

# Trace contracts

CAMERON MOY and MATTHIAS FELLEISEN

Northeastern University, Boston, MA, USA

(e-mails: [camoy@ccs.neu.edu](mailto:camoy@ccs.neu.edu), [matthias@ccs.neu.edu](mailto:matthias@ccs.neu.edu))

---

## Abstract

Behavioral software contracts allow programmers to strengthen the obligations and promises that they express with conventional types. They lack expressive power, though, when it comes to invariants that hold across several function calls. Trace contracts narrow this expressiveness gap. A trace contract is a predicate over the sequence of values that flow through function calls and returns. This paper presents a principled design, an implementation, and an evaluation of trace contracts.

---

## 1 Multi-call constraints for APIs

Conventional type systems lack the power to express all the obligations and promises that an API imposes on, or promises to, client modules. Some language designers cover this expressiveness gap with contracts (Meyer, 1988, 1992), dubbed *behavioral contracts* in the literature. Simply put, a contract is a Boolean-valued assertion that governs some aspect of an API. Say a programmer wishes to narrow the set of acceptable inputs to a function from integers to primes. A type combined with a contract, say  $\{p:\text{Int} \mid \text{isPrime } p\}$ , expresses this concisely. A proof assistant might discharge this assertion at compile time or a run-time check might monitor it during execution.

While contracts can easily express logical constraints on function signatures, other constraints pose challenges. Temporal properties in particular are difficult to express. Due to this expressiveness gap, APIs come with sequence diagrams, protocol descriptions, and other informal specifications. The Unix I/O API is a standard example: “open a file before reading from it.” A framework for specifying static-analysis passes may state that the given transfer functions must be monotone. A GUI framework may allow the registration of callback objects and promise to call them back in the order of registration.

This paper presents *trace contracts*, an extension of contract systems that permits the functional specification of constraints across multiple function and method calls. A *trace* reifies the sequence of values that flow through certain interception points of a contract system (Dimoulas et al., 2016), say, function calls. A *trace contract* inspects this reified trace with a predicate that decides whether a property holds.

Concretely, this paper reports two contributions. The first is a principled blueprint of trace contracts (section 4), including the design of a compiler to ordinary contracts with a correctness theorem (section 6). Working through the blueprint points to the central challenge of extending existing systems with trace contracts: on the one hand, specifications

47 should remain functional, while on the other hand, collecting a trace of values necessarily  
48 involves mutable state. Managing this state while maintaining ordinary contract composi-  
49 tion is key. Our insight is to separate value-interception time from the point when a value  
50 crosses from one component to another.

51 The second contribution is a practical and efficient implementation of the blueprint in  
52 Racket, which could be ported to any other language that satisfies some basic requirements  
53 (section 7). The implementation supports both predicates over full traces (as streams) as  
54 well as the use of efficient, bespoke data structures. For example, the creator of a static-  
55 analysis pass could state the monotonicity obligation as a predicate either across a full trace  
56 of all input-output pairs or a special-purpose, tree-based data structure. A performance  
57 evaluation shows that the fixed-cost overhead of trace contracts is between 1% and 17%  
58 on average (section 8).

## 60 2 Pedagogic trace-contract examples

62 Constraints on sequences of function calls are common. Sometimes these constraints cover  
63 just one function, but more commonly they involve several. In a functional language such  
64 as Racket, they also govern higher-order functions. This section introduces the Racket  
65 implementation of trace contracts with pedagogic examples of such constraints. It demon-  
66 strates how the integration of trace contracts with Racket’s higher-order contract system  
67 facilitates authoring maintainable specifications.

### 69 2.1 A naive look at trace contracts

71 In 2020, a developer reported a bug to Racket’s mailing list about the  
72 `current-memory-use` function.<sup>1</sup> The documentation states that the function “returns an  
73 estimate of the total number of bytes allocated since start up, including bytes that have  
74 since been reclaimed by garbage collection” (Flatt and PLT, 2010). Given this description,  
75 one might expect that the series of return values from `current-memory-use` would  
76 increase over time. However, a memory-consumption plot for a long-running system  
77 showed periodic dips.

78 In a language with a conventional type system, such as Java, this function would have  
79 the following signature:

```
81 // Returns the number of bytes allocated since start up,  
82 // including those deallocated during garbage collection.  
83 int currentMemoryUse();
```

85 The comment mentions two unchecked constraints. First, the function’s result cannot be  
86 negative, so `int` is imprecise. In Racket, the API author could improve on this type with a  
87 run-time-checked contract such as `(-> natural?)`. This notation denotes the signature of  
88 a function that takes no arguments and returns natural numbers. Second, the documentation  
89

90 <sup>1</sup> <https://groups.google.com/g/racket-users/c/xq0Y8uevGzE/m/mBtHeq2jAwAJ>  
91

implies that every call returns a number that is greater than or equal to the result of all previous calls. Existing contract systems cannot express this constraint easily.

With trace contracts, it is possible to express this second constraint directly:

```
(provide
  (contract-out
    [current-memory-use
      (trace/c ([y natural?])1
        (-> y)2
        (full (y) sorted?)3)]))
```

This contract captures both of the constraints that conventional type systems could not express. As the highlighting and subscripts indicate, a trace contract consists of three parts: (1) a sequence of *trace variable declarations*, including one behavioral contract for each; (2) a contract expression, dubbed the *body contract*; and (3) a sequence of *predicate clauses*, in this example introduced with `full`.

Here, there is a single trace variable, `y`, associated with `natural?`. The body contract is `(-> y)`, which specifies ordinary, single-call constraints placed on values protected by the trace contract. When a client module calls `current-memory-use`, the contract system ensures that the returned value is a natural number and, if so, collects the value in a data structure associated with `y`. This data structure is called a *trace*. Additionally, the trace contract specifies a `full` predicate clause that depends on `y`. For `full`, the trace data structure is a stream. Every time the contract system collects a value in the `y` trace, it applies the function specified in the predicate clause—`sorted?`—to the stream of values. The trace contract fails if `sorted?` returns `false`, indicating a dip in the sequence.

Note that `sorted?` is a pure function in the host language, just like ordinary first-order behavioral contracts. One immediate advantage is that a developer can test contracts like any other piece of code—an important property considering that all code, including specification code, may have bugs. Testing builds confidence in the correctness of the specification itself.

With this contract in place, violations are detected as soon as they occur. Moreover, the trace contract blames the appropriate party for the violation:

```
> (current-memory-use)
100
> (current-memory-use)
200
> (current-memory-use)
; current-memory-use: broke its own contract
;   produced: 0
;   ...
;   blaming: current-memory-use
```

In this interaction, `current-memory-use` returns increasing values for the first two calls. On the third call it produces 0, causing a contract error. Since the problematic value was

collected from the module that defined `current-memory-use`, the function itself is to blame. Developers confronted with this error message can immediately report a bug in the run-time library, knowing with confidence that their code is not responsible for the fault.

## 2.2 A less naive look: tolerable performance

In its current form, the `current-memory-use` contract comes with a steep performance cost. While any contract can slow down a program, naive trace contracts can be especially expensive because they execute code every time a value is added to a trace. Programmers should be mindful of this expense. In particular, `sorted?` iterates through the entirety of the `y` trace every time a new value is collected. Thus, checking this trace contract is quadratic in the number of calls to `current-memory-use`. To reduce this overhead, a trace-contract system must hand developers fine-grained control over the trace data structure.

Fine-grained control means that developers can choose a custom representation of the trace instead of the naive, stream data structure. When choosing, a developer must: (1) decide on a data structure, (2) pick an initial value, and (3) supply an operation that incorporates a value into the existing trace representation or signals a failure. This kind of predicate clause is introduced with `accumulate` and the data structure is referred to as the *accumulator*. Note that the function given to `accumulate` is no longer a predicate. Instead, it receives two values: the current accumulator and the newly collected values. It returns the new accumulator on success or a designated failure value otherwise.

For the running example, it suffices to use a single number as the accumulator. A simple comparison between any collected value and the accumulator is enough to enforce the promised behavior:

```
(trace/c ([y natural?])
  (-> y)
  (accumulate 0
    [(y) (λ (acc cur)
      (if (<= acc cur) cur (fail)))]))
```

The `accumulate` clause specifies an initial accumulator value of 0 and an accumulating function. When `y` receives a new value, the latter is applied to the current accumulator and the latest value. If the current accumulator is smaller than the new value, then the new value is returned and becomes the next accumulator.<sup>2</sup> Otherwise, the function's result is `(fail)`, the designated failure value.

Every trace contract can be expressed with `accumulate` instead of `full`. In fact, `full` is just syntactic sugar over an `accumulate` clause with a stream accumulator. While `full` is a useful tool to understand trace contracts conceptually, in practice programmers should almost always use `accumulate` combined with an efficient trace data structure.

<sup>2</sup> If `current-memory-use` were to return a non-numeric result, an error would be raised even without the `natural?` check on `y` because `<=` expects two numbers. The error message, however, would blame the contract itself for violating the precondition of `<=`, instead of `current-memory-use`. Thus, to generate practical error messages, the `natural?` check must remain.

### 2.3 Checking all calls to one function

Consider a compiler pass that computes a live-variables analysis via fixed-point iteration. The interface to such an analysis, using ordinary contracts, might look like this:

```

185
186
187
188
189 (provide
190   (contract-out
191     ;; The transfer function must be monotonically increasing.
192     [live-vars (-> (-> set? set?) label? set?)]))
193

```

Given a monotonically increasing transfer function and a program label, `live-vars` returns the set of live variables at that label (Nielson et al., 2005). Unlike the simplistic example from the preceding section, this constraint involves a higher-order function. A comment describes the constraint, but it is not enforced. Since an incorrectly computed least fixed point can lead to a silent failure, this problem may be especially difficult to debug.

A trace contract can replace the informal comment, enforcing monotonicity:

```

200
201
202 (provide
203   (contract-out
204     [live-vars (-> (monotone/c set? set? subset?) label? set?)]))
205
206 ;; Contract Contract (Set Set -> Boolean) -> Contract
207 (define (monotone/c dom/c cod/c leq?)
208   (trace/c ([x dom/c] [y cod/c])
209     (-> x y)
210     (accumulate (red-black-tree leq?)
211       [(x y) (monotone-func leq?)])))
212

```

The `monotone/c` function consumes two contracts and a comparison function; it returns a function contract that checks monotonicity with respect to the given comparison function. When a client module imports `live-vars` and invokes it, the highlighted contract is attached to the supplied transfer function. This contract stipulates that the transfer function takes and returns sets and is monotone with respect to set inclusion. During fixed-point iteration, the trace contract observes all input-output pairs of the transfer function and builds an extensional representation of the function. Violations are detected by ensuring that no two input-output pairs fail monotonicity.

While a stream containing all input-output pairs would work, it would be inefficient. An order-aware data representation can reduce the time needed to determine whether monotonicity holds from  $O(n^3)$  to  $O(n \log n)$ , where  $n$  is the number of calls to the transfer function. One possible choice is a red-black tree as it can quickly determine the immediate predecessor and successor of an ordered element.<sup>3</sup>

<sup>3</sup> Ordinarily this works only for a total order, not a partial order such as set inclusion. However, since fixed-point iteration always explores comparable elements, a red-black tree is acceptable. A general-purpose contract for monotonicity that supports partial orders would require a different data structure. Assuming that fixed-point iteration climbs the lattice *in order*, as it usually does, a contract like the one from section 2.2 would also work.

185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230

Every time the trace contract monitors a new value, it initializes a new accumulator. If `live-vars` is invoked twice, two separate accumulators are created, one for each given transfer function. This policy allows trace contracts to compose sensibly with other contract combinators.

Here is the (curried) function that finishes the definition of `monotone/c`:

```

231
232
233
234
235
236 ;; Acc = [Ordered-Dict Set Set]
237 ;; Comparator -> (Acc Set Set -> [Or Acc Fail])
238 (define ((monotone-func leq?) acc x y)
239   (cond
240     [(dict-has-key? acc x)
241      (if (equal? y (dict-ref acc x)) acc (fail))]
242     [else
243      (define pred-y (dict-pred acc x))
244      (define succ-y (dict-succ acc x))
245      (if (and (=> pred-y (leq? pred-y y))
246             (=> succ-y (leq? y succ-y)))
247          (dict-set acc x y)
248          (fail)))]))
249
250

```

When the transfer function returns, `monotone-func` is applied to the current accumulator `acc`, the latest input `x`, and the latest output `y`. It determines the transfer function's predecessor and successor results for `x` and, if they exist, checks that they properly relate to the current output `y`. Just two comparisons suffice: by transitivity there are no other monotonicity violations. If successful, `monotone-func` returns the next accumulator, relating the new input-output pair in the augmented red-black tree.

