad \rho \cdot \rho', e, \alpha \not\vdash_T v$		
$\rho, \text{let } h' \text{ in } e, \alpha \not\vdash_{\text{let } h \text{ in } T} v$			$\rho, h, \alpha \not\vdash \rho'$		
$\not\vdash$ -rec-defs			h forward-generates to ρ' in ρ and α		
$v_i = \text{cl}(\rho, \vec{x}: \vec{\sigma}, \alpha, \sigma_i) \quad (\forall i \in \vec{x})$			$\not\vdash$ -rec-defs		
$\rho, \vec{x}: \vec{\sigma}, \alpha \not\vdash \vec{x}: \vec{v}$			$\rho, \vec{x}: \vec{\sigma}, \alpha \not\vdash \vec{x}: \vec{v}$		

Fig. 10. Forward evaluation

forward-dependency however they choose, with the caveat that § 3.3.2 will also require ϕ to provide a backward-dependency function for any input \vec{n} , and § 3.4 will require these to be related in a certain way for the consistency of the whole system to be guaranteed.

Other rules. The remaining rules follow the general pattern. Variable lookup disregards α , simply preserving the selection on the value extracted from the environment. The lambda rule captures α in the closure along with the environment; the letrec rule passes α on to $\not\vdash$ so it can be captured by recursive closures as well. Record projection is more interesting, disregarding not only the argument availability α but also the availability β of the record itself. This is because containers are considered to be independent of the values they contain: here, v_i has its own internal availability which is preserved by projection, but there is no implied dependency of the field on the record from which it was projected. Record construction also reflects this principle, preserving the field selections unchanged into the resulting record selection. But since this rule also constructs a partial value — the record itself — it must specify an availability on that output. The availability is set to $\alpha \sqcap \alpha'$, reflecting the dependency of the constructed container on both the constructing expression and the active argument match. The rules for nil, cons, integers and Booleans are similar, since they also construct values.

$\rho, \kappa, \alpha \rightsquigarrow_w v, \sigma$			ρ and κ , with argument demand α , backward-match along w to v and σ		
$\rightsquigarrow_{\text{true}}$			$\rightsquigarrow_{\text{false}}$		$\rightsquigarrow_{\text{var}}$
$\frac{}{\varepsilon, \kappa, \alpha \rightsquigarrow_{\text{true}} \text{true}_\alpha, \{\text{true}: \kappa, \text{false}: \square\}}$			$\frac{}{\varepsilon, \kappa, \alpha \rightsquigarrow_{\text{false}} \text{false}_\alpha, \{\text{true}: \square, \text{false}: \kappa\}}$		
$\rightsquigarrow_{\text{unit}}$			$\rightsquigarrow_{\text{nil}}$		
$\frac{}{\varepsilon, \kappa, \alpha \rightsquigarrow_{()} ()_\alpha, \{(): \kappa\}}$			$\frac{}{\varepsilon, \kappa, \alpha \rightsquigarrow_{[]} []_\alpha, \{[]: \kappa, (:): \square\}}$		
$\rightsquigarrow_{\text{record}}$			$\rightsquigarrow_{\text{cons}}$		
$\frac{\rho', \kappa, \alpha \rightsquigarrow_{w'} u, \sigma \quad \rho, \sigma, \alpha \rightsquigarrow_{(\bar{x}: \bar{w})} (\bar{x}: \bar{v})_\beta, \tau}{\rho \cdot \rho', \kappa, \alpha \rightsquigarrow_{(\bar{x}: \bar{w}: y: w')} (\bar{x}: \bar{v}: y: u)_\alpha, \{(\bar{x}: \bar{v}: y: \tau)\}}$			$\frac{\rho', \kappa, \alpha \rightsquigarrow_{w'} v', \sigma \quad \rho, \sigma, \alpha \rightsquigarrow_w v, \tau}{\rho \cdot \rho', \kappa, \alpha \rightsquigarrow_{w: w'} v: v', \{[]: \square, (:): \tau\}}$		

Fig. 11. Backward match

Hole cases. Environments have no special \square form. However, a hole rule is needed to allow forward evaluation to continue in the event that e is \square ; this is essential because subsequent steps may result in non- \square outputs (for example by extracting non- \square values from ρ). The rule is similar to the hole rules for \rightsquigarrow_w and again can be understood operationally as using the information in T to expand \square sufficiently for another rule to apply, with a result which is unique up to \doteq . In addition, application and record projection must accommodate the case where the selection on the closure or record being eliminated is represented by \square . In these rules $\nearrow_T \doteq$ is used to denote the relational composition of \nearrow_T and \doteq .

LEMMA 3.4 (MONOTONICITY OF \nearrow_T). *Suppose $T :: \rho, e \Rightarrow v$ with $\rho, e, \alpha \nearrow_T v$ and $\rho', e', \alpha' \nearrow_T v'$. If $(\rho, e, \alpha) \sqsubseteq (\rho', e', \alpha')$ then $v \sqsubseteq v'$.*

3.3 Backward data dependency

The backward dependency function \rightsquigarrow_T “rewinds” evaluation, turning output demand into input demand, with T guiding the analysis backward. We start with the auxiliary function \rightsquigarrow_w which is used for backward-analysing a pattern-match.

3.3.1 Backward match. Figure 11 defines a family of *backward-match* functions \rightsquigarrow_w of type $(\text{Sel}_{\rho, \kappa} \mathcal{A}) \times \mathcal{A} \rightarrow \text{Sel}_{v, \sigma} \mathcal{A}$ for any $w :: v, \sigma \rightsquigarrow \rho, \kappa$. Backward-match rewinds the match witnessed by w , turning demand on the environment and continuation into demand on the value and eliminator that were originally matched. The additional input α represents the downstream demand placed on any resources that were constructed in the context of this match; \rightsquigarrow_w transfers this to the matched portion of v , establishing a backwards link from resources produced to resources consumed in a given function context. We call α the *argument demand* since it represents the demand to be pushed backwards onto the matched part of a function argument.

In the variable case, the empty part of v was matched, so α is disregarded. The rule need only ensure that the demand v in the singleton environment $x: v$ is propagated backward. If a Boolean constant was matched, α becomes the demand on that constant, and κ , capturing the demand on the continuation, is used to construct the demand on the original eliminator, with \square used to represent the absence of demand on the non-taken branch. (This use of \square explains why matches w need only retain information about taken branches.) The nil case is similar.

For a cons match $w: w'$, we split the environment into ρ and ρ' , using the fact that there is a unique well-typed decomposition. We then backward-match w and w' recursively to obtain v and

v' , representing the demand on the head and tail of the list. These are combined into the demand on the entire list, using α as the demand on the root cons node. The eliminator selection σ represents the demand on the interim eliminator used to match the tail, and τ the demand on the eliminator used to match the head; these are then combined into a demand on the eliminator used to match the whole list, with \square again used to represent the absence of demand on the nil branch. Records are similar, except that there is only a single branch. The selection state β computed for the initial part of the record is an artefact of processing records recursively, and is disregarded.

LEMMA 3.5 (MONOTONICITY OF \mathcal{Z}_w). *Suppose $w :: \mathbf{v}, \sigma \rightsquigarrow \rho, \kappa$, with $\rho, \kappa, \alpha \mathcal{Z}_w \mathbf{v}, \sigma$ and $\rho', \kappa', \alpha' \mathcal{Z}_w \mathbf{v}', \sigma'$. If $(\rho, \kappa, \alpha) \sqsubseteq (\rho', \kappa', \alpha')$ then $(\mathbf{v}, \sigma) \sqsubseteq (\mathbf{v}', \sigma')$.*

3.3.2 Backward evaluation. Figure 12 defines a family of *backward-evaluation* functions \mathcal{N}_T of type $\text{Sel}_v \mathcal{A} \rightarrow (\text{Sel}_{\rho, e} \mathcal{A}) \times \mathcal{A}$ for any $T :: \rho, e \Rightarrow \mathbf{v}$. Backward evaluation rewinds T , using the output selection $v \in \text{Sel}_v \mathcal{A}$ to determine an input selection $(\rho, e) \in \text{Sel}_{\rho, e} \mathcal{A}$ and an argument demand $\alpha \in \mathcal{A}$ which will eventually be pushed back onto the argument of the dynamically innermost function call. (At the outermost level, where there are no active function calls, the argument demand is discarded.) The rules resemble those of the evaluation relation \Rightarrow with inputs and outputs flipped. The general pattern is that each backward rule takes the join of the demand attached to any partial values constructed at that step, and the argument demand associated with any subcomputations, and passes it upwards as the new argument demand. The output environment is constructed similarly, by joining the demand flowing back through the environment copies used to evaluate subcomputations. Demand is also attached to the source expression when it is the expression form responsible for the construction of a demanded value.