## 2.4 Global initialization of traces

The following warning from Racket's documentation tells developers about an essential constraint that the language does not enforce:

*“If a key in an equal?-based hash table is mutated (e.g., a key string is modified with string-set!), then the hash table's behavior for insertion and lookup operations becomes unpredictable.”*

Time and again, however, programmers—especially novices—fail to heed this warning, experience arbitrary program behavior, and have a difficult time debugging such mistakes. Trace contracts can enforce such constraints:

```

269 (provide
270   (contract-out
271     [hash-set hash-set/c]
272     [string-set! (-> mutable/c natural? char? void?)]))
273
274
275
276

```

```

277 (define-values (hash-set/c mutable/c)
278   (trace/c ([t any/c])
279     #:global
280     (values (-> hash? (list/t 'set t) any/c void?)
281             (list/t 'mut t))
282     (full (t) not-interfere?)))

```

283 This trace contract makes use of a few features. First, the body contract produces two  
284 values using Racket’s values function, which allows an expression to return multiple  
285 values (Ashley and Dybvig, 1994). Because the property relates different functions, i.e.  
286 `hash-set` and `string-set!`, their contracts need to be created within the same `trace/c`.  
287 Second, the `#:global` option causes the state of the trace contract to be initialized at  
288 *definition* time, not the usual *attachment* time. Without `#:global`, the `hash-set/c` and  
289 `mutable/c` contracts would be initialized separately and could never interact. Finally, the  
290 `list/t` function alters the given collector to tag incoming values with a symbol. Here, the  
291 symbol is used to indicate the operation.

292 The `not-interfere?` predicate ensures that no key is modified after it becomes a key  
293 in a hash table:

```

294
295 (define/match (not-interfere? xs)
296   [((stream))
297    true]
298   [((stream* '(mut ,x) xt))
299    (not-interfere? xt)]
300   [((stream* '(set ,x) xt))
301    (and (not (stream-member? xt '(mut ,x)))
302         (not-interfere? xt)))]
303
304

```

## 305 2.5 The full grammar of trace contracts

306 In summary, the trace contract library extends Racket’s grammar with a `trace/c` form  
307 that constructs trace contracts. Figure 1 displays the extension to Racket’s grammar. As the  
308 preceding examples motivate, each piece of the trace contract (trace variable declarations,  
309 the body contract expression, and predicate clauses) come with enhancements that make  
310 the system practical:

312 **Trace Variable Declarations** The trace-variable declarations  $[x e_i]$  determine how many  
313 traces the contract creates. Each declaration comes with a contract  $e_i$  that governs  
314 newly collected values.

316 **Body-Contract Expression** When a trace contract is attached to a value, the body-  
317 contract expression  $e_b$  is evaluated in an environment where trace variables are  
318 bound to *collectors*. A *collector* is a contract that gathers values that flow through  
319 the corresponding points in the body contract. These points are called *interception*  
320 *points*, e.g., argument or return positions. Once collected, values are added to all  
321 dependent trace data structures.

```

323   e ∈ Expr = ... | (trace/c ([x et] ...+) o eb c ...+)
324   o ∈ Opt = #:global | ε
325   c ∈ Clause = (accumulate e [(x ...+) ea] ...+)
326               | (full (x ...+) ep)
327               | (track e c ...+)
328   et, eb ∈ Exprκ = {e | e evaluates to a contract}
329   ea ∈ Expra = {e | e evaluates to an accumulating function}
330   ep ∈ Exprp = {e | e evaluates to a predicate}
331   x ∈ Var

```

Fig. 1. The Extended Racket Grammar for Trace Contracts

If `trace/c` comes with the `#:global` option, then the collectors are initialized only once, namely, when the contract is created. The default behavior, as demonstrated in section 2.3, initializes collectors each time the trace contract is attached to a value.

The body-contract expression may produce multiple values, which is useful in conjunction with `#:global`. Programmers should use the `#:global` option when more than one contract must share a trace or multiple traces, as seen in section 2.4.

**Predicate Clauses** A predicate clause  $c$  is responsible for determining how the trace should be updated when a new value is collected and whether the contract is violated. The implementation supports three types: `accumulate`, `full`, and `track`.

The `accumulate` clause consists of several subclauses that determine how the accumulator is updated when a new value is collected. A subclause consists of a dependency specification and an expression  $e_a$ , which must evaluate to a function. When a subclause depends on more than one collector, the contract system waits until all values have been collected before applying the function. If a collector receives more than one value before the other collectors are ready, then all but the last are discarded.<sup>4</sup> The corresponding accumulating function must return either an updated accumulator or a value indicating failure.

The `full` clause evaluates the expression  $e_p$  to a predicate and applies this predicate to a time-ordered stream of collected values. Instead of triggering when *all* the dependent collectors have new values, the predicate is applied when *any* of the dependent collectors have new values.

The `track` clause augments the error message of other clauses with information about all the parties that contributed values to the trace. Section 7.1 describes this feature in detail.

### 3 Real-world trace-contract examples

This section provides two real-world examples of trace contracts. The first comes from Racket’s drawing library and the second comes from code written as part of the grading infrastructure for an undergraduate course.

<sup>4</sup> Alternative choices are expressible by having multiple `accumulate` subclauses with one dependency each. The accumulator would store collected values and then the accumulating function would determine the policy.



### 3.1 Reusing trace contracts

Racket comes with a built-in library, `racket/draw`, for drawing images. The library provides a thin wrapper around a low-level graphics API written in C. As such, the wrapper must protect against client behavior that would induce undefined behavior at the C level. One instance of undefined behavior occurs with drawing context (DC) objects.

To produce an image with `racket/draw`, a developer must first choose a DC representing the desired output device. There are many such contexts, but they all share a common interface. Part of this interface is a collection of methods that manages the pages of a document: `start-doc`, `start-page`, `end-page`, `end-doc`. Clients must call these methods in a particular order. It does not make sense to call, e.g., `end-doc` before `start-doc`. Moreover, all drawing commands must occur within a page.

Here is a regular expression that describes a valid *complete* sequence of method calls:

```
start-doc, (start-page, draw*, end-page)*, end-doc
```

This regular expression is not suitable for trace-contract monitoring. A trace contract also checks every *incomplete* sequence of method calls, not just the complete sequence. So, this regular expression has to be adapted to accept any prefix of the complete sequence.

Here is an adapted version of the regular expression above, described using Racket's automata library (McCarthy, 2011):

```
(define SINGLE-PAGE
  (re (seq/close 'start-page (star 'draw) 'end-page)))
(define DC-RE
  (re (seq/close 'start-doc (star ,SINGLE-PAGE) 'end-doc)))
```

The `re` form compiles a finite-state automaton that accepts the given regular expression. Within `re`, `seq/close` denotes a regular expression that accepts not just the given sequence, but any prefix of that sequence.

The following trace contract enforces the protocol using DC-RE:

```
(provide
  (contract-out [make-ps-dc (-> (dc/c DC-RE))]))
(define (dc/c aut)
  (trace/c ([s symbol?])
    (object/c
      [start-doc (apply/c [s 'start-doc])]
      [start-page (apply/c [s 'start-page])]
      [draw-point (apply/c [s 'draw])]
      [end-page (apply/c [s 'end-page])]
      [end-doc (apply/c [s 'end-doc'])])
    (accumulate aut
      [(s) (λ (acc x)
              (define acc* (acc x))
              (if (machine-accepting? acc*) acc* (fail))))]))))
```

Given a finite-state automaton, `dc/c` produces a contract for a DC where the method call sequence is governed by the regular expression. In the body of `dc/c`, a trace contract is wrapped around an object contract specifying each of the DC methods. There is only a single collector, `s`, that collects symbols corresponding to the method calls. The `apply/c` combinator provides the collector with a constant value each time a protected method is called. To check the protocol, the trace predicate uses the state of the automaton as the accumulator. So long as the automaton is accepting, the contract is satisfied. The trace contract is then used in the codomain of `make-ps-dc`, which produces PostScript (PS) drawing contexts.

As mentioned before, there is more than one kind of DC. In particular, an Encapsulated PostScript (EPS) drawing context has a slightly different constraint than an ordinary PS context. Since an EPS file is intended to be embedded in a larger document, it can only have a single page. Supporting EPS is easy since `dc/c` abstracts over the regular expression. Checking a different protocol requires only passing in a different regular expression to `dc/c`:

```
(provide (contract-out [make-eps-dc (-> (dc/c EPS-RE))]))
(define EPS-RE
  (re (seq/close 'start-doc ,SINGLE-PAGE 'end-doc)))
```

### 3.2 Protocols for many methods

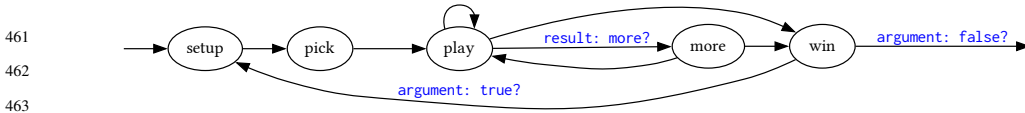
Imagine a board-game framework that pits AI player components against one another. In a typical board game, players (1) receive their game pieces; (2) take turns, which may consist of several interactions with the board; and (3) determine which ones won and lost. Winners of a game can move on to the next round of a tournament while losers are left behind.

A natural implementation of an AI player is as an object with methods that correspond to these game stages. Each player expects that these methods are called in a certain order, which may depend on the state of the game. In short, the methods relate to each other according to a value-dependent, multi-function, temporal property.

Programmers often use state-transition diagrams to document such multi-function protocols. Figure 2 displays a diagram for an AI board-game player (top), together with a matching trace-contract specification (bottom). States in this diagram indicate which method the referee component must call next. Labeled edges represent transitions that depend on either an argument value or a return value. Unlabeled edges represent independent transitions. Since there are several possible transitions for some states, this is a non-deterministic automaton.

Specifically, this diagram dictates that players must implement five methods:

1. A `setup` method that delivers the game pieces.
2. A `pick` method that asks a player to choose some game objectives.
3. A `play` method that grants a player the right to take a turn. The result is either a request to perform an *action* on the game state or a request for *more* game pieces.



```

461
462
463
464
465 (provide (contract-out [player-factory (-> strategy/c player/c)]))
466
467 (define PLAYER-NFA
468   (nfa (setup) (setup pick play more win done)
469     [setup (['(setup ,_) (pick)])]
470     [pick (['(pick ,_) (play)])]
471     [play (['(play ,(? action?)) (play win)]
472           ['(play ,(? more?)) (play more win)])]
473     [more (['(more ,_) (play win)])]
474     [win (['(win ,true) (setup)]
475          ['(win ,false) (done)])]
476     [done ()])
477
478 (define player/c
479   (trace/c ([x any/c])
480     (object/c
481       [setup (->m game-map? (list/t 'setup x))]
482       [pick (->m set? (list/t 'pick x))]
483       [play (->m state? (list/t 'play x))]
484       [more (->m list? (list/t 'more x))]
485       [win (->m (list/t 'win x) any/c)])
486     (accumulate PLAYER-NFA
487       [(x) (λ (acc x)
488             (define acc* (acc x))
489             (if (machine-accepting? acc*) acc* (fail))))]))
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506

```

Fig. 2. The state-machine contract for AI players, with a transition diagram

4. If the referee gets this second kind of request in response to play, it may invoke the player's more method. But, it may also skip this call, depending on the game state.
5. The player is granted turns and more pieces until the referee discovers an end-game condition and then informs the player whether it won or lost. The player may participate in the next game only if win is called with true.

In this particular software system, a factory function creates AI players from a strategy and returns player objects that implement the above five methods. The contract on this factory method attaches a trace contract to each player object. As a result, every instance of the player class must obey the order of method calls specified in the sequence diagram. Otherwise, the system raises an error with a blame-assignment message that informs the

507 developer of the player that was mistreated; once again, the trace-contract design greatly  
508 benefits from a tight integration with higher-order behavioral contracts.

509 This protocol is a language over the alphabet containing the names of methods, along  
510 with the specific arguments or return values of two of them: `play` and `win`. For example,  
511 the following sequence of method calls is correct so long as `play` returned a value satis-  
512 fying the `more?` predicate: `setup, pick, play, more`. If `play` returned a value satisfying  
513 `action?` instead, then that sequence of method calls is invalid, and a contract error should  
514 be raised on the call to `more`.

515 To check this protocol, the trace contract once again simulates the finite-state machine  
516 with `accumulate`. Unlike the automaton in section 3.2, this machine inspects pieces  
517 of data. For `setup`, `pick`, and `more`, the transition is independent of run-time values.  
518 However, `play` and `win` have value-dependent transitions. For example, `play` uses the  
519 `action?` and `more?` predicates to determine the next set of states. It does so using Racket's  
520 `(? p)` match pattern, which matches a value if the predicate `p` holds.

### 521 3.3 Contracts are better than ad-hoc checks

522 As mentioned previously, these two examples come from real-world projects. In the origi-  
523 nal code, both contained ad-hoc protocol checks instead of trace contracts. Given that, it is  
524 worth reviewing why contracts are preferable to such handwritten checks:  
525  
526  
527  
528  
529  
530

- 531 1. Contracts cleanly separate specification code and implementation code—with ad-  
532 hoc checks the two are intertwined. This makes programs difficult to read, and thus  
533 hard to maintain (Meyer, 1988, 1992). Additionally, the code needed to correctly  
534 check a specification is often repetitive and tedious. Getting it wrong is inevitable.
- 535 2. As a direct consequence of separating specification and implementation, contracts  
536 enable static and dynamic analyses. For example, the contract library supports pro-  
537 filing (Andersen et al., 2018) to determine which contracts are slowing down a  
538 program. Static techniques (Nguyễn et al., 2018) can verify whether a program  
539 satisfies a contract. These kinds of tools are impossible with ad-hoc checks.
- 540 3. The contract library automatically supports detailed error messages with blame that  
541 points to the module that violated the contract. This information is exceptionally  
542 useful for debugging (Lazarek et al., 2020).
- 543 4. Programmers have fine-grained control over the scope of a contract, i.e., which mod-  
544 ules get checks and which ones do not. Trusted modules may not need checks. Thus,  
545 the balance between correctness and performance can be tuned precisely. This also  
546 allows tools to automatically bypass contracts in certain cases, for instance, when  
547 they are statically proven to be unnecessary (Moy et al., 2021).
- 548 5. Finally, contracts permit specification reuse. In section 3.1, repetitive blocks of ad-  
549 hoc checking code are replaced with `make-ps-dc` and `make-eps-dc`; abstracting  
550 over the contract eliminates duplicate code.  
551  
552

$\Lambda$  Surface Syntax

$e \in \text{Expr} = b \mid x \mid f \mid oe \mid \text{if } eee$   
 $\quad \mid ee \mid \text{queue} \mid \text{add! } ee$   
 $b \in \text{Bool} = \text{true} \mid \text{false}$   
 $f \in \text{Fun} = \lambda x.e$   
 $o \in \text{Op} = \text{null?} \mid \text{head} \mid \text{tail}$   
 $x, y, z \in \text{Var}$

 $\Lambda$  Evaluation Syntax

$\zeta \in \text{Conf} = \langle e, \sigma \rangle$   
 $e \in \text{Expr} = \dots \mid \alpha \mid \text{err}_j^k$   
 $v \in \text{Val} = b \mid f \mid \alpha$   
 $E \in \text{Ctx} = \square \mid oE \mid \text{if } Eee \mid Ee$   
 $\quad \mid vE \mid \text{add! } Ee \mid \text{add! } vE$   
 $u \in \text{SVal} = \text{null} \mid \text{cons } v \alpha$   
 $\sigma \in \text{Store} = \text{Addr} \rightarrow_{\text{fin}} \text{SVal}$   
 $\alpha \in \text{Addr}$   
 $j, k, l \in \text{Lab}$

Fig. 3. Surface and Evaluation Syntax of  $\Lambda$ 

## 4 A model of trace contracts

A design requires a rigorous blueprint so that implementors of other languages can understand the idea and adapt it. This section presents a model of the  $\lambda$ -calculus extended with trace contracts. To keep the formalism accessible, the model is developed and explained incrementally using five languages:  $\Lambda$ ,  $\Lambda_B$ ,  $\Lambda_C$ ,  $\Lambda_T$ ,  $\Lambda_U$ . Additionally, some of the pragmatic features of section 2 have been omitted to reduce the complexity of the final model. Section 5 presents formal properties of these models.

### 4.1 A functional base

Figure 3 (left) defines the surface syntax of  $\Lambda$ , the call-by-value  $\lambda$ -calculus (Plotkin, 1975) extended with Booleans and mutable queues. The final model represents traces using queues. The nullary constructor `queue` builds a new instance and `add!` puts an element into a queue. Primitive operations allow functions to walk over queues similar to immutable lists. All the remaining syntax is standard.

Figure 3 (right) defines the evaluation syntax of  $\Lambda$ . Along with a grammar of values and evaluation contexts, the syntax contains errors and queue-specific stores.

Errors come with two labels:  $j$  names the party that specified the violated contract and  $k$  names the party that violated the contract. There are two special labels:  $\circ$  refers to the language runtime itself and  $\dagger$  refers to the read-eval-print-loop (REPL). Since  $\Lambda$  does not have user-defined contracts, the only possible error is  $\text{err}_{\circ}^{\dagger}$ .

Stores map addresses to either an empty queue (`null`) or a cons cells that combines a head value with an address containing the remaining elements. This choice facilitates functional iteration over queues.

Next, figure 4 defines the reduction relation for  $\Lambda$  with the supporting metafunctions provided in figure 5. Conditionals and application are standard. For functional primitive operations, the  $\delta$  metafunction (Barendregt, 1981) is used to compute the result. Constructing a new queue uses the next free address in the store and sets it to the empty queue. Adding to an existing queue updates the store, replacing the empty queue at the end

**$\Lambda$  Reduction Relation**

599	$\langle E[\text{if } v \ e_1 \ e_2], \sigma \rangle \mapsto \langle E[e_1], \sigma \rangle$ if $v \neq \text{false}$	IF-TRUE
600	$\langle E[\text{if false } e_1 \ e_2], \sigma \rangle \mapsto \langle E[e_2], \sigma \rangle$	IF-FALSE
601	$\langle E[(\lambda x.e) \ v], \sigma \rangle \mapsto \langle E[e[v/x]], \sigma \rangle$	APP
602	$\langle E[o \ v], \sigma \rangle \mapsto \langle E[\delta(o, v, \sigma)], \sigma \rangle$	PRIM
603	$\langle E[\text{queue}], \sigma \rangle \mapsto \langle E[\alpha], \sigma[\alpha \mapsto \text{null}] \rangle$ if $\alpha = \text{next}(\sigma)$	QUEUE
604	$\langle E[\text{add! } \alpha \ v], \sigma \rangle \mapsto \langle E[\alpha], \text{add}(\sigma, \alpha, v) \rangle$	ADD!
605	$\langle E[v_f \ v], \sigma \rangle \mapsto \langle E[\text{err}_o^\dagger], \sigma \rangle$ if $v_f \notin \text{Fun}$	ERR-APP
606	$\langle E[\text{add! } v_q \ v], \sigma \rangle \mapsto \langle E[\text{err}_o^\dagger], \sigma \rangle$ if $v_q \notin \text{Addr}$	ERR-ADD!
607	$\langle E[\text{err}_j^k], \sigma \rangle \mapsto \langle \text{err}_j^k, \sigma \rangle$ if $E \neq \square$	ERR-UNWIND
608		
609		
610		
611		

Fig. 4. Reduction Relation of  $\Lambda$ 

$$\begin{aligned}
\text{add}(\sigma, \alpha, v) &= \begin{cases} \sigma[\alpha \mapsto \text{cons } v \ \alpha'][\alpha' \mapsto \text{null}] & \text{if } \alpha' = \text{next}(\sigma), \sigma(\alpha) = \text{null} \\ \text{add}(\sigma, \alpha', v) & \text{if } \sigma(\alpha) = \text{cons } v_h \ \alpha' \end{cases} \\
\text{next}(\sigma) &= \max(\text{dom}(\sigma)) + 1 \\
\delta(o, v, \sigma) &= \begin{cases} \text{true} & \text{if } o = \text{null?}, \sigma(v) = \text{null} \\ \text{false} & \text{if } o = \text{null?}, \sigma(v) = \text{cons } v \ \alpha' \\ \text{err}_o^\dagger & \text{if } o = \text{head}, \sigma(v) = \text{null} \\ v & \text{if } o = \text{head}, \sigma(v) = \text{cons } v \ \alpha' \\ \text{err}_o^\dagger & \text{if } o = \text{tail}, \sigma(v) = \text{null} \\ \alpha' & \text{if } o = \text{tail}, \sigma(v) = \text{cons } v \ \alpha' \\ \text{err}_o^\dagger & \text{if } v \notin \text{Addr} \end{cases}
\end{aligned}$$

Fig. 5. Metafunctions of  $\Lambda$ 

with a cons cell containing the new value. The last three rules deal with error conditions. Errors to do with primitive operations are handled by  $\delta$  itself.

## 4.2 The classic contract model

Figure 6 defines the surface and evaluation syntax for  $\Lambda_B$ , a model of higher-order contracts based on that of Dimoulas and Felleisen (2011) and Dimoulas et al. (2011). The surface syntax extends  $\Lambda$  with two new elements: dependent function contracts  $e_d \rightarrow_i e_c$  and monitors  $\text{mon}_j^{k,l} e_\kappa e_c$ . A dependent function contract can describe properties of functions where the codomain contract depends on the argument to the protected function.<sup>5</sup> A monitor is then used to attach a contract to a value. So,  $\text{mon}_j^{k,l} e_\kappa e_c$  attaches  $e_\kappa$  to  $e_c$ . The value of  $e_c$

<sup>5</sup> This paper uses the abbreviation  $e_d \rightarrow e_c$  to stand for an independent function contract, i.e.,  $e_d \rightarrow_i (\lambda\_e.c)$ .

$$\begin{array}{l}
\boxed{\Lambda_B \text{ Surface Syntax}} \text{ extends } \Lambda \\
e \in \text{Expr} = \dots \mid e \rightarrow_i e \mid \text{mon}_j^{k,l} e e \\
j, k, l \in \text{Lab} \\
\boxed{\Lambda_B \text{ Evaluation Syntax}} \text{ extends } \Lambda \\
v \in \text{Val} = \dots \mid \kappa \\
\kappa \in \text{Con} = b \mid \lambda x.e \mid v \rightarrow_i v \\
E \in \text{Ctx} = \dots \mid E \rightarrow_i e \mid v \rightarrow_i E \\
\quad \mid \text{mon}_j^{k,l} E e \mid \text{mon}_j^{k,l} v E
\end{array}$$

Fig. 6. Surface and Evaluation Syntax of  $\Lambda_B$ 

is dubbed the *carrier* of the contract. Monitors also come with labels naming the parties that agreed to the contract: the contract-defining module  $j$ , the server module  $k$ , and the client module  $l$ .

In addition to dependent function contracts, the evaluation syntax reveals that Booleans and functions can be used as contracts. When used as a contract, `true` permits any value and `false` forbids all values. These correspond to Racket’s `any/c` and `none/c` contracts, respectively. When used as a contract, a function checks first-order properties of the carrier. This corresponds to Racket’s flat contracts.

Here is an example program with a contract:<sup>6</sup>

$$\text{mon}_{\text{ctc}}^{\text{lib,main}} (\text{true} \rightarrow_i (\lambda x. \lambda y. x = y)) (\lambda z. z) \quad (4.1)$$

This example contains a contract fully specifying the behavior of the identity function. Since the domain contract is `true`, every argument is accepted. When the function returns, the output value is checked against the codomain contract  $\lambda y. x = y$ , ensuring that it is equal to the input value.

Figure 7 shows the reduction relation for  $\Lambda_B$ . The first four rules describe the checks performed by each kind of contract. For `true` and `false`, the check immediately succeeds or immediately fails, respectively. For a flat contract  $\lambda x.e$ , the result of applying this function to the carrier is then used as the new contract. Thus, if  $\lambda x.e$  is a predicate, this corresponds exactly to a first-order check because `true` and `false` are themselves contracts.

While  $\lambda x.e$  may return a Boolean, there is nothing in the semantics that forces it to be one. In particular, it could return a function contract. This can be used to create *cascading contracts* that combine arbitrary first-order checks with higher-order contracts.