Function application. The application rule is where the argument demand is used and the function context changes, so we start here. The rule essentially runs the forward evaluation rule in reverse, using the trace T' to backward-evaluate the function body. The argument demand β associated with T' is the join of the demand on any resources constructed directly by that function invocation, and is transferred to the matched part of the function argument by the backward-match function \mathcal{Z}_w . The argument demand passed upwards into the enclosing function context is $\alpha \sqcup \alpha'$, representing the resources needed along T and U . The auxiliary function $\mathcal{N}_{\rho, h} : \text{Sel}_{\rho'} \mathcal{A} \rightarrow (\text{Sel}_{\rho, h} \mathcal{A}) \times \mathcal{A}$ for any $\rho, h \rightarrow \rho'$ defined at the bottom of Figure 12 is used to turn ρ_2 , capturing the demand flowing back through any recursive uses of the function and any others with which it was mutually defined, into information that can be merged back into the demand on the closure. The function $\mathcal{N}_{\rho, h}$ is also used in the letrec rule, which otherwise follows the general pattern described above.

Primitive application. Each primitive operation $\phi : \text{Int}^i \rightarrow \text{Int}$ must provide a backward-dependency function $\phi_{\vec{n}}^* : \mathcal{A} \rightarrow \mathcal{A}^i$ for every $\vec{n} \in \text{Int}^i$ which specifies how to turn the output selection α' on $\phi(\vec{n})$ into an input selection $\vec{\alpha} \in \mathcal{A}^i$ on \vec{n} . The rule for primitive application uses this information to pair each argument n_i with its demand α_i and then backwards-evaluate the argument. The argument demand passed upward is the join of those arising from these subcomputations, and is unrelated to the execution of the primitive itself, similar to a function application. Here $\sqcup \vec{\beta}$ means the fold of \sqcup (with unit \perp) over the sequence of selection states $\beta_1 \cdot \dots \cdot \beta'_{|\vec{\alpha}|}$. Environment demands $\vec{\rho} = \rho_1 \cdot \dots \cdot \rho_{|\vec{\alpha}|}$ are joined (pointwise) in a similar fashion.

Other rules. In the variable case, no partial values were constructed during evaluation and there are no subcomputations, so the argument demand is \perp , the unit for \sqcup . The returned environment selection demands v for the variable x and \square for all other variables, using the family of *backwards lookup* functions $-\ni_{\rho} x : -$ of type $\text{Sel}_v \mathcal{A} \rightarrow \text{Sel}_{\rho} \mathcal{A}$ for any $x : \mathbf{v} \in \rho$ also defined in Figure 12. (The output of the function is on the left in the relational notation.) For atomic values such as integers, Booleans and nil, the argument demand is simply the demand α associated with the

$v \Downarrow_T \rho, e, \alpha$ v backward-evaluates along T to ρ and e , with argument demand α

$\Downarrow\text{-hole}$ $\frac{\Box \doteq v \quad v \Downarrow_T \rho, e, \alpha}{\Box \Downarrow_T \rho, e, \alpha}$	$\Downarrow\text{-var}$ $\frac{\rho' \ni_\rho x: v}{v \Downarrow_x \rho', x, \perp}$	$\Downarrow\text{-lambda}$ $\frac{}{\text{cl}(\rho, \varepsilon, \alpha, \sigma) \Downarrow_{\lambda\sigma'} \rho, \lambda\sigma, \alpha}$	$\Downarrow\text{-int}$ $\frac{}{n_\alpha \Downarrow_n \Box_\rho, n_\alpha, \alpha}$
$\Downarrow\text{-true}$ $\frac{}{\text{true}_\alpha \Downarrow_{\text{true}} \Box_\rho, \text{true}_\alpha, \alpha}$	$\Downarrow\text{-false}$ $\frac{}{\text{false}_\alpha \Downarrow_{\text{false}} \Box_\rho, \text{false}_\alpha, \alpha}$	$\Downarrow\text{-record}$ $\frac{v_i \Downarrow_{T_i} \rho_i, e_i, \alpha'_i \quad (\forall i \leq \vec{x})}{(\vec{x}: \vec{v})_\alpha \Downarrow_{(\vec{x}: \vec{T})} \Box \vec{\rho}, (\vec{x}: \vec{e})_\alpha, \alpha \sqcup \Box \vec{\alpha}'}$	
$\Downarrow\text{-project}$ $\frac{\vec{x}: \vec{u} \ni_{\vec{x}: \vec{v}} y: v' \quad (\vec{x}: \vec{u})_\perp \Downarrow_T \rho, e, \alpha}{v' \Downarrow_{\vec{x}: \vec{v}. y} \rho, e, y, \alpha}$		$\Downarrow\text{-nil}$ $\frac{}{[]_\alpha \Downarrow_{[]} \Box_\rho, []_\alpha, \alpha}$	
$\Downarrow\text{-cons}$ $\frac{v \Downarrow_T \rho, e, \alpha \quad v' \Downarrow_U \rho', e', \alpha'}{v: \beta v' \Downarrow_{T: U} \rho \sqcup \rho', e: \beta e', \beta \sqcup \alpha \sqcup \alpha'}$		$\Downarrow\text{-let-rec}$ $\frac{v \Downarrow_T \rho \cdot \rho_1, e, \alpha \quad \rho_1 \Downarrow \rho', h', \alpha'}{v \Downarrow_{\text{let } h \text{ in } T} \rho \sqcup \rho', \text{let } h' \text{ in } e, \alpha \sqcup \alpha'}$	
$\Downarrow\text{-apply-prim}$ $\frac{n_{i\alpha_i} \Downarrow_{U_i} \rho_i, e_i, \beta_i \quad (\forall i \in \vec{n})}{m_{\alpha'} \Downarrow_{\phi(\vec{U}_n)} \Box \vec{\rho}, \phi(\vec{e}), \Box \vec{\beta} \quad \phi_{\vec{n}}^*(\alpha') = \vec{\alpha}}$			
$\Downarrow\text{-apply}$ $\frac{v \Downarrow_{T'} \rho_1 \cdot \rho_2 \cdot \rho_3, e, \beta \quad \rho_3, e, \beta \Downarrow_w v', \sigma \quad v' \Downarrow_U \rho, e_2, \alpha' \quad \rho_2 \Downarrow \rho'_1, h, \beta' \quad \text{cl}(\rho_1 \sqcup \rho'_1, h, \beta \sqcup \beta', \sigma) \Downarrow_T \rho', e_1, \alpha'}{v \Downarrow_{T \triangleright w: T'} \rho \sqcup \rho', e_1 e_2, \alpha \sqcup \alpha'}$			

$\rho' \ni_\rho x: v$ $\Downarrow\text{-head}$ $\frac{}{(\Box_\rho \cdot x: u) \ni_{\rho \cdot x: v} x: u}$	ρ' contains $x: v$ $\Downarrow\text{-tail}$ $\frac{\rho' \ni_\rho x: u \quad x \neq y}{(\rho' \cdot y: \Box) \ni_{\rho \cdot y: v} x: u}$	$\rho \Downarrow \rho', h, \alpha$ ρ backward-generates to ρ', h, α $\Downarrow\text{-rec-defs}$ $\frac{v_i = \text{cl}(\rho_i, h_i, \alpha_i, \sigma_i) \quad (\forall i \in \vec{x})}{\vec{x}: \vec{v} \Downarrow \Box \vec{\rho}, \vec{x}: \vec{\sigma} \sqcup \Box \vec{h}, \Box \vec{\alpha}}$
---	--	---

Fig. 12. Backward evaluation

constructed value, which is also attached to the corresponding expression, and the environment demand has \Box for every variable in the original environment ρ , written \Box_ρ .