Consider this example:

$$\lambda f. \text{if} (\text{arity } f = 1) (\text{int?} \rightarrow \text{int?}) \text{false}$$

Assuming an arity primitive, this cascading contract checks a first-order constraint, namely that the carrier has arity one. If successful, the higher-order contract `int? → int?` protects the carrier. Otherwise, the contract fails.

In Racket, function contracts perform arity checks eagerly, exactly in this manner. The model from Dimoulas and Felleisen (2011) cannot encode this behavior. Cascading contracts are essential for defining the compiler in section 6.

<sup>6</sup> These example programs are intended to illustrate a point, and therefore may use language features that are not formally defined. The meaning should always be clear from context.

$\Lambda_B$  Reduction Relation extends  $\Lambda$

691	$\langle E[\text{mon}_j^{k,l} \text{ true } v], \sigma \rangle \mapsto \langle E[v], \sigma \rangle$	MON-TRUE
692		
693	$\langle E[\text{mon}_j^{k,l} \text{ false } v], \sigma \rangle \mapsto \langle E[\text{err}_j^k], \sigma \rangle$	MON-FALSE
694		
695	$\langle E[\text{mon}_j^{k,l} (\lambda x.e) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^{k,l} ((\lambda x.e) v) v], \sigma \rangle$	MON-FLAT
696	$\langle E[\text{mon}_j^{k,l} (v_d \rightarrow_i v_c) v], \sigma \rangle \mapsto \langle E[\lambda x.\text{let } x_j = \text{mon}_j^{l,j} v_d x \text{ in}$	MON-FUN
697	$\text{let } x_k = \text{mon}_j^{l,k} v_d x \text{ in}$	
698	$\text{mon}_j^{k,l} (v_c x_j) (v x_k)], \sigma \rangle$	
699		
700	$\langle E[\text{mon}_j^{k,l} v_\kappa v], \sigma \rangle \mapsto \langle E[\text{err}_o^\dagger], \sigma \rangle$ if $v_\kappa \notin \text{Con}$	ERR-MON
701		

Fig. 7. Reduction Relation of  $\Lambda_B$

Finally, MON-FUN describes the indy semantics of dependent function contracts (Dimoulas et al., 2011). The key insight of indy is that the contract itself can be inconsistent, and therefore must be subject to checks.

Here is an example that illustrates this point:

$$(\text{bool?} \rightarrow \text{bool?}) \rightarrow_i (\lambda f.f \ 42)$$

While the domain contract states that the input is a function over Booleans, generating the codomain contract violates that assumption by applying  $f$  to a number. In this case, indy raises an error blaming the contract itself.

### 4.3 A revised contract model

As is,  $\Lambda_B$  cannot accommodate contracts with effects, such as trace contracts. When used as the domain of a function, a contract's effects are erroneously duplicated.

Take the following variation on program (4.1):

$$\text{mon}_{\text{ctc}}^{\text{lib,main}} ((\lambda x.\text{print } x; \text{true}) \rightarrow_i (\lambda x.\lambda y.x = y)) (\lambda z.z)$$

The only difference is the presence of an effect in the domain contract. As the following reduction sequence demonstrates, `print` is executed twice:

$$\langle (\text{mon}_{\text{ctc}}^{\text{lib,main}} ((\lambda x.\text{print } x; \text{true}) \rightarrow_i (\lambda x.\lambda y.x = y)) (\lambda z.z)) \ 42, \emptyset \rangle$$

By MON-FUN, the monitor produces a wrapper function that checks the arguments against the domain contract and the return value against the codomain contract.

$$\begin{aligned} \mapsto & \langle (\lambda x.\text{mon}_{\text{ctc}}^{\text{lib,main}} ((\lambda x.\lambda y.x = y) (\text{mon}_{\text{ctc}}^{\text{main,ctc}} (\lambda y.\text{print } y; \text{true}) x)) \\ & ((\lambda z.z) (\text{mon}_{\text{ctc}}^{\text{main,lib}} (\lambda y.\text{print } y; \text{true}) x))) \ 42, \emptyset \rangle \end{aligned}$$

The wrapper function is applied to 42.

$$\begin{aligned} \mapsto & \langle \text{mon}_{\text{ctc}}^{\text{lib,main}} ((\lambda x.\lambda y.x = y) (\text{mon}_{\text{ctc}}^{\text{main,ctc}} (\lambda y.\text{print } y; \text{true}) 42)) \\ & ((\lambda z.z) (\text{mon}_{\text{ctc}}^{\text{main,lib}} (\lambda y.\text{print } y; \text{true}) 42)), \emptyset \rangle \end{aligned}$$



To produce the codomain contract, the argument is first checked against the domain contract with the contract-defining party (ctc) as the client label. This prints 42.

737  $\mapsto^+ \langle \text{mon}_{\text{ctc}}^{\text{lib,main}} ((\lambda x. \lambda y. x = y) 42)$   
 738  $\quad ((\lambda z. z) (\text{mon}_{\text{ctc}}^{\text{main,lib}} (\lambda y. \text{print } y; \text{true}) 42)), \emptyset \rangle$

Once the argument is checked, the codomain contract can be created.

743  $\mapsto \langle \text{mon}_{\text{ctc}}^{\text{lib,main}} (\lambda y. 42 = y)$   
 744  $\quad ((\lambda z. z) (\text{mon}_{\text{ctc}}^{\text{main,lib}} (\lambda y. \text{print } y; \text{true}) 42)), \emptyset \rangle$

The argument has to be checked against the domain contract once more. This time the client label is lib. Again, 42 is printed.

746  $\mapsto^+ \langle \text{mon}_{\text{ctc}}^{\text{lib,main}} (\lambda y. 42 = y) ((\lambda z. z) 42), \emptyset \rangle$

Now the carrier is applied to 42. Since the carrier is the identity function, it returns 42.

750  $\mapsto^+ \langle \text{mon}_{\text{ctc}}^{\text{lib,main}} (\lambda y. 42 = y) 42, \emptyset \rangle$

The returned value is checked against the generated codomain contract. In this case, the contract is satisfied and is discharged.

752  $\mapsto \langle 42, \emptyset \rangle$

Effect duplication is a major problem for trace contracts. If a collector is used as the domain of a function, then it will collect duplicate values.

To understand the source of the problem, consider the contractum of MON-FUN. It contains two  $v_d$  monitors that differ only in their client label: one uses  $j$  and the other uses  $k$ . A simple let binding cannot be used to eliminate the duplicated effect since each of the monitors may produce wrappers that contain different labels.

The conclusion to draw is that  $\Lambda_B$  conflates *interception time* and *crossing time*. Interception time occurs when the contract system intercepts a value from the monitored program, i.e., when a value flows through an interception point. Crossing time occurs when an intercepted value moves to another component.

Consider a wrapper for the contract  $v_d \rightarrow v_c$ . Every time the wrapper is applied, it must perform two tasks related to the argument. First,  $v_d$  must be used to check first-order properties of the argument. Second, if  $v_d$  is a higher-order contract, wrappers must be created for every client of the argument. In the case of indy, there are two such clients, by convention labeled  $j$  and  $l$ . Interception time corresponds to when task one occurs and crossing time corresponds to when task two occurs.<sup>7</sup> Since  $\Lambda_B$  has only one mon form, both tasks are its responsibility.

Splitting the three-labeled monitor into two forms separates these responsibilities. Figure 8 defines the syntax of  $\Lambda_C$ , a revised contract language. While the surface syntax

<sup>7</sup> Often, interception-time coincides with first-order checks and crossing-time coincides with higher-order wrapping. There are exceptions, however. For example, in Racket the `unconstrained-domain->` contract makes no demand on function arguments. Because such a contract is guaranteed never to blame clients, its wrapper can be constructed at interception time. For simplicity, though, this paper blurs the distinction.

$\Lambda_C$ Surface Syntax extends $\Lambda$	$\Lambda_C$ Evaluation Syntax extends $\Lambda$
$e \in \text{Expr} = \dots \mid e \rightarrow_i e \mid \text{mon}_j^{k,l} e e$ $j, k, l \in \text{Lab}$	$e \in \text{Expr} = \dots \mid \text{mon}_j^k e e \mid \text{grd}_j^k \omega v$ $\mid e \cdot l$ $v \in \text{Val} = \dots \mid \kappa \mid \text{grd}_j^k \omega v$ $\kappa \in \text{Con} = b \mid \lambda x. e \mid v \rightarrow_i v$ $\omega \in \text{Wrap} = \text{true} \mid v \rightarrow_i v$ $E \in \text{Ctx} = \dots \mid E \rightarrow_i e \mid v \rightarrow_i E$ $\mid \text{mon}_j^k E e \mid \text{mon}_j^k v E \mid E \cdot l$

Fig. 8. Surface and Evaluation Syntax of  $\Lambda_C$ 

is the same as  $\Lambda_B$ , the evaluation syntax has a few differences (highlighted): two-labeled monitors  $\text{mon}_j^k e_\kappa e_c$ , guarded values  $\text{grd}_j^k \omega v$ , and label applications  $e_g \cdot l$ . Reduction of  $\text{mon}_j^k e_\kappa e_c$  corresponds to interception time, when first-order properties of the carrier are checked. Reduction of  $(\text{grd}_j^k \omega v) \cdot l$  corresponds to crossing time and produces a wrapper for client  $l$ .

Figure 9 displays the reduction relation for  $\Lambda_C$ . The first rule, MON-APPLY, decomposes the surface-level monitor into a two-labeled monitor applied to the client label. If successful, the two-labeled monitor produces a guarded value. The next four rules are responsible for the first-order checks of each contract. In the case of MON-TRUE and MON-FALSE, the first-order check is all that needs to occur.

Below the monitor rules, there are two rules for guarded values: GRD-TRUE and GRD-FUN. For true there is no wrapper needed so the carrier is produced directly. A wrapper is needed for function contracts, though. The wrapper in the contractum of GRD-FUN exploits the two-stage process. Instead of two  $v_d$  monitors, there is now only one, with its result bound to  $x_g$ . Effects caused by checking  $v_d$  occur only once while binding  $x_g$ . In the scope of this let binding, two wrappers are produced by applying  $x_g$  to the two client labels. Constructing these wrappers is not effectful.

#### 4.4 The trace contract model

Finally, figure 10 defines the trace contract model  $\Lambda_T$  that extends  $\Lambda_C$ . The surface syntax contains only one new form:  $\text{tr } e_\kappa e_p$ . This represents a trace contract with body-contract constructor  $e_\kappa$  and trace predicate  $e_p$ . A body-contract constructor is a function that, when provided with a collector, returns the body contract. The evaluation syntax contains one new form:  $\text{co } \alpha v_p$ . This represents a collector with trace address  $\alpha$  and trace predicate  $v_p$ .

The reduction relation for  $\Lambda_T$  is presented in figure 11. MON-TRACE performs two tasks. First, it allocates a queue for storing the trace. Second, it creates a collector and provides it to the body-contract constructor. MON-COL produces code that adds a new value to the trace and checks it using the trace predicate.