For closures, the argument demand is unpacked along with the other components, preserving any internal selections on ρ and σ . Composite values such as records and cons cells follow the general pattern; thus for records, the argument demands α'_i from the subcomputations are joined with the α on the record itself to produce the argument demand passed upward. Record projection never demands the record constructor itself, but simply promotes the field demand into a record demand, using $\ni_{\vec{x}: \vec{v}}$ to demand fields other than y with \Box .

Hole rule. The hole rule, as elsewhere, ensures that the function is defined when v is \Box , and it is easy to show that \Downarrow_T preserves \sqsubseteq , and thus \doteq .

LEMMA 3.6 (MONOTONICITY OF \Downarrow_T). *Suppose $T :: \rho, \mathbf{e} \Rightarrow \mathbf{v}$ with $v \Downarrow_T \rho, e, \alpha$ and $v' \Downarrow_T \rho', e', \alpha'$. If $v \sqsubseteq v'$ then $(\rho, e, \alpha) \sqsubseteq (\rho', e', \alpha')$.*

3.4 Round-tripping properties of \nearrow_T and \searrow_T

We now establish more formally the round-tripping properties, alluded at the beginning of the section, that relate \nearrow_T to \searrow_T . For the analyses to be coherent, we expect $\nearrow_T(\searrow_T(v))$ to produce a value selection $v' \sqsupseteq v$, and $\searrow_T(\nearrow_T(\rho, e))$ to produce an input selection $(\rho', e') \sqsubseteq (\rho, e)$. Pairs of (monotonic) functions $f : X \rightarrow Y$ and $g : Y \rightarrow X$ that are related in this way are called *Galois connections*. Galois connections generalise isomorphisms: f and g are not quite mutual inverses, but are the nearest to an inverse each can get to the other. We will present a visual example of some of these round-tripping properties in § 4.2; here we establish the relevant theorems.

Definition 3.7 (Galois connection). Suppose X and Y are sets equipped with partial orders \leq_X and \leq_Y . Then monotonic functions $f : X \rightarrow Y$ and $g : Y \rightarrow X$ form a *Galois connection* $(f, g) : X \rightarrow Y$ iff $g(f(x)) \geq_X x$ and $f(g(y)) \leq_Y y$.

Galois connections are also adjoint functors between poset categories, with left and right adjoints f and g usually called the *lower* and *upper* adjoints, because f approximates an inverse of g from below, and g an inverse of f from above. Galois connections compose component-wise, so it is useful to think of them as having a type $X \rightarrow Y$, with the direction (by convention) given by the lower adjoint. If $\gamma : X \rightarrow Y$ is a Galois connection, we will write γ^* and γ_* for the lower and upper adjoints respectively; an important property we will return to is that γ^* preserves joins and γ_* preserves meets. We now show that, for any \mathcal{A} , \searrow_T and \nearrow_T form a Galois connection (Theorem 3.11), by first establishing that the relevant auxiliary functions also form Galois connections.

THEOREM 3.8 (GALOIS CONNECTION FOR PATTERN-MATCHING). Suppose $w :: \mathbf{v}, \sigma \rightsquigarrow \rho, \kappa$. Then $(\searrow_w, \nearrow_w) : (\text{Sel}_{\rho, \kappa} \mathcal{A}) \times \mathcal{A} \rightarrow \text{Sel}_{\mathbf{v}, \sigma} \mathcal{A}$ is a Galois connection.

PROOF. Included with [supplementary materials](#). □

LEMMA 3.9 (GALOIS CONNECTION FOR ENVIRONMENT LOOKUP). Suppose $x : \mathbf{v} \in \rho$. Then $(-\exists_\rho x : -, \in_\rho x : -) : \text{Sel}_{\mathbf{v}} \mathcal{A} \rightarrow \text{Sel}_\rho \mathcal{A}$ is a Galois connection.

PROOF. Included with [supplementary materials](#). □

THEOREM 3.10 (GALOIS CONNECTION FOR RECURSIVE BINDINGS). Suppose $\rho, \mathbf{h} \twoheadrightarrow \rho'$. Then $(\searrow_{\rho, \mathbf{h}}, \nearrow_{\rho, \mathbf{h}}) : \text{Sel}_{\rho'} \mathcal{A} \rightarrow (\text{Sel}_{\rho, \mathbf{h}} \mathcal{A}) \times \mathcal{A}$ is a Galois connection.

PROOF. Included with [supplementary materials](#). □

We assume (rather than prove) that the backward and forward dependency functions $\phi_{\vec{n}}^*$ and $\phi_{\vec{n}*}$ provided for every primitive operation $\phi : \text{Int}^i \rightarrow \text{Int}$ and every \vec{n} of length i form a Galois connection of type $\mathcal{A} \rightarrow \mathcal{A}^i$. Under this assumption the following holds.

THEOREM 3.11 (GALOIS CONNECTION FOR EVALUATION). Suppose $T :: \rho, \mathbf{e} \Rightarrow \mathbf{v}$. Then $(\searrow_T, \nearrow_T) : \text{Sel}_{\mathbf{v}} \mathcal{A} \rightarrow (\text{Sel}_{\rho, \mathbf{e}} \mathcal{A}) \times \mathcal{A}$ is a Galois connection.

PROOF. Included with [supplementary materials](#). □

Establishing that (\searrow_T, \nearrow_T) is an adjoint pair might seem rather weak as a correctness property: it merely ensures that the two analyses are related in a sensible way, not that they actually capture any useful information. This is a familiar problem from other approximate analyses like type systems and model checking, where properties like soundness or completeness are essential but do not by themselves guarantee utility. One could certainly define versions of \searrow_T and \nearrow_T that are too coarse grained to be useful, yet still satisfy Theorem 3.11. However Galois connections do at least require that every tightening or tweak to the forward analysis is paired with a corresponding adjustment to the backward analysis, and vice-versa. In § 6 we consider how other ideas from provenance and program slicing might be adapted to provide additional correctness criteria.

4 DE MORGAN DEPENDENCIES FOR BRUSHING AND LINKING

§ 3 addresses the first kind of question we motivated in the introduction (§ 1.1). In particular \bowtie_T can answer questions like: “what data is needed to compute this bar in a bar chart?”, and indeed we were able to use our implementation to generate Figure 1. The second problem we set ourselves was how to link selections between *cognate* outputs, i.e. outputs computed from the same data (§ 1.2). This is called “brushing and linking” [3] in data visualisation, and has been extensively studied as an interaction paradigm, but with little emphasis on techniques for automation. Intuitively, the problem has a bidirectional flavour: one must consider how dependencies flow backward from a selection in one output to a selection v in the common data, and then forward from the selected data v to a corresponding selection in the other output. A natural question then is whether the analysis established in § 3 can supply the information required to support an automated solution.

An immediate problem is that the flavour of the forward dependency required here differs from that provided by the forward analysis \nearrow_T defined in § 3.2. That was able to answer the question: what can we compute given only the data selected in v ? But to identify the related data in another output, we must determine not what the input selection v is sufficient for, but what it is necessary for: those parts of the other output that depend on v . In fact the question can be formulated as a kind of dual: what would we *not* be able to compute if the data selected in v were *unavailable*?