$\Lambda_C$  Reduction Relation extends  $\Lambda$

829		
830	$\langle E[\text{mon}_j^{k,l} e_\kappa e], \sigma \rangle \mapsto \langle E[(\text{mon}_j^k e_\kappa e) \cdot l], \sigma \rangle$	MON-APPLY
831	$\langle E[\text{mon}_j^k \text{true } v], \sigma \rangle \mapsto \langle E[\text{grd}_j^k \text{true } v], \sigma \rangle$	MON-TRUE
832		
833	$\langle E[\text{mon}_j^k \text{false } v], \sigma \rangle \mapsto \langle E[\text{err}_j^k], \sigma \rangle$	MON-FALSE
834	$\langle E[\text{mon}_j^k (\lambda x.e) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^k ((\lambda x.e) v) v], \sigma \rangle$	MON-FLAT
835	$\langle E[\text{mon}_j^k (v_d \rightarrow_i v_c) v], \sigma \rangle \mapsto \langle E[\text{grd}_j^k (v_d \rightarrow_i v_c) v], \sigma \rangle$	MON-FUN
836	$\langle E[(\text{grd}_j^k \text{true } v) \cdot l], \sigma \rangle \mapsto \langle E[v], \sigma \rangle$	GRD-TRUE
837		
838	$\langle E[(\text{grd}_j^k (v_d \rightarrow_i v_c) v) \cdot l], \sigma \rangle \mapsto \langle E[\lambda x. \text{let } x_g = \text{mon}_j^l v_d x \text{ in}$	GRD-FUN
839	$\text{let } x_j = x_g \cdot j \text{ in}$	
840	$\text{let } x_k = x_g \cdot k \text{ in}$	
841	$\text{mon}_j^{k,l} (v_c x_j) (v_c x_k)], \sigma \rangle$	
842		
843	$\langle E[\text{mon}_j^k v_\kappa v], \sigma \rangle \mapsto \langle E[\text{err}_o^\dagger], \sigma \rangle$ if $v_\kappa \notin \text{Con}$	ERR-MON
844		
845		

Fig. 9. Reduction Relation of  $\Lambda_C$

$\Lambda_T$  Surface Syntax extends  $\Lambda_C$

$\Lambda_T$  Evaluation Syntax extends  $\Lambda_C$

849		
850		
851	$e \in \text{Expr} = \dots \mid \text{tree } e$	$e \in \text{Expr} = \dots \mid \text{co } \alpha v$
852		$\kappa \in \text{Con} = \dots \mid \text{tr } v v \mid \text{co } \alpha v$
853		$E \in \text{Ctx} = \dots \mid \text{tr } E e \mid \text{tr } v E$
854		

Fig. 10. Surface and Evaluation Syntax of  $\Lambda_T$

$\Lambda_T$  Reduction Relation extends  $\Lambda_C$

859	$\langle E[\text{mon}_j^k (\text{tr } v_b v_p) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^k (v_b (\text{co } \alpha v_p)) v], \sigma[\alpha \mapsto \text{null}] \rangle$	MON-TRACE
860	if $\alpha = \text{next}(\sigma)$	
861		
862	$\langle E[\text{mon}_j^k (\text{co } \alpha v_p) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^k (v_p (\text{add! } \alpha v)) v], \sigma \rangle$	MON-COL
863		

Fig. 11. Reduction Relation of  $\Lambda_T$

Here is a translation of the current-memory-use example from section 2.1 into this model:

$$\text{tr } (\lambda y. \text{true} \rightarrow y) \text{ sorted?} \quad (4.2)$$

As mentioned earlier, the body-contract constructor consumes a collector  $k$  and returns a contract:  $\text{true} \rightarrow k$ . That is, the generated contract does not impose any precondition on

the argument of the carrier; the collector itself serves as the function's codomain contract. The trace predicate `sorted?` consumes and inspects a queue to ensure that it is sorted.<sup>8</sup>

Here is an example reduction sequence generated by protecting a function with this contract and applying it to `false`:

875  
876  
877  
878  
879  
880  $\langle \text{let } f = \text{mon}_{\text{lib}}^{\text{lib,main}} (\text{tr } (\lambda y. \text{true} \rightarrow y) \text{ sorted?}) (\lambda x. \text{---}) \text{ in } f \text{ false}, \emptyset \rangle$

881 A three-labeled mon becomes a two-labeled mon that is immediately applied to the  
882 client label. All other monitor reductions are defined only on the two-labeled form.

883  
884  $\mapsto \langle \text{let } f = (\text{mon}_{\text{lib}}^{\text{lib}} (\text{tr } (\lambda y. \text{true} \rightarrow y) \text{ sorted?}) (\lambda x. \text{---})) \cdot \text{main in } f \text{ false}, \emptyset \rangle$

885 MON-TRACE allocates a fresh queue for the trace and constructs a collector to give  
886 to the body-contract constructor.

887  
888  $\mapsto \langle \text{let } f = (\text{mon}_{\text{lib}}^{\text{lib}} ((\lambda y. \text{true} \rightarrow y) (\text{co } \alpha_0 \text{ sorted?})) (\lambda x. \text{---})) \cdot \text{main in}$   
889  $f \text{ false}, [\alpha_0 \mapsto \text{null}] \rangle$

890 In this step, the first argument to the trace contract produces the body contract—  
891 filling in the appropriate spot with the collector.

892  
893  $\mapsto \langle \text{let } f = (\text{mon}_{\text{lib}}^{\text{lib}} (\text{true} \rightarrow (\text{co } \alpha_0 \text{ sorted?})) (\lambda x. \text{---})) \cdot \text{main in}$   
894  $f \text{ false}, [\alpha_0 \mapsto \text{null}] \rangle$

895 The monitor contains a function contract, so the first-order check succeeds and  
896 produces a guarded value by MON-FUN.

897  
898  $\mapsto \langle \text{let } f = (\text{grd}_{\text{lib}}^{\text{lib}} (\text{true} \rightarrow (\text{co } \alpha_0 \text{ sorted?})) (\lambda x. \text{---})) \cdot \text{main in}$   
899  $f \text{ false}, [\alpha_0 \mapsto \text{null}] \rangle$

900 After several let-based steps, the elided function is applied to `false`.

901  
902  $\mapsto^+ \langle (\text{mon}_{\text{lib}}^{\text{lib}} (\text{co } \alpha_0 \text{ sorted?}) ((\lambda x. \text{---}) \text{ false})) \cdot \text{main}, [\alpha_0 \mapsto \text{null}] \rangle$

903 Assume that the elided function produces 42.

904  
905  $\mapsto^+ \langle (\text{mon}_{\text{lib}}^{\text{lib}} (\text{co } \alpha_0 \text{ sorted?}) 42) \cdot \text{main}, [\alpha_0 \mapsto \text{null}] \rangle$

906 MON-COL appends the newly received value, 42, to the trace. It then arranges for  
907 the trace predicate to be checked.

908  
909  $\mapsto^+ \langle (\text{mon}_{\text{lib}}^{\text{lib}} (\text{sorted? } \alpha_0) 42) \cdot \text{main}, [\alpha_0 \mapsto \text{cons } 42 \ \alpha_1, \alpha_1 \mapsto \text{null}] \rangle$

910 Since the singleton queue containing just 42 is sorted, the predicate succeeds.

911  
912  $\mapsto^+ \langle (\text{mon}_{\text{lib}}^{\text{lib}} \text{true } 42) \cdot \text{main}, [\alpha_0 \mapsto \text{cons } 42 \ \alpha_1, \alpha_1 \mapsto \text{null}] \rangle$

913 The result is just the return value of the function.

914  
915  $\mapsto^+ \langle 42, [\alpha_0 \mapsto \text{cons } 42 \ \alpha_1, \alpha_1 \mapsto \text{null}] \rangle$

916  
917  
918 <sup>8</sup> This model's syntax does not support trace variable declarations, so the `natural?` constraint from section 2.1  
919 is missing. Section 4.5 demonstrates how to add this feature to the model.

$\Lambda_U$ Surface Syntax	$\Lambda_U$ Evaluation Syntax
extends $\Lambda_C$	extends $\Lambda_C$
$e \in \text{Expr} = \dots \mid \text{tr } e e e$	$e \in \text{Expr} = \dots \mid \text{co } v \alpha v$
	$\kappa \in \text{Con} = \dots \mid \text{tr } v v v \mid \text{co } v \alpha v$
	$E \in \text{Ctx} = \dots \mid \text{tr } E e e \mid \text{tr } v E e$
	$\mid \text{tr } v v E$

Fig. 12. Surface and Evaluation Syntax of  $\Lambda_U$ 

$\Lambda_U$ Reduction Relation	extends $\Lambda_C$
$\langle E[\text{mon}_j^k(\text{tr } v_\kappa v_b v_p) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^k(v_b(\text{co } v_\kappa \alpha v_p)) v], \sigma[\alpha \mapsto \text{null}] \rangle$	MON-TRACE
	if $\alpha = \text{next}(\sigma)$
$\langle E[\text{mon}_j^k(\text{co } v_\kappa \alpha v_p) v], \sigma \rangle \mapsto \langle E[\text{let } x_v = \text{mon}_j^k v_\kappa v \text{ in}$	MON-COL
	$\text{let } x_j = x_v \cdot j \text{ in}$
	$\text{add! } \alpha x_j ; \text{mon}_j^k(v_p \alpha) v ; x_v], \sigma \rangle$

Fig. 13. Reduction Relation of  $\Lambda_U$ 

#### 4.5 Extending the model

While the Racket implementation pairs each trace variable with a contract that governs collected values, the model omits this capability. To illustrate the versatility of the model, this subsection shows how to add this feature. To do so is relatively simple: one tweak to the syntax and another to MON-COL. Other adaptations to the model—making it more faithful to the implementation—are similarly straightforward.

The revised surface syntax, shown in figure 12, adds contracts to the body-contract constructor; an analogous change augments collectors with contracts to protect collected values. Figure 13 shows the modified reduction relation. The MON-TRACE rule is just adapted for the new argument, while the revised MON-COL reduction has some new behavior. In the contractum, a `let` expression binds  $x_v$  to the collected value  $v$ , monitored with contract  $v_\kappa$ . The second binding, for  $x_j$ , applies the monitored value  $x_v$  to  $j$  because the consumer of the trace is the *contract-defining party*. At this point, the value is added to the trace, and the trace is tested with the predicate. If the predicate succeeds, the monitored value  $x_v$  becomes the result of the `let` expression.

This variant of MON-COL demands careful construction. First, it requires the proper management of blame parties. Monitoring the to-be-collected value is the responsibility of the contract-defining party, but using the value remains the responsibility of the client, which is the context. Second, the right-hand side may not duplicate the monitoring expression because a contract may have effects—after all, it could be another collector. So, like GRD-FUN, this rule is arranged such that the effects of  $v_\kappa$  are performed only once.

## 5 Semantic properties

Here is an evaluation function that can be used for all the languages defined in section 4:

$$\text{eval}_{\mathcal{L}} : \text{Prog} \rightarrow \text{Ans}$$

$$\text{eval}_{\mathcal{L}}(e) = \begin{cases} b & \text{if } \langle e, \emptyset \rangle \mapsto^* \langle b, \sigma \rangle \\ \text{opaque} & \text{if } \langle e, \emptyset \rangle \mapsto^* \langle v, \sigma \rangle, v \notin \text{Bool} \\ \text{err}_j^k & \text{if } \langle e, \emptyset \rangle \mapsto^* \langle \text{err}_j^k, \sigma \rangle \end{cases}$$

The  $\text{eval}_{\mathcal{L}}$  function takes *programs* as input. A program is a closed surface expression. If the reduction relation connects the program to a Boolean, then  $\text{eval}_{\mathcal{L}}$  produces the same Boolean. If the reduction relation connects the program to any other value, then  $\text{eval}_{\mathcal{L}}$  produces `opaque`, just like the REPL does for a  $\lambda$  expression. Finally, the evaluator produces an error token with two labels when the reduction relation does too.

The  $\text{eval}_{\mathcal{L}}$  relation is a partial function. Therefore, a deterministic interpreter can be defined.

**Theorem 5.1 (Functional Evaluator).**  $\text{eval}_{\mathcal{L}}$  is a partial function.

*Proof.* See appendix B. □

Moreover, the only time  $\text{eval}_{\mathcal{L}}$  is undefined is when it diverges.

**Theorem 5.2 (Uniform Evaluator).** Either  $\text{eval}_{\mathcal{L}}(e)$  is defined or the reduction sequence starting with  $\langle e, \emptyset \rangle$  is unbounded.

*Proof.* See appendix C. □

Finally, the revised contract semantics is equivalent to the original model in the absence of mutations.

**Definition (Mutation Free).** An expression  $e$  is *mutation free* if for all  $e'$  such that  $\langle e, \emptyset \rangle \mapsto^* \langle e', \sigma \rangle$  it must be that  $\sigma = \emptyset$ .

**Theorem 5.3 (Evaluator Equivalence).** If  $e$  is mutation free, then  $\text{eval}_{\Lambda_B}(e) = \text{eval}_{\Lambda_C}(e)$ .

*Proof.* See appendix D. □

## 6 Implementation in principle

The semantics of section 4 suggests a macro-style compilation of trace contracts into a mix of plain contracts and queue manipulations. Such a translation requires the timely initialization of traces, strict control of effects (i.e., queue manipulation), the injection of run-time checks, and proper blame assignment. Compiler correctness follows from a theorem like the one Findler and Felleisen (2002) prove for plain contracts.

### 6.1 Theoretical compiler

Consider the following compiler that translates a  $\Lambda_T$  program into a  $\Lambda_C$  program:

$$\mathcal{C}(\text{tr } e_b e_p) = \begin{cases} \text{let } x_b = \mathcal{C}(e_b) \text{ in} \\ \text{let } x_p = \mathcal{C}(e_p) \text{ in} \\ \lambda\_.\text{let } x_\alpha = \text{queue in} \\ \quad x_b (\lambda y.x_p (\text{add! } x_\alpha y)) \end{cases}$$

Since there is only one construct related to trace contracts in the surface syntax,  $\mathcal{C}$  has only one interesting case and is otherwise a homomorphism.

For a trace contract, the compiler sets up two bindings in a `let` expression:  $x_b$  and  $x_p$ . These stand for the compilations of the body-contract constructor and the trace predicate, respectively. The body of the `let` expression is a flat contract. Like `MON-TRACE`, it creates a fresh queue, and then an instance of the body contract by applying  $x_b$  to (the compilation of) a collector. The flat contract is used as a mechanism to initialize the queue at attachment time. Similarly, the compilation of the collector yields a flat contract that simulates `MON-COL`. Specifically, it adds the given element to the queue and then passes the extended queue to the trace predicate.

Here is the compilation of program (4.2):

$$\begin{aligned} & \text{let } x_b = \lambda y.\text{true} \rightarrow y \text{ in} \\ & \text{let } x_p = \text{sorted? in} \\ & \lambda\_.\text{let } x_\alpha = \text{queue in} \\ & \quad x_b (\lambda y.x_p (\text{add! } x_\alpha y)) \end{aligned} \tag{6.1}$$

### 6.2 Compiler correctness

Compare the reduction sequences for program (4.2) with that of program (6.1):

$$\begin{aligned} & \langle \text{let } f = \text{mon}_{\text{lib}}^{\text{lib,main}} (\text{let } x_b = \lambda y.\text{true} \rightarrow y \text{ in} \\ & \quad \text{let } x_p = \text{sorted? in} \\ & \quad \lambda\_.\text{let } x_b = \text{queue in} \\ & \quad \quad x_\kappa (\lambda y.x_p (\text{add! } x_\alpha y))) \\ & (\lambda x. \text{---}) \text{ in } f \text{ false, } \emptyset \rangle \end{aligned}$$

Following left-to-right evaluation, the compilation uses a sequence of `let` expressions to evaluate the arguments of the trace contract.

$$\begin{aligned} \mapsto^+ & \langle \text{let } f = \text{mon}_{\text{lib}}^{\text{lib,main}} (\lambda\_.\text{let } x_\alpha = \text{queue in} \\ & \quad (\lambda y.\text{true} \rightarrow y) (\lambda y.\text{sorted?} (\text{add! } x_\alpha y))) \\ & (\lambda x. \text{---}) \text{ in } f \text{ false, } \emptyset \rangle \end{aligned}$$

The three-labeled `mon` becomes a two-labeled `mon` applied to the client label.

$$\begin{aligned} \mapsto & \langle \text{let } f = (\text{mon}_{\text{lib}}^{\text{lib}} (\lambda\_.\text{let } x_\alpha = \text{queue in} \\ & \quad (\lambda y.\text{true} \rightarrow y) (\lambda y.\text{sorted?} (\text{add! } x_\alpha y))) \\ & (\lambda x. \text{---})) \cdot \text{main in } f \text{ false, } \emptyset \rangle \end{aligned}$$

The flat contract constructs a new queue and then produces an application of the body-contract constructor to the compiled collector.

$$\mapsto^+ \langle \text{let } f = (\text{mon}_{\text{lib}}^{\text{lib}} ((\lambda y. \text{true} \rightarrow y) (\lambda y. \text{sorted?} (\text{add! } \alpha_0 y))) \\ (\lambda x. \text{---})) \cdot \text{main in } f \text{ false}, [\alpha_0 \mapsto \text{null}] \rangle$$

Substituting gives a function contract with the compiled collector in the codomain position.

$$\mapsto \langle \text{let } f = (\text{mon}_{\text{lib}}^{\text{lib}} (\text{true} \rightarrow (\lambda y. \text{sorted?} (\text{add! } \alpha_0 y))) \\ (\lambda x. \text{---})) \cdot \text{main in } f \text{ false}, [\alpha_0 \mapsto \text{null}] \rangle$$

After a few steps, the elided function produces 42 by assumption. This must be checked against the compiled collector.

$$\mapsto^+ \langle (\text{mon}_{\text{lib}}^{\text{lib}} (\lambda y. \text{sorted?} (\text{add! } \alpha_0 y)) 42) \cdot \text{main}, [\alpha_0 \mapsto \text{null}] \rangle$$

The compiled collector adds the given value to the associated trace.

$$\mapsto \langle (\text{mon}_{\text{lib}}^{\text{lib}} (\text{sorted? } \alpha_0) 42) \cdot \text{main}, [\alpha_0 \mapsto \text{cons } 42 \alpha_1, \alpha_1 \mapsto \text{null}] \rangle$$

Finally, the trace predicate is run to ensure that the trace is sorted. Since it is, the final value is the result of the function: 42.

$$\mapsto \langle 42, [\alpha_0 \mapsto \text{cons } 42 \alpha_1, \alpha_1 \mapsto \text{null}] \rangle$$

This comparison suggests a proof that the compiled trace contract simulates the original behavior. Indeed, evaluating the compiled code always yields the same answer as the uncompiled source code, including divergence and errors.

**Theorem 6.1 (Compiler Correctness).**  $\text{eval}_{\Lambda_T} = \text{eval}_{\Lambda_C} \circ \mathcal{C}$

*Proof.* See appendix E. □

## 7 Implementation in practice

A principled design (section 4) specifies when traces are initialized, when they are updated, and when a predicate evaluates their validity. The design gives rise to a principled implementation (section 6), which clarifies how to translate key features into a kernel language. But, developers do not live by principles alone; pragmatics matter just as much.

One pragmatic concern is contract blame. Contracts help enforce basic correctness claims, and contract failures alert developers to problems. Finder and Felleisen (2002) insist on precise blame assignment in failure messages. The design of the trace contract system carefully reuses the blame assignment mechanism from the underlying contract system. Experience suggests that for trace contracts, developers may need additional information beyond what standard blame provides (section 7.1).

Another concern is the availability of contract combinators. Working with the trace contract system pointed to limitations in the existing behavioral contract system. In particular, additional combinators are needed to support the specification of interception points



relevant to trace contracts. Fortunately, these pragmatically important combinators are orthogonal additions to the base system (section 7.2).

Finally, an implementation effort also informs designers of what is needed in a target host language to add a new feature. While the use of Racket’s macro system greatly facilitates the addition of macro-expressible features, it should not be much more effort to extend existing compilers directly with support for trace contracts, provided the target language supports certain features (section 7.3).

### 7.1 *Blame and suspects*

When a contract system discovers a contract violation, it raises an exception, including a witness value and a pointer to the responsible component. This is dubbed *blame assignment*. Section 2.1 illustrates this point with an example of a violated trace contract.

As Lazarek et al. (2020) show in the context of behavioral contracts, blame assignment comes with enough information to almost always locate the actual source of the bug. They simulate tens of thousands of buggy programs by introducing a targeted fault via mutation. In most cases, following blame assignment leads to the source of the bug. For the few hundred cases where blame fails to identify the bug, Lazarek et al. (2020) reduce the failure to a lack of multi-call contracts. One of their examples is the DUNGEON program. As section 8 explains, strengthening the behavioral contract to a trace contract for DUNGEON provides exactly the needed blame information.

Trace contracts also complicate the situation, however. By default, blame goes to the party that added a value to the trace just before the predicate fails. Since all prefixes of the trace satisfied the predicate, this blame assignment seems to make sense. Yet, debugging real scenarios suggests that neither the blame correctness property (Dimoulas and Felleisen, 2011) nor the complete monitoring property (Dimoulas et al., 2012) are as useful for trace contracts as they are for behavioral ones.

Imagine a scenario with five components (A, B, C, D, E), where each contributes a number to a trace in increasing order ( $\leq$ ). Here is an execution:

Component	A	B	C	E	D
Contribution	1.41	2.71	3.14	5.00	4.67

The model blames D because it contributes 4.67, causing the  $\leq$  relation to fail. But, E might have made a call to the API out of order, and blaming just D does not even indicate a suspicion that some other component could be at fault. It is often useful to know the source of *all* values in a trace. After all, the idea behind traces is to subject multi-function interactions to contractual obligations.

A careful reader may argue that the problem is not with the blame assignment system, but with the predicate. Perhaps  $\leq$  does not capture the specification to a sufficient degree. This claim is already true about behavioral contracts because a predicate may always be weaker than the intended property. And if the predicate is weaker than the intended property, the contract system may blame the wrong party.

This argument, however, overlooks the key premise of contract-system design: blame assignment must help developers narrow the search space for bugs, *regardless of the*

1151 *strength of the predicate*. To explain this idea rigorously, Lazarek et al. (2020) turn folk  
 1152 wisdom into two properties: *blame trail* and *search progress*. The blame trail property  
 1153 states that either (i) blame is assigned to the buggy component or (ii) blame can be shifted  
 1154 to another component by strengthening contracts. The search progress property states that  
 1155 blame shifting always points to a component closer to the bug than before the modification.

1156 For trace contracts, both properties can be violated in practice. In the example, strength-  
 1157 ening contracts on D is unlikely to shift the blame, meaning the blame trail property is  
 1158 violated. When strengthening a trace predicate, the violating trace may decrease in length,  
 1159 but there is no reason to think *a priori* that the last contributor to a trace is always closest  
 1160 to the source of a bug, violating the search progress property. In short, the current blame  
 1161 assignment scheme points to the broken contract, but more information is needed to help  
 1162 developers identify the fault.

1163 To address this problem, the implementation comes with three different ways of express-  
 1164 ing blame assignments. Let a *suspect* be any party that contributes to a trace. Here are the  
 1165 three mechanisms used to express blame:

- 1166 1. By default, the `trace/c` implementation does not report suspects. Instead, the error  
 1167 message merely mentions the violated contract and its parties.
- 1168 2. The `setof-suspect` option forces the trace-contract system to track the set of all  
 1169 suspects and report that information when assigning blame. Frequently, there are just  
 1170 two parties to a contract. Without `#:global`, a two-party contract has a suspect set  
 1171 with at most two elements.
- 1172 3. The `listof-suspect` option causes the trace-contract system to report the exact  
 1173 sequence of suspects, one per value in the trace. This option supplies the most com-  
 1174 prehensive information, but it requires a large amount of memory and makes for  
 1175 large error messages.

1176  
 1177 Whether all of these strategies are useful in practice, only some of them, or some in certain  
 1178 circumstances and some in other circumstances, is left as an open research question.

## 1180 7.2 Supporting functionality

1181  
 1182 The trace contract library comes with additional functions for manipulating interception  
 1183 points, resetting state explicitly, transforming collectors, and augmenting error messages  
 1184 with additional information.

1185 Unlike behavioral contracts, trace contracts occasionally need to note events even in the  
 1186 absence of an informative value flow. For example, when a function receives no arguments,  
 1187 there is no natural interception point. The trace contract library supplies some combinators  
 1188 to create interception points for such situations (e.g., `apply/c`, `return/c`). See section 3.1  
 1189 for sample uses.

1190 Collector transformers wrap a collector and compute the value to be added to a trace  
 1191 from the given one. An example is `list/t`, which allows a programmer to tag values  
 1192 before they go into a trace. Typically, this tag adds information about the interception point.  
 1193 See section 3.2 for an example. Another one is `map/t`, which applies a given function to  
 1194 the captured value before adding it to a trace.

In practical situations, the `fail` function may have to perform more tasks than just inform the contract system of a failure. A software system may have to recover from a contract failure, and in those cases, a failure should reset accumulators to certain values. The author of a trace contract may also wish to add information about the rationale behind a failure. To this end, the trace-contract system supports augmenting error messages.

### 7.3 Implementing trace contracts in general

While the implementation is based on Racket’s contract system, the design is language independent. An implementor of another programming language may thus wonder what it takes to add trace contracts. Our experience suggests a few criteria.

A trace is a data structure representing the sequence of values collected from various interception points. In the context of a functional language, function calls and returns are obvious interception points. Similarly, in an object-oriented language, this same role is played by methods. Generally speaking, an implementor’s first business is to decide where to intercept and how to monitor the flow of values. The rest of this section assumes that call-and-return points suffice.

#### 7.3.1 Monitoring higher-order values

In a higher-order language, functions, objects, modules, and classes may be first-class values. This implies that a contract system cannot determine statically where a particular call or return takes place. It is the task of the target language’s runtime to support the monitoring of value flows. The Racket implementation employs proxy values (Strickland et al., 2012)—invisible wrappers—for interception. With such wrappers, it is straightforward to intercept values even in the presence of higher-order values.