4.1 De Morgan duality

Why \nearrow_T is unsuitable as a forward dependency relation for linking cognate outputs can also be understood in terms of compositionality. Suppose \mathcal{V}_1 and \mathcal{V}_2 are the lattices of selections for two views computed from a shared input source, and \mathcal{D} is the lattice of selections for the shared input. Using the procedure given in § 3, we can obtain two Galois connections $\gamma : \mathcal{V}_1 \rightarrow \mathcal{D}$ and $\delta : \mathcal{V}_2 \rightarrow \mathcal{D}$ as shown in Figure 13a below. (The reader can ignore Figure 13b for the moment.)

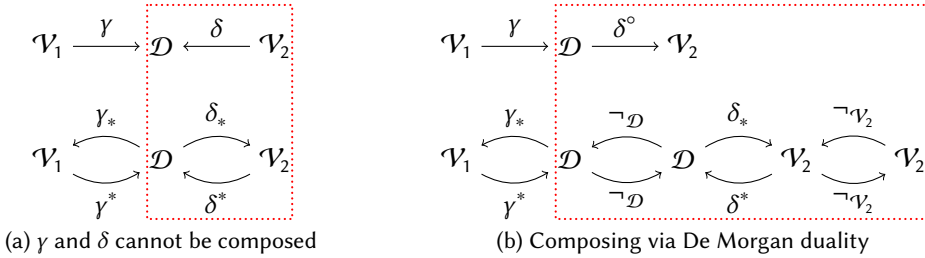


Fig. 13. Dualising $\delta : \mathcal{V}_2 \rightarrow \mathcal{D}$ for composition with $\gamma : \mathcal{V}_1 \rightarrow \mathcal{D}$

Unfortunately, γ and δ are not composable, as their types makes clear. While the upper adjoint $\delta_* : \mathcal{D} \rightarrow \mathcal{V}_2$ has the correct type to compose with the lower adjoint $\gamma^* : \mathcal{V}_1 \rightarrow \mathcal{D}$, the result is not a Galois connection: δ_* preserves meets, whereas γ^* preserves joins. However, it turns out that if selections are closed under complement, we can derive an analysis of what is *necessary* for a given input selection from an analysis of what it is *sufficient* for. The effect is to invert δ , yielding a Galois connection δ° with a type that allows it to compose with γ . Then the composite $\delta^\circ \circ \gamma$ is a Galois connection linking \mathcal{V}_1 to \mathcal{V}_2 via \mathcal{D} , as shown in Figure 13b, offering a general mechanism for brushing and linking, with nice round-tripping properties. We now unpack this in more detail.

First we shift settings from the lattices used in § 3 to Boolean lattices (or Boolean algebras) $\mathcal{A} = \langle \mathcal{A}, \top, \perp, \sqcap, \sqcup, \neg \rangle$, which are lattices equipped with an involution $\neg : \mathcal{A} \rightarrow \mathcal{A}$ called *complement*. Boolean algebras satisfy complementation laws $x \sqcap \neg x = \perp$ and $x \sqcup \neg x = \top$ and De Morgan laws $\neg x \sqcap \neg y = \neg(x \sqcup y)$ and $\neg x \sqcup \neg y = \neg(x \sqcap y)$. If \mathcal{A} is a Boolean algebra, then $\text{Sel}_{\mathcal{V}} \mathcal{A}$

```

1  let zero n = const n;
2  wrap n n_max = ((n - 1) `mod` n_max) + 1;
3  extend n = min (max n 1);
4  nth2 i j xss = nth (j - 1) (nth (i - 1) xss);
5
6  let convolve image kernel method =
7    let ((m, n), (i, j)) = (dims image, dims kernel);
8    (half_i, half_j) = (i `quot` 2, j `quot` 2);
9    area = i * j
10   in < let weightedSum = sum [
11       image!(x, y) * kernel!(i' + 1, j' + 1)
12       | (i', j') ← range (0, 0) (i - 1, j - 1),
13       let x = method (m' + i' - half_i) m,
14       let y = method (n' + j' - half_j) n,
15       x ≥ 1, x ≤ m, y ≥ 1, y ≤ n
16     ] in weightedSum `quot` area
17     | (m', n') in (m, n) >;

```

```

1  let emboss = [[-2, -1, 0],
2               [-1, 1, 1],
3               [ 0, 1, 2]];
4  filter = < nth2 i j emboss
5           | (i, j) in (3, 3) >;
6  image' = [[15, 13, 6, 9, 16],
7            [12, 5, 15, 4, 13],
8            [14, 9, 20, 8, 11],
9            [ 4, 10, 3, 7, 19],
10           [ 3, 11, 15, 2, 9]];
11 image = < nth2 i j image'
12         | (i, j) in (5, 5) >
13 in convolve image filter zero

```

Fig. 14. Matrix convolution example, with methods zero, wrap and extend for dealing with boundaries

is also a Boolean algebra, with the Boolean operations, and in particular $\neg_v : \text{Sel}_v \mathcal{A} \rightarrow \text{Sel}_v \mathcal{A}$, defined pointwise. An additional distinguished value selection \blacksquare serves as the negation of \square . The two-point lattice 2 we used to illustrate § 3 is also a Boolean algebra $\langle \{tt, ff\}, tt, ff, \wedge, \vee, \neg \rangle$ with \neg corresponding to logical negation.

It is an easy consequence of the complementation and De Morgan laws that any meet-preserving operation $g : \mathcal{A} \rightarrow \mathcal{B}$ on Boolean algebras has a join-preserving De Morgan dual $g^\circ : \mathcal{A} \rightarrow \mathcal{B}$ given by $\neg_{\mathcal{B}} \circ g \circ \neg_{\mathcal{A}}$, and any join-preserving operation h has a meet-preserving De Morgan dual h° defined similarly. Moreover if h is the lower adjoint of g , then g° is the lower adjoint of h° . Thus Galois connections on Boolean algebras also admit a (contravariant) notion of De Morgan duality, defined component-wise.

Definition 4.1 (De Morgan dual of a Galois connection). Suppose \mathcal{A} and \mathcal{B} are Boolean algebras and $\gamma : \mathcal{A} \rightarrow \mathcal{B}$ is a Galois connection (γ^*, γ_*) . Define the *De Morgan dual* γ° of γ to be the Galois connection $(\gamma_*^\circ, \gamma^{*\circ}) : \mathcal{B} \rightarrow \mathcal{A}$.

Dualising a Galois connection flips the direction of the arrow by swapping the roles of the upper and lower adjoints. So while $\gamma : \mathcal{A} \rightarrow \mathcal{B}$ and $\delta : \mathcal{C} \rightarrow \mathcal{B}$ are not composable, γ and $\delta^\circ : \mathcal{B} \rightarrow \mathcal{C}$ are, and the composition is achieved by transforming δ_* from something which determines what we can compute with v into something which determines what we cannot compute without v . This offers a principled basis for an automated brushing and linking feature between cognate computations T and U . When the user selects part of the output of T , we can use \Downarrow_T to compute the needed data v , and then use \Uparrow_U° to compute the parts of the output of U that depend on v . This is the approach implemented in Fluid, and we used this to generate Figure 2 in § 1.2.

4.2 Example: matrix convolution

We now illustrate the $(\Uparrow_T^\circ, \Downarrow_T^\circ)$ Galois connection, contrasting it with $(\Downarrow_T, \Uparrow_T)$, using an example which computes the convolution of a 5×5 matrix with a 3×3 kernel. Convolution has an intuitive dependency structure and the values involved have an easy visual presentation, making it useful for conveying the flavour of the four distinct (but connected) dependency relations that arise in the framework. The source code for the example is given in Figure 14, and shows the convolve

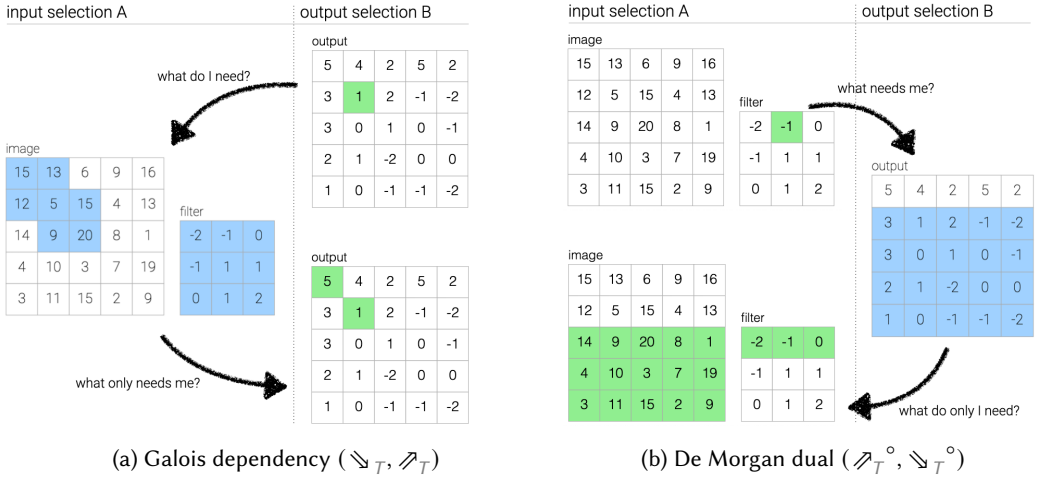


Fig. 15. Upper and lower pairs are dual; left and right pairs are adjoint

function, plus zero, wrap and extend which provide different methods for handling the boundaries of the input matrix. The angle-bracket notation is used to construct matrices, which were omitted from § 2. (The formal treatment is similar to records.)

Fluid was used to generate the diagrams in Figure 15, which show the four dependency relations and two of their four possible round-trips. Figure 15a shows the $(\mathbb{M}_T, \mathbb{M}_T^\circ)$ Galois connection defined in § 3.4. In the upper figure, the user selects (in green) the output cell at position (2, 2) (counting rows downwards from 1). This induces a demand (via the lower adjoint \mathbb{M}_T) on the input matrix image and the kernel filter, revealing (in blue) that the entire kernel was needed to compute the value 1, but only some of the input matrix. In particular the elements at (1, 3) and (3, 1) in image were not needed, because of zeros present in filter. If we then “round-trip” that input selection, computing the corresponding availability on the output using the upper adjoint \mathbb{M}_T° , the green selection grows: it turns out that the data needed to make (2, 2) available are sufficient to make (1, 1) available as well.

Figure 15b shows the De Morgan dual $(\mathbb{M}_T^\circ, \mathbb{M}_T^\circ)$. In the upper part of the figure, the user selects (green) kernel cell (1, 2) to see the output cells that depend on it. This is computed using the De Morgan dual \mathbb{M}_T° . First we negate the input selection, marking (1, 2) as unavailable, and all other inputs as available. Then we forward-analyse with \mathbb{M}_T to determine that with this data selection, we can only compute the top row of the output. (If it seems odd that we can compute even the top row, notice that the example uses the method zero for dealing with boundaries; wrap or extend would give a different behaviour.) Then we negate that top row selection to produce the (blue) output selection shown in the figure. These are exactly the output cells which depend on kernel cell (1, 2) in the sense that they cannot be computed if that input is unavailable.

We can then round-trip this output selection using the De Morgan dual \mathbb{M}_T° . We first negate the blue output selection (selecting the top row of the output again), and then use \mathbb{M}_T to determine the needed inputs, which turn out to be the top two rows of image, and the top row of filter. Negating again produces the green output selection shown in the lower figure. Thus the backwards De Morgan dual computes the inputs that would *not* be needed if the selected outputs were not needed: more economically, the inputs that are *only* needed for the selected output. Here the round-trip reveals that if kernel cell (1, 2) is unavailable, then the entire top row of the kernel might as well have been unavailable too, and similarly for the bottom 3 rows of the input.

```

1  let (x, y) = (2, 3);
2    sum = x + y in
3  if sum == 0
4    then (0, 0)
5    else (x / sum, y / sum)
6  ⇒ (0.4, 0.6)

```

(a) Original program

```

1  let (x, y) = (2, 3);
2    sum = x + y in
3  if sum == 0
4    then □
5    else (x / sum, □)
6  ⇒ (0.4, □)

```

(b) Backward slice for (0.4, □)

```

1  let (x, y) = (2, 3);
2    sum = x + y in
3  if sum == 0
4    then □
5    else (□, □)
6  ⇒ (□, □)

```

(c) Backward slice for spine (□, □)

```

1  let (x, y) = (2, 3);
2    sum = x + y in
3  if sum == 0
4    then (0, 0)
5    else (x / sum, y / sum)
6  ⇒ (0.4, 0.6)

```

(d) Differential backward slice for (0.4, □)

Fig. 16. Differential Galois slicing selects input (blue) needed *only* for selected output (green)

4.3 Relationship to Galois slicing

The De Morgan dual puts us in a better position to consider the relationship between the present system and earlier work on *Galois slicing*, a program slicing technique that has been explored for pure functional programs [23, 22], functional programs with effects [28], and π -calculus [24]. We consider other related work in § 6.1.

Galois slicing operates on lattices of *slices*, which are programs (or values) where parts deemed irrelevant are replaced by a hole \square . (If we think of the notion of selection defined in § 3.1.1 as picking out a subset of the paths in a term, then slices resembles selections which are prefix-closed, meaning that if a given path in a term is selected, then so are all of its prefixes.) For a fixed computation, a meet-preserving *forward-slicing* function is defined which takes input slices to output slices, discarding parts which cannot be computed because the needed input is not present, plus a join-preserving *backward-slicing* function taking output slices to input slices, retaining the parts needed for the output slice. For example Figure 16a shows a computation with output (0.4, 0.6), and Figure 16b gives the backward slice for output slice (0.4, \square). Forward and backward slicing, for a given computation, form a Galois connection, giving the analyses the nice round-tripping properties we motivated in § 3.4.

Unfortunately, the notion of slice does not lend itself to computing dependencies where the needed input or output is a proper part of a value, such as a component of a tuple. *Differential slicing* [23] improves on this by using Galois slicing to compute a pair of input slices (e, e') for a pair of output slices (v, v') where $v \sqsubseteq v'$. By monotonicity, $e \sqsubseteq e'$. This can be used to compute a (differential) slice for an arbitrary subtree, by setting v to be the “spine” of the original output up to the location of the subtree, and v' to be v with the subtree of interest plugged back in. Here we could focus on the value 0.4 in the output by computing the backward slice for (\square, \square) (Figure 16c) and then comparing it with the backward slice for (0.4, \square), generating a differential slice where the parts that are different are highlighted (Figure 16d). But although it supports a notion of selection which is closer to what we need, the differential slice highlights only the program parts that are needed *exclusively* by the selected output, and as such underapproximates the dependency information needed for data linking. (In fact differential slicing is similar to the De Morgan dual \bowtie_T° .) Because in this example 2 and 3 are needed to compute the spine as well (in order to decide which conditional branch to execute), they are excluded from the differential slice, whereas our backward analysis \bowtie_T is able to directly determine that both 2 and 3 are needed to compute 0.4.