Wrappers are not the only option. For instance, the weaving mechanism from aspect-oriented programming (Kiczales et al., 1997) could be used for a similar purpose. Roughly speaking, weaving injects code into the program at specifiable program points. Although weaving is powerful, it is not clear whether weaving can efficiently intercept values in a higher-order language, as needed by the proposed design.

#### 7.3.2 Mutation within contracts

Trace-contract checking is effectful. When a collector receives a value, it mutably adds this value to a trace. Even though, as some of the examples in section 2 show, the component itself can be purely functional. Hence, the underlying language must allow side effects in contracts, even though trace predicates themselves are pure functions.<sup>9</sup>

Formally, section 6 validates that trace contracts are expressible as shorthand in an underlying language with higher-order contracts and a mutable data structure. In the terminology of Felleisen (1991), the new feature is macro expressible. Theorem 6.1 shows that this translation completely preserves the specified behavior. Though, Felleisen (1991) also

<sup>9</sup> Since collectors mutate traces, checking a collector is not idempotent. While idempotence is sometimes considered an important property of contract systems (Degen et al., 2009; Findler and Blume, 2006), it often fails to hold for other reasons. For example, Owens (2012) and Hinze et al. (2006) observe violations of idempotence in several useful contexts.

shows that imperative assignment increases the expressive power of a pure host language. By implication, trace contracts are *not* expressible in such a setting.

### 7.3.3 Interception and crossing times

As mentioned in section 4.3, a trace-contract system assumes that crossing and interception time in the target contract system are separate. As it turns out, the implementation of trace contracts exposed the lack of this separation in Racket’s contract system. Racket fails to separate the two points in one combinator: the depended-upon argument contract in `->i` (Dimoulas et al., 2013). A change to Racket’s contract system allows trace contracts to distinguish these boundary crossings, meaning that a collector may ignore arguments passing through a boundary that has an indy (third) party.<sup>10</sup> This is sufficient to eliminate the duplicate-collection problem.

### 7.3.4 Macros not needed

An implementor can easily add trace contracts to a language with a rich macro system, such as a Racket. Including all the practical features mentioned in section 2 makes this macro rather large and complex. While macros are a convenient implementation mechanism for trace contracts, they are not a requirement. The implementor of a functional language such as SML, which elaborates surface syntax into a small kernel, can add trace contracts with a similar addition to the front-end elaborator.

## 8 Usability and performance evaluation

Usability questions concern the ease with which programmers can write trace-contract properties for their programs and what performance penalty the system imposes.

Section 8.2 gives a qualitative assessment of our experience writing trace contracts. This assessment suggests two opposite insights. On the one hand, trace contracts enable developers to use the entire underlying programming language. Hence, developing a trace-contract property is just like developing an ordinary predicate in an ordinary language, using all available tools—especially unit and property-testing frameworks. On the other hand, as experience with ordinary higher-order contracts shows, contracts are a special-purpose domain. Such domains call for specific, tailor-made notations to eliminate boilerplate code. Developing such notations remains future work.

As for performance, the only relevant question is what kind of *fixed cost* the mechanism itself imposes on programs, not the *variable cost* of the programmer-defined predicates.<sup>11</sup> Trace initialization, trace updates, and calls to predicates are all included in this fixed cost. The results of measuring the performance of trace contracts, presented in section 8.3, are quite encouraging.

<sup>10</sup> Thanks to Robby Findler for help with this change to Racket’s contract system.

<sup>11</sup> The performance evaluation cannot answer questions concerning the *variable cost* of trace predicates. Trace contracts are *property agnostic*, so the variable cost of a trace contract depends largely on the property being checked. In other words, this cost is solely under the purview of the programmer, not the trace-contract system.

### 8.1 Benchmark programs

The selected benchmarks represent real-world uses of Racket that offer opportunities for adding trace contracts. `MEMORY` turns the example from section 2.1 into a pathological stress test. `FUTURE` is a large existing Racket library equipped with trace contracts, plus an application that stresses the functionality. Four of the benchmark programs (`DUNGEON`, `JPEG`, `LNМ`, `TETRIS`) are variants on programs from the standard gradual typing benchmark suite (Greenman et al., 2019). Three (`DATAFLOW`, `FISH`, `TICKET`) are programs developed for use in university courses. All the benchmarks have been adapted so that they do not measure I/O operations.

**DATAFLOW** Computes a constant propagation analysis for a simple imperative language. A trace contract, similar to the one from section 2.3, checks the monotonicity of a transfer function during fixed-point iteration.

**DUNGEON** Generates the specification of a maze. A trace contract on the random-number generator ensures that it does not exhaust a fixed pool of random numbers. In the original program, resizing the random number pool caused a contract violation that failed to provide helpful blame information (Lazarek et al., 2020, sec. 5.1). With a trace contract, this same bug produces an error message with a blame assignment that directly points to the problem. The contract must keep track of how many times the random function is called, so its accumulator is just a natural number and the check is cheap.

**FISH** Runs a “That’s My Fish” board game tournament. There are two trace contracts: a referee contract and a player contract.

The referee contract ensures that the referee calls back players in the specified order unless the game state does not permit the player to take a turn. The contract is a promise made by the referee to all the players. To enforce this promise, the contract is placed on the referee’s list of player objects. A collector receives a new value every time the referee calls the `take-turn` method on any player. The trace contract then checks that this is in accordance with the promised callback order on the *players*, including skipping over players that are momentarily prohibited from taking a turn. The player contract enforces a sequence property on its method calls. In other words, the player components ensure that their individual *methods* are called in the specified order. This contract is similar to the value-dependent temporal protocol example from section 3.2. It is independent of, and orthogonal to, the referee contract.

**FUTURE** Visualizes the performance of a futures benchmark. Futures are a run-time mechanism for incrementally adding parallelism to programs (Swaine et al., 2010). The future visualizer (Swaine et al., 2012) uses Racket’s drawing library, which has been equipped with trace contracts to enforce multi-call properties. A full list of these properties is enumerated in appendix F. Some of the properties were monitored by the drawing library using ad-hoc checks and others were not checked at all.

**JPEG** Parses a JPEG input stream and writes it to an output stream. A trace contract guarantees that operations on the output stream occur in the correct order. Like the example in section 3.2, it checks every stream-related function call against a finite

Benchmark	SLOC	Protects	Checks	Disabled	Enabled	Predicate	Overhead
DATAFLOW	502	1	584	83 ± 3	87 ± 2	274 ± 3	5%
DUNGEON	589	0	538,000	2441 ± 38	2715 ± 46	2713 ± 33	11%
FISH	1,452	2,698	63,175	7780 ± 70	8340 ± 82	8366 ± 80	7%
FUTURE	1,721	16,360	234,444	6075 ± 54	7083 ± 83	7502 ± 86	17%
JPEG	1,481	0	54,556	276 ± 5	303 ± 6	316 ± 6	10%
LNM	564	168	3,248	522 ± 8	532 ± 9	534 ± 9	2%
MEMORY	59	0	10,000	141 ± 4	164 ± 4	164 ± 4	16%
TETRIS	334	6,807	125,570	3040 ± 24	3566 ± 36	3927 ± 43	17%
TICKET	1,427	384	15,794	13062 ± 149	13186 ± 170	13199 ± 182	1%

Table 4. Basic Metrics and Performance Measurements

automaton. Formulating the trace contract involves creating several contracts that share the same accumulator, the state of the finite automaton.

**LNM** Draws plots of the performance measurements of a gradual type system. Like **FUTURE**, this benchmark uses a variant of Racket’s drawing library equipped with trace contracts.

**MEMORY** Reports memory use, including garbage-collected blocks. The trace contract from section 2.1 ensures that `current-memory-use` returns increasing numbers over time; it is called 10,000 times in a tight loop, the results of which are graphed on a line chart using Racket’s `plot` (Toronto and Harsányi, 2011) library.

**TETRIS** Simulates and displays a recording of the game of Tetris. This benchmark also uses a variant of Racket’s drawing library equipped with trace contracts.

**TICKET** Runs a “Ticket to Ride” board game tournament. Like **FISH**, **TICKET** has both a referee and a player contract. The referee contract enforces a promise that the referee calls back players in the specified order. This trace contract is significantly simpler than the one for **FISH**, because every player can execute an action in every game state. The player-side trace contract enforces the correct sequence of method calls. The example presented in section 3.2 is a simplified version of this contract.

## 8.2 Benchmark summary

Table 4 first lists the number of essential lines of source code (SLOC) for each program, including the trace contract and its auxiliary functions.

None of the trace contracts require much code. **FISH** and **TICKET** contain the most complex ones, but the others are relatively simple. Even the most complex trace contracts are concise. Indeed, the contract for **TICKET** is shown nearly verbatim in section 3.2. Since predicates are ordinary code, they can make use of existing data structure libraries, and those libraries serve as workhorses in many cases. For example, **JPEG** uses an existing FSM package that renders its temporal constraint predicate practically a one-liner.

Tight integration with the existing contract system makes writing many trace contracts natural. Since the trace contract mechanism manages state behind the scenes, contract composition and contract abstraction work as expected. Developers can write trace contracts as ordinary code, compose them as usual, and even abstract over them.

1381 Programming trace contracts for these benchmark programs also points to limitations.  
1382 For example, placing collector contracts can be awkward and repetitive. Consider the trace  
1383 contracts in sections 3.1 and 3.2, both of which contain several nearly identical lines. A  
1384 macro can eliminate the repetition in each case individually, but it is not obvious if there is  
1385 a general-purpose DSL that could reduce such repetitive code across many cases.  
1386

### 1387 8.3 Performance measurements

1388 The performance measurements on the right side of table 4 were recorded on a dedicated  
1389 Linux machine with an Intel Xeon E3 processor running at 3.10 GHz with 32 GB of RAM  
1390 and with Racket 8.6 CS. Each benchmark configuration was repeated 100 times with a  
1391 maximum timeout of two minutes.  
1392

1393 The Protects column reports the number of times a trace contract protects a new value  
1394 during the steady state of a program's execution. Each time, there is some overhead due to  
1395 allocating references for accumulators and creating collector contracts. Some benchmarks  
1396 have a zero entry because all of the trace contracts are initialized before the main body of  
1397 the program begins, for example, when dependencies are being loaded.

1398 The Checks column states the number of times each trace predicate is checked. As men-  
1399 tioned, this evaluation is concerned with the *fixed cost* of trace contracts. Therefore, each  
1400 trace predicate is replaced with the trivial predicate that always returns true. Benchmarks  
1401 were executed at two levels: Disabled where trace contracts are disabled, and Enabled  
1402 where they are enabled. These measurements are the mean number of milliseconds it takes  
1403 to run each benchmark, averaged over 100 samples, along with the standard deviation. The  
1404 Predicate column lists the performance numbers where trace contracts are enabled and the  
1405 predicate actually checks the desired property. Despite it not being the primary means of  
1406 evaluation, these numbers are provided for context. Such predicates are straightforward  
1407 implementations and are not heavily optimized. Finally, the Overhead column shows the  
1408 percent overhead of Enabled compared to Disabled.

1409 The overhead of the trace-contract mechanism is relatively low, somewhere between 1%  
1410 and 17%. As is, the setups basically simulate worst-case scenarios. For example, MEMORY  
1411 just calls a simple function in a tight loop, so contract checking takes up a large portion  
1412 of total execution time. By contrast, benchmarks that are closer to real-world programs,  
1413 such as TICKET, incur a low overhead. Thus, the evidence suggests that the trace-contract  
1414 mechanism itself does not exhibit any performance pathology.

1415 These measurements do not exercise an industrial-strength implementation of trace con-  
1416 tracts, but rather a direct translation of the design. This implementation serves as a vehicle  
1417 for exploration. With some performance engineering, it is likely to perform significantly  
1418 better. While this evaluation can provide some first impression of the performance of trace  
1419 contracts, it is not enough to generalize to other settings or languages.  
1420  
1421  
1422  
1423  
1424  
1425  
1426



## 9 Related work

Prior work is in the tradition of software contracts or runtime verification (RV). Specifically, this paper leverages the development of higher-order dependent contracts (Findler and Felleisen, 2002; Blume and McAllester, 2006; Findler and Blume, 2006; Greenberg et al., 2010; Dimoulas et al., 2012); the temporal contract system of Disney et al. (2011) is the most directly comparable piece of work from this area. Within the runtime verification area, the most similar approach is the monitor-oriented programming framework (Meredith et al., 2011; Chen and Roşu, 2007; Chen et al., 2005).

These two bodies of research have distinct philosophies about expressing and checking properties. Trace contracts borrow the notion of traces from RV to extend a higher-order behavioral contract system. They seek to bridge the gap between the two areas. Eventually this bridge should make many results from RV available to contract programmers, and it may inject new ideas into RV.

### 9.1 Runtime verification, generally

Traditional contract systems and RV systems differ along several dimensions. Most importantly, as Meyer (1992) observes, contracts are a design tool for the developer; in contrast, RV is a tool for the quality assurance stage of the development process.

#### 9.1.1 Scope

Contracts are *modular*. A programmer attaches contracts to the interface of a “server” component. When a “client” component imports a server component, it is forced to agree to the contract. Similarly, a client component may impose a contract on imported pieces of functionality to protect itself from a misbehaving service component. In the first case, clients do not need to be adapted to the service contract, and in the second case, service components remain unaware of the client’s protective contract. Put differently, it is possible to compile these components in either order or, even better, to link pre-compiled binary objects.

RV is *whole program*. A programmer specifies events of interest and properties about event traces. The RV system converts this specification into an executable monitor and weaves interception code into the host program to communicate first-order data about events to a separate monitor process (Bartocci et al., 2018).

Monitoring higher-order values is possible with RV, but the encoding uses a complex protocol between the server and the client module; it requires source modification to both components. Implementing the protocol on a modular basis is either impossible, which precludes the binary-linking approach available with contracts, or requires complex extensions (Xiang et al., 2015).

#### 9.1.2 Language

Contracts are linguistic elements that are *inside* the language. The programmer uses the same language—and the exact same tools—for writing code and contracts. Extending the notation for contracts in a domain-specific manner (via macros in Racket) is useful; the



1473 -> abbreviation for function contracts is one example. Racket also treats contracts as first-  
1474 class objects, meaning they can be put into lists, passed and returned from functions, and  
1475 composed at run time.

1476 RV is extra-linguistic; that is, RV systems exist *outside* the language. Specifications are  
1477 usually written in a distinct, external logic language and tend to make temporal statements  
1478 about sequences of first-order data (Havelund et al., 2018). While this language may con-  
1479 tain fragments of host-language code, it is only loosely connected with the host language  
1480 and its tool chain.

### 1481 9.1.3 Violations

1482 As a consequence of the differences along the linguistic axis, contracts and RV differ in two  
1483 ways concerning the violation of specifications: recovery and error-location information.

1484 When a contract system discovers a violation of an assertion, it raises an exception that  
1485 includes information about the parties that agreed to the contract and which of them vio-  
1486 lated it—blame information. By raising an exception at the very point where a contract  
1487 violation is discovered, the contract system gives the program a chance to recover immedi-  
1488 ately and with a response targeted to the problem. In a language with resumable exceptions,  
1489 such as Common Lisp (Steele, 1990), a program may even resume its execution at the exact  
1490 place where the violation occurred.

1491 The *precise* error information in violation messages enables the developer to understand  
1492 the cause of a violation. Lazarek et al. (2020) show that this blame information is effec-  
1493 tive at narrowing the search space during debugging. It is also a well-founded concept;  
1494 Dimoulas et al. (2012) provide a framework for proving that blame information points to  
1495 the component which supplies a value that does not meet the specification.

1496 Traditionally, RV systems report violations of specifications with delay and do not  
1497 contain blame information (Swords, 2019). The delay is due to the underlying process-  
1498 communication arrangement between the program proper and its monitor. This poses  
1499 a problem for tracking the provenance of values and for assigning blame. Hence, RV  
1500 makes it difficult to restart programs with a problem-specific, localized response, unless  
1501 an additional “diagnosis layer” is supplied (Leucker and Schallhart, 2009).

### 1502 9.1.4 Properties

1503 Contracts are property *agnostic*. Any predicate, including one that tries to decide a  
1504 recursively-enumerable property, can be used as a contract. This is maximally expressive  
1505 but can be computationally expensive.

1506 RV is property *sensitive*. Much of RV research focuses on the development of specifica-  
1507 tion languages that can express properties of interest concisely and that can be compiled  
1508 into efficient monitoring code (Leucker and Schallhart, 2009). Often these are variants of  
1509 temporal logic. These specialized logics can provide hard guarantees about time and space  
1510 efficiency, at the cost of expressive power.

1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518

### 9.2 Runtime verification, specifically

1519  
1520 Within the landscape of RV tools, JavaMOP is the best point for comparison. It is the  
1521 most versatile implementation in the family of monitor-oriented programming (MOP)  
1522 systems (Meredith et al., 2011). A selling feature of JavaMOP is that it is generic; the  
1523 programmer can choose the events of interest, specification logic, and violation handler  
1524 code. Chen and Roşu (2007) argue that there is no logic suitable to express all properties,  
1525 and thus JavaMOP developers must engineer external logic “plugins” (Chen et al., 2005).

1526 Trace contracts, by contrast, allow programmers to take full advantage of the host lan-  
1527 guage. If this host language comes with expressive meta-programming facilities, such as  
1528 the macros of Racket (Ballantyne et al., 2020; Felleisen et al., 2018; Flatt, 2002), develop-  
1529 ers can easily add a custom notation for trace contracts. Consider section 3.2 which uses  
1530 Racket’s automata package (McCarthy, 2011) and significantly improves the readability of  
1531 the trace predicate without external tooling. With the visual-interactive syntax of Andersen  
1532 et al. (2020), a developer could even edit and view the NFA graphically.

1533 For an example of cross-pollination, consider trace slicing. This idea is due to the RV  
1534 community (Chen and Roşu, 2007). In the RV world, this operation is *not* exposed to users  
1535 of RV systems; rather, an efficient slicing algorithm is derived from data quantifiers in  
1536 the specification logic. The trace contract library supports trace slicing via tagging and  
1537 ordinary stream functions. In keeping with the philosophy of contract-system design, the  
1538 power is handed to programmers.

### 9.3 Higher-order contracts, specifically

1540  
1541  
1542 While higher-order contracts are typically *independent* of state, trace contracts manage  
1543 state behind the scenes to support a mostly functional view of specifications. Others  
1544 show that contracts could occasionally benefit from a modicum of state (Tov and Pucella,  
1545 2010; Moore et al., 2016; Waye et al., 2017), though these systems do not come with the  
1546 expressiveness of trace contracts.

1547 The higher-order temporal contracts of Disney et al. (2011) are the closest prior work to  
1548 trace contracts. Their research focuses on two aspects: an operational theory of temporal  
1549 event sequences and the specification of properties. On the theory side, the work intro-  
1550 duces a novel approach to operational semantics that formalizes the meaning of modules as  
1551 automata that create trees of observable events, similar to game-based denotational seman-  
1552 tics. The semantics satisfies a non-interference theorem, meaning that streams of values  
1553 are kept separate. On the practical side, the work focuses on specifying properties of event  
1554 sequences as regular expressions *without* giving programmers access to a data represen-  
1555 tation of traces. Trace contracts come with more expressive power, yet do not necessarily  
1556 sacrifice efficiency.

1557 At first glance, computational contracts (Scholliers et al., 2015) look similar to higher-  
1558 order temporal contracts. But, computational contracts go far beyond any classical contract  
1559 classification scheme (Beugnard et al., 1999, 2010), providing unprecedented power and  
1560 imposing a similarly high cost. A computational contract system empowers programmers  
1561  
1562  
1563  
1564

to impose arbitrary restrictions on components from the outside and in a post-hoc manner. Thus, computational contracts depart from the idea that contracts are assertions at the boundary between black-box components, instead turning components into glass boxes.

#### 9.4 *Typestate and type systems*

Researchers often try to move from dynamically checked contracts to statically checked types, because discovering general mistakes during compile time is safer than discovering specific mistakes at run time, perhaps even after a program has been deployed. This subsection deals with two distantly related ideas from the world of static checking.

The work of Strom and Yemini (1986) on typestate systems, recently resumed in various forms (Jaspan and Aldrich, 2009; Pucella and Tov, 2008; Wolff et al., 2011), directly addresses simple but common affinity restrictions in APIs. For example, typestate systems can check constraints such as “method *m* may be called at most once” and even “method *m* must be called before method *n*.” These constraints are restricted to regular properties, i.e., those that can be expressed using a finite-state machine.

Honda et al. (1998)’s notion of session type is a closely related idea. Recently this field has experienced rapid growth. Roughly speaking, session types for objects come with the same expressive power as typestate (Gay et al., 2010).

Effect systems are also capable, in a limited way, of constraining the order in which effects can be performed. Ordinary effect systems do not consider the order of effects, but sequential effect systems (Tate, 2013; Koskinen and Terauchi, 2014) can. Further extensions can statically verify some temporal logic propositions (Gordon, 2017).

No existing static technique can express all of the trace-contract examples. By combining traces with plain code, a programmer can formulate arbitrary predicates and check value-dependent constraints on traces. Trace predicates can look for specific values or use specific values to express a constraint, which is impossible with these type systems. Dependent session types (Toninho et al., 2011) may be able to do better, but are still limited to statically decidable properties. Trace contracts, by monitoring programs at run time, are able to take advantage of the precision that run-time checking offers. A combination of session types and contracts (Bocchi et al., 2010) can refine the content of messages passed between parties, but the structure of the protocol remains fixed. This approach also does not naturally extend to contracts on higher-order values.

## 10 Trace contracts for rich specifications

Engineering complex software requires mechanisms for expressing and enforcing component specifications. Types, contracts, run-time verification—each has been successful in its own way, but major expressiveness gaps remain.

This paper introduces trace contracts as a novel, practical, and well-founded element of this spectrum. Specifically, trace contracts enable developers to protect the elements of their API across multiple function and method calls. The trace contract system provides traces of argument and result values as a first-class piece of data. Hence, trace contracts can express protocols that are ubiquitous in practice, but are usually specified informally.

1611 In addition to a principled design, this paper describes an implementation of trace  
1612 contracts, along with an evaluation. The implementation addresses a good number of prag-  
1613 matic concerns, especially those of performance. On the question of blame assignment,  
1614 the implementation supports several natural strategies with different precision and memory  
1615 consumption trade-offs.

1616 Critically, the trace-contract design separates the concept of a value trace from the  
1617 language of enforced properties. In other words, trace contracts separate the low-level col-  
1618 lection mechanism from the high-level property formulation. Hence, the design enables an  
1619 investigation of trace-collection performance, independent of an exploration of problem-  
1620 specific notations for expressing the properties of traces. Racket, with its powerful tools for  
1621 creating embedded and extensible DSLs (Ballantyne et al., 2020), is a convenient platform  
1622 for this kind of research.

1623 Plenty of work remains. Section 7.1 proposes three blame strategies but gives no theo-  
1624 retical or empirical justification for any of them. What are the tradeoffs between these  
1625 approaches with regard to theory (blame correctness), implementation (memory use),  
1626 and pragmatics (debugging violations)? Protocols are common in concurrent programs  
1627 but are often informally described. Can trace contracts be adapted to monitor protocols  
1628 in concurrent applications? Techniques exist to statically verify functional contracts in  
1629 Racket (Nguyễn et al., 2018). Is static verification practical for trace contracts? Section 9  
1630 compares trace contracts to other research results. How many of these systems can be  
1631 implemented on top of trace contracts? If they can, what are the benefits of doing so? If  
1632 they cannot, how can trace contracts be extended to accommodate such systems?

1633 Even though future work is needed to turn trace contracts into a truly practical technol-  
1634 ogy, hopefully the foundation put forth in this paper is sufficient to advance the practice of  
1635 software specification in Racket and beyond.

### 1637 **Acknowledgements**

1639 This work was supported by National Science Foundation grant SHF 2116372. The authors  
1640 thank anonymous POPL and JFP reviewers for their comments.

### 1643 **Artifact**

1644 The implementation of trace contracts has been released as an open-source library. Right  
1645 now, Racket developers can use trace contracts to fortify their programs.

### 1648 **Conflicts of Interest**

1649 None.

### 1653 **References**

1654 Andersen, L., Ballantyne, M. & Felleisen, M. (2020) Adding Interactive Visual Syntax to Textual  
1655

- Code. Object-Oriented Programming, Systems, Languages and Applications (OOPSLA).
- 1657 Andersen, L., St-Amour, V., Vitek, J. & Felleisen, M. (2018) Feature-Specific Profiling. *Transactions*  
1658 *on Programming Languages and Systems (TOPLAS)*.
- 1659 Ashley, J. M. & Dybvig, R. K. (