Identifier		Recursive function	
$x, y ::= \dots$		$g ::= x: \vec{c}$	
\oplus	operator name	Clause	
Surface term		$c ::= \vec{p} = s$	
$s, t ::= \dots$		Pattern	
(\oplus)	first-class operator	$p ::= x$	variable
$s \oplus s'$	binary application	$(\vec{x}: \vec{p})$	record
$\text{let } \vec{g} \text{ in } s$	recursive functions	$[]$	nil
$\text{if } s \text{ then } s \text{ else } s$	if	$p : p$	cons
$\text{match } s \text{ as } \vec{p} = \vec{s}$	match	$[p \ o$	non-empty list
$\text{let } p = s \text{ in } s$	structured let	List rest pattern	
$[_\alpha \ s \ r]$	non-empty list	$] \quad \text{end}$	
$[s \dots s]$	list enum	$, p \ o \quad \text{cons}$	
$[s \mid \vec{q}]_\alpha$	list comprehension	Qualifier	
List rest term		$q ::= s$	guard
$r ::= [_\alpha$	end	$\text{let } p = s$	declaration
$,_\alpha \ s \ r$	cons	$p \leftarrow s$	generator

Fig. 17. Syntax for surface language, with selection states

5 GALOIS CONNECTIONS FOR DESUGARING

Elaborating a richer surface language into a simpler core is a common pattern with well known benefits. It can, however, make it harder to express certain information to the programmer in terms of the surface language. We face this problem with the analysis in § 3, which links outputs not only to inputs, but also to expressions responsible for introducing data. We could use this information in an IDE to link structured outputs to relevant code fragments, but only if we are able to map term selections back to the surface program. We now sketch a bidirectional desugaring procedure which addresses this, and which composes with the Galois dependency analysis defined in § 3.

5.1 Surface language syntax

The surface language, Fluid, extends the core syntax with list notation $[s, \dots, s']$, Haskell 98-style list comprehensions [17], list enumerations, first-class primitives and piecewise function definitions and pattern-matching, as shown in Figure 17. Typing rules are included with the supplementary materials. We attach selection states α to surface terms s, t that desugar to core terms with selections, and let s, t range over “raw” surface terms, which are isomorphic to the term selections where the type of selection states is the unit lattice 1.

Figure 18 shows how the end-to-end mapping would appear to a user. (For illustrative purposes the library function `map` and some raw data are included in the same source file.) On the left, the user selects a cons cell (green) in the output; by backwards evaluating and then backwards desugaring, we are able to highlight the list comprehension, the cons in the second clause of `map`, and both occurrences of the constant “Hydro”. These last two are highlighted because the selected cons cell was constructed by eliminating a Boolean that was in turn constructed by the primitive `==` operator, which consumed the two strings. The user might then conjecture that the two occurrences of “Geo” were somehow responsible for the inclusion of the third cons cell in the output; they can confirm this by making the green selection on the right. (Highlighting `==` too would clearly be helpful here; we discuss this possibility in § 6.1.) The grey selection is included to contrast the cons highlighting with the data demanded by the list elements themselves, which is quite different.


```

1  let map f [] = [];
2  map f (x : xs) = f x : map f xs;
3  let data = [
4    { energyType: "Bio", output: 6.2 },
5    { energyType: "Hydro", output: 260 },
6    { energyType: "Solar", output: 19.9 },
7    { energyType: "Wind", output: 91 },
8    { energyType: "Geo", output: 14.4 }
9  ];
10 let xs = [ row.output
11   | type ← ["Hydro", "Solar", "Geo"],
12   row ← data, row.energyType == type
13 ] in
14 map (fun x → floor (x / sum xs * 100)) xs
15 ⇒ (88 : (6 : (4 : [])))

```

```

1  let map f [] = [];
2  map f (x : xs) = f x : map f xs;
3  let data = [
4    { energyType: "Bio", output: 6.2 },
5    { energyType: "Hydro", output: 260 },
6    { energyType: "Solar", output: 19.9 },
7    { energyType: "Wind", output: 91 },
8    { energyType: "Geo", output: 14.4 }
9  ];
10 let xs = [ row.output
11   | type ← ["Hydro", "Solar", "Geo"],
12   row ← data, row.energyType == type
13 ] in
14 map (fun x → floor (x / sum xs * 100)) xs
15 ⇒ (88 : (6 : (4 : [])))

```

Fig. 18. Source selections (blue) resulting from selecting different list cells (green)

5.2 Forward desugaring

To define the forward evaluation function \mathcal{R}_T in § 3.2, we performed a regular evaluation using \Rightarrow to obtain a trace T , and then defined \mathcal{R}_T by recursion over T , with T guiding the analysis in the presence of \square . There are no holes in the surface language, so we can take a simpler approach, defining a single *forward desugaring* relation \mathcal{R} , and then showing that for every raw surface term $\mathbf{t} \mathcal{R} \mathbf{e}$, there is a monotonic function $\mathcal{R}_t: \text{Sel}_t \mathcal{A} \rightarrow \text{Sel}_e \mathcal{A}$, which is simply \mathcal{R} domain-restricted to $\text{Sel}_t \mathcal{A}$. The full definition of \mathcal{R} is included with the supplementary materials; Figure 19 gives a representative selection of the rules.

The definition follows a similar pattern to \mathcal{R}_T . At each step, we take the meet of the availability on any parts of s being consumed at that step, and use that as the availability of any parts of e being generated at that step. Thus the rules for list notation simply transfer the selection state α on the opening and closing brackets $[_\alpha$ and $]_\alpha$ to the corresponding cons and nil of the resulting list, and those on intervening delimiters $,_\alpha$ to the corresponding cons. List comprehensions $[s \mid \vec{q}]_\alpha$ have a rule for each kind of qualifier q at the head of \vec{q} , plus a rule for when \vec{q} is ε . The general pattern is to push the α on the comprehension itself through recursively, so it ends up on all core terms generated during its elaboration: in particular the *false* branch when q is a guard, and the singleton list when \vec{q} is empty. Auxiliary relations \mathcal{R} and \mathcal{R}_p (included with the supplementary materials) transfer availability on guards and generators onto the eliminators they elaborate into.

5.3 Backward desugaring

The backwards analysis is then defined as a family of *backward desugaring* functions $\mathcal{R}_t: \text{Sel}_e \mathcal{A} \rightarrow \text{Sel}_t \mathcal{A}$ for any $\mathbf{t} \mathcal{R} \mathbf{e}$, with the raw surface term \mathbf{t} guiding the analysis backwards. (The role of \mathbf{t} in disambiguating the backwards rules should be clear if you consider that e typically matches multiple rules but only one for a given \mathbf{t} .) Figure 19 gives some representative rules; the full definition is included with the supplementary materials. To reverse a desugaring step, we take the join of the demand on any parts of e which were constructed at this step, and use that as the demand on the parts of s which were consumed at this step, turning demand on the core term into (minimal) demand on the surface term. Thus the effect of the list comprehension rules and auxiliary judgements is to set the demand on the comprehension itself to be the join of the demand of all the syntax generated during the elaboration of the comprehension, using auxiliary judgements \searrow_p and \searrow_p to transfer demand from eliminators back onto the guards and generators.

$s \not\approx e$		s forward-desugars to e	
$\not\approx$ -nil	$\not\approx$ -cons	$\not\approx$ -non-empty-list	$\not\approx$ -list-comp-done
$\frac{}{[]_\alpha \not\approx []_\alpha}$	$\frac{s \not\approx e \quad s' \not\approx e'}{s :_\alpha s' \not\approx e :_\alpha e'}$	$\frac{s \not\approx e \quad r \not\approx e'}{[_\alpha s r \not\approx e :_\alpha e']}$	$\frac{s \not\approx e}{[s \mid \varepsilon]_\alpha \not\approx e :_\alpha []_\alpha}$
$\not\approx$ -list-comp-gen		$\not\approx$ -list-comp-decl	
$\frac{[s \mid \vec{q}]_\alpha \not\approx e \quad p, e \succ \sigma \quad \sigma, \alpha \nearrow_p \sigma' \quad s' \not\approx e'}{[s \mid p \leftarrow s' \cdot \vec{q}]_\alpha \not\approx \text{concatMap } \lambda \sigma' e'}$		$\frac{[s \mid \vec{q}]_\alpha \not\approx e \quad p, e \succ \sigma \quad s' \not\approx e}{[s \mid \text{let } p = s' \cdot \vec{q}]_\alpha \not\approx \lambda \sigma e}$	
$\not\approx$ -list-comp-guard			
$\frac{[s \mid \vec{q}]_\alpha \not\approx e \quad s' \not\approx e'}{[s \mid s' \cdot \vec{q}]_\alpha \not\approx \lambda\{\text{true}: e, \text{false}: []_\alpha\} e'}$			
$e \Downarrow_t s$		e backward-desugars along t to s	
\Downarrow -nil	\Downarrow -cons	\Downarrow -non-empty-list	\Downarrow -list-comp-done
$\frac{}{[]_\alpha \Downarrow_t []_\alpha}$	$\frac{e \Downarrow_t s \quad e' \Downarrow_{t'} s'}{e :_\alpha e' \Downarrow_{t : t'} s :_\alpha s'}$	$\frac{e \Downarrow_t s \quad e' \Downarrow_{t'} r'}{e :_\alpha e' \Downarrow_{[t \mid r]} [_\alpha s r']}$	$\frac{e \Downarrow_t s}{e :_\alpha []_\alpha \Downarrow_{[t \mid \varepsilon]} [s \mid \varepsilon]_{\alpha \sqcup \alpha'}}$
\Downarrow -list-comp-gen		\Downarrow -list-comp-decl	
$\frac{e \Downarrow_t s \quad \sigma \searrow_p \sigma', \beta \quad \sigma' \searrow_p e' \quad e' \Downarrow_{[t' \mid \vec{q}]} [s' \mid \vec{q}']_{\beta'}}{\text{concatMap } \lambda \sigma e \Downarrow_{[t' \mid p \leftarrow t \cdot \vec{q}]} [s' \mid p \leftarrow s \cdot \vec{q}']_{\beta \sqcup \beta'}}$		$\frac{\sigma \searrow_p e' \quad e' \Downarrow_{[t' \mid \vec{q}]} [s' \mid \vec{q}']_{\beta} \quad e \Downarrow_t s}{\lambda \sigma e \Downarrow_{[t' \mid \text{let } p = t \cdot \vec{q}]} [s' \mid \text{let } p = s \cdot \vec{q}']_{\beta}}$	
\Downarrow -list-comp-guard			
$\frac{e' \Downarrow_{t'} s' \quad e \Downarrow_{[t \mid \vec{q}]} [s \mid \vec{q}']_{\beta}}{\lambda\{\text{true}: e, \text{false}: []_\alpha\} e' \Downarrow_{[t \mid t' \cdot \vec{q}]} [s \mid s' \cdot \vec{q}']_{\alpha \sqcup \beta}}$			
$r \not\approx e$	r forward-desugars to e	$e \Downarrow_r r'$	e backward-desugars along r to r'
$\not\approx$ -list-rest-end	$\not\approx$ -list-rest-cons	\Downarrow -list-rest-end	\Downarrow -list-rest-cons
$\frac{}{[_\alpha] \not\approx []_\alpha}$	$\frac{s \not\approx e \quad r \not\approx e'}{(\cdot, \alpha s r) \not\approx e :_\alpha e'}$	$\frac{}{[]_\alpha \Downarrow_r []_\alpha}$	$\frac{e \Downarrow_t s \quad e' \Downarrow_{t'} r'}{e :_\alpha e' \Downarrow_{(\cdot, t r)} (\cdot, \alpha s r')}$

Fig. 19. Forwards and backwards desugaring (selected rules only)

5.4 Round-tripping properties and compositionality

It is easy to verify that $\not\approx_t$ and \Downarrow_t are monotonic. Moreover they form an adjoint pair.

THEOREM 5.1 (GALOIS CONNECTION FOR DESUGARING). *Suppose $t \not\approx e$. Then $(\Downarrow_t, \not\approx_t) : \text{Sel}_e \mathcal{A} \rightarrow \text{Sel}_t \mathcal{A}$ is a Galois connection.*

PROOF. Included with [supplementary materials](#). □

The $(\Downarrow_t, \not\approx_t)$ Galois connection readily composes with $(\Downarrow_t, \not\approx_t)$ to produce surface-language selections like the ones shown in Figure 18. This is useful, although somewhat monolithic. In future work we will investigate techniques for backwards desugaring at arbitrary steps in the computation, perhaps by interleaving desugaring with execution in the style of Pombrio and Krishnamurthi [25], as well as presenting selections on intermediate values (such as lists) in the surface language, even though they were not obtained via desugaring.

6 CONCLUSION

Our research² was motivated by the goal of making computational outputs which are automatically able to reveal how they relate to data in a fine-grained way. A casual reader who wants to understand or fact-check a chart, or a scientist evaluating another’s work, should be able to do so by interacting directly with an output. Recent work by Walny et al. suggests that developers would also benefit from such a feature while implementing visualisations, for example to check whether a quantity is represented by diameter or area in a bubble chart [32].

Galois connections provide an appealing setting for this problem because of their elegant round-tripping properties. However, existing dynamic analysis techniques based on Galois connections do not lend themselves to richly structured outputs like visualisations and matrices. We presented an approach that allows focusing on arbitrary substructures, which also means data selections can be inverted. This enables linking not just of outputs to data, but of outputs to other outputs, providing a mathematical basis for a widely used (but so far ad hoc) feature in data visualisation. We implemented our approach in *Fluid*, a realistic high-level functional programming language.

6.1 Other related work and future directions

We close by considering some limitations and opportunities in the context of other related work. Galois slicing [23, 28, 24] was considered in § 4.3.

Executable slicing. Executable slices [14] are programs with some parts removed, but which are still executable. Our approach computes data selections, not executable slices, but such a notion has obvious relevance in data science: “explaining” part of a result should (arguably) entail being able to recompute it. *Expression provenance* [1] explains how primitive values are computed using only primitive operations; however, this still omits crucial information, and does not obviously generalise to structured outputs. Work on executable slicing in term rewriting [11] could perhaps be adapted to structured data and combined with dependency tracking for higher-order data (§ 3.1).

Dynamic program analysis. Dynamic analysis techniques like dataflow analysis [8, 27] and taint tracking [27] tend to focus on variables, rather than parts of structured values, and lack round-tripping properties; Galois dependencies have a clear advantage here. A limitation of dynamic techniques which is shared by our approach is that they can usually only reveal *that* certain dependencies arise, not *why*, which requires analysing path conditions [15]. In a data science setting this would clearly be valuable too, and it would be interesting to see if the benefits of the Galois framework can be extended to techniques for computing dynamic path conditions.

Brushing and linking. Brushing and linking has been extensively studied in the data visualisation community [19, 3], but although Roberts and Wright [29] argued it should be ubiquitous, no automated method of implementation has been proposed to date. Geospatial applications like GeoDa [2] hard-code view coordination features into specific views, and libraries like d3.js and Plotly support ad hoc linking mechanisms, with varying degrees of programmer effort required. No existing approach provides automation or round-tripping guarantees, or is able to provide data selections explaining why visual selections are linked.

Data provenance in data visualisation. A recent vision paper by Psallidas and Wu [26] is the only work we are aware of that proposes that brushing and linking, and related view coordination features like cross-filtering, can be understood in terms of data provenance. In a relational (query processing) setting, where the relevant notion of provenance is lineage, they propose backward-analysing to data, and then forward-analysing to another view, although again without the round-tripping features of Galois connections. Moreover theirs is primarily a concept paper, proposing a research programme, rather than solving a specific problem.

²Perera and Petricek were supported by EPSRC grants EP/T001569/1 & EP/N510129/1. Wang was supported by EP/T008911/1.

REFERENCES

- [1] Umut A. Acar, Amal Ahmed, James Cheney, and Roly Perera. 2012. A Core Calculus for Provenance. In *Proceedings of the First International Conference on Principles of Security and Trust* (Tallinn, Estonia) (POST '12). Springer-Verlag, Berlin, Heidelberg, 410–429. https://doi.org/10.1007/978-3-642-28641-4_22
- [2] Luc Anselin, Ibnu Syabri, and Youngihn Kho. 2006. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38, 1 (2006), 5–22. <https://doi.org/10.1111/j.0016-7363.2005.00671.x>
- [3] Richard A. Becker and William S. Cleveland. 1987. Brushing Scatterplots. *Technometrics* 29, 2 (May 1987), 127–142. <https://doi.org/10.1080/00401706.1987.10488204>
- [4] Richard Bird and Lambert Meertens. 1998. Nested datatypes. In *Mathematics of Program Construction*, Johan Jeuring (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 52–67.
- [5] Gérard Boudol and Ilaria Castellani. 1989. Permutation of transitions: An event structure semantics for CCS and SCCS. In *Linear Time, Branching Time and Partial Order in Logics and Models for Concurrency*, J.W. Bakker, W.-P. Roever, and G. Rozenberg (Eds.). Lecture Notes in Computer Science, Vol. 354. Springer, 411–427. <https://doi.org/10.1007/BFb0013028>
- [6] Nadië Bremer and Marlieke Ranzijn. 2015. Urbanization in East Asia between 2000 and 2010. <http://nbremer.github.io/urbanization/>.
- [7] Peter Buneman, Sanjeev Khanna, and Wang-Chiew Tan. 2001. Why and Where: A Characterization of Data Provenance. In *Proceedings of the 8th International Conference on Database Theory (ICDT '01)*. Springer-Verlag, London, UK, 316–330.
- [8] TY Chen and PC Poole. 1988. Dynamic dataflow analysis. *Information and Software Technology* 30, 8 (1988), 497–505. [https://doi.org/10.1016/0950-5849\(88\)90146-2](https://doi.org/10.1016/0950-5849(88)90146-2)
- [9] Richard H. Connelly and F. Lockwood Morris. 1995. A generalization of the trie data structure. *Mathematical Structures in Computer Science* 5, 3 (1995), 381–418. <https://doi.org/10.1017/S0960129500000803>
- [10] A. De Lucia, A.R. Fasolino, and M. Munro. 1996. Understanding function behaviors through program slicing. In *WPC '96. 4th Workshop on Program Comprehension*. 9–18. <https://doi.org/10.1109/WPC.1996.501116>
- [11] John Field and Frank Tip. 1998. Dynamic Dependence in Term Rewriting Systems and its Application to Program Slicing. *Information and Software Technology* 40, 11–12 (November/December 1998), 609–636.
- [12] Jeremy Gibbons. 2017. APlicative Programming with Naperian Functors. In *European Symposium on Programming (Lecture Notes in Computer Science, Vol. 10201)*, Hongseok Yang (Ed.). 568–583. https://doi.org/10.1007/978-3-662-54434-1_21
- [13] Sebastian Graf, Simon Peyton Jones, and Ryan G Scott. 2020. Lower your guards: a compositional pattern-match coverage checker. *Proceedings of the ACM on Programming Languages* 4, ICFP (2020), 1–30.
- [14] Robert J. Hall. 1995. Automatic extraction of executable program subsets by simultaneous dynamic program slicing. *Automated Software Engineering* 2 (1995), 33–53. <https://doi.org/10.1007/BF00873408>
- [15] Christian Hammer, Martin Grimme, and Jens Krinke. 2006. Dynamic path conditions in dependence graphs. *Proceedings of the ACM SIGPLAN Symposium on Partial Evaluation and Semantics-Based Program Manipulation*, 58–67. <https://doi.org/10.1145/1111542.1111552>
- [16] Ralf Hinze. 2000. Generalizing generalized tries. *Journal of Functional Programming* 10, 4 (2000), 327–351. <https://doi.org/10.1017/S0956796800003713>
- [17] Simon L. Peyton Jones. 2003. Haskell 98: Introduction. *Journal of Functional Programming* 13, 1 (2003), 0–6.
- [18] Gary A. Kildall. 1973. A Unified Approach to Global Program Optimization. In *Proceedings of the 1st Annual ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages* (Boston, Massachusetts) (POPL '73). Association for Computing Machinery, New York, NY, USA, 194–206. <https://doi.org/10.1145/512927.512945>
- [19] John Alan McDonald. 1982. *Interactive graphics for data analysis*. Ph.D. Dissertation.
- [20] Greg Miller. 2006. A Scientist's Nightmare: Software Problem Leads to Five Retractions. *Science* 314, 5807 (2006), 1856–1857. <https://doi.org/10.1126/science.314.5807.1856>
- [21] James Newsome and Dawn Song. 2005. Dynamic taint analysis for automatic detection, analysis, and signature generation of exploits on commodity software. In *Network and Distributed Systems Security Symposium*.
- [22] Roland Perera. 2013. *Interactive Functional Programming*. Ph.D. Dissertation. University of Birmingham, Birmingham, UK. <http://etheses.bham.ac.uk/4209/>.
- [23] Roly Perera, Umut A. Acar, James Cheney, and Paul Blain Levy. 2012. Functional Programs That Explain Their Work. In *Proceedings of the 17th ACM SIGPLAN International Conference on Functional Programming* (Copenhagen, Denmark) (ICFP '12). ACM, New York, NY, USA, 365–376. <https://doi.org/10.1145/2364527.2364579>
- [24] Roly Perera, Deepak Garg, and James Cheney. 2016. Causally Consistent Dynamic Slicing. In *Concurrency Theory, 27th International Conference, CONCUR '16 (Leibniz International Proceedings in Informatics (LIPIcs))*, Josée Desharnais and Radha Jagadeesan (Eds.). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany. <https://doi.org/10.4230/LIPIcs.CONCUR.2016.18>

- [25] Justin Pombrio and Shiram Krishnamurthi. 2014. Resugaring: Lifting Evaluation Sequences through Syntactic Sugar. *SIGPLAN Notices* 49, 6 (Jun 2014), 361–371. <https://doi.org/10.1145/2666356.2594319>
- [26] Fotis Psallidas and Eugene Wu. 2018. Provenance for Interactive Visualizations. In *Workshop on Human-In-the-Loop Data Analytics (HILDA 2018)*. ACM.
- [27] Thomas Reps, Susan Horwitz, and Mooly Sagiv. 1995. Precise Interprocedural Dataflow Analysis via Graph Reachability. In *Proceedings of the 22nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages* (San Francisco, California, USA) (*POPL '95*). Association for Computing Machinery, New York, NY, USA, 49–61. <https://doi.org/10.1145/199448.199462>
- [28] Wilmer Ricciotti, Jan Stolarek, Roly Perera, and James Cheney. 2017. Imperative Functional Programs That Explain Their Work. *Proceedings of the ACM on Programming Languages* 1, ICFP, Article 14 (2017), 28 pages. <https://doi.org/10.1145/3110258>
- [29] J. C. Roberts and M. A. E. Wright. 2006. Towards Ubiquitous Brushing for Information Visualization. In *Tenth International Conference on Information Visualisation (IV'06)*. 151–156. <https://doi.org/10.1109/IV.2006.113>
- [30] A. Sabelfeld and A. C. Myers. 2003. Language-Based Information-Flow Security. *IEEE Journal on Selected Areas in Communications* 21, 1 (Jan 2003), 5–19. <https://doi.org/10.1109/JSAC.2002.806121>
- [31] Jacob VanderPlas, Brian E. Granger, Jeffrey Heer, Dominik Moritz, Kanit Wongsuphasawat, Arvind Satyanarayan, Eitan Lees, Ilia Timofeev, Ben Welsh, and Scott Sievert. 2018. Altair: Interactive Statistical Visualizations for Python. *The Journal of Open Source Software* 3, 32 (2018). <https://doi.org/10.21105/joss.01057>
- [32] Jagoda Walny, Christian Frisson, Mieka West, Doris Kosminsky, Søren Knudsen, Sheelagh Carpendale, and Wesley Willett. 2019. Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff. *IEEE Transactions on Visualization and Computer Graphics* PP (08 2019), 1–1. <https://doi.org/10.1109/TVCG.2019.2934538>
- [33] Mark Weiser. 1981. Program slicing. In *Proceedings of the 5th International Conference on Software Engineering* (San Diego, California, USA) (*ICSE '81*). IEEE Press, Piscataway, NJ, USA, 439–449. <https://doi.org/10.5555/800078.802557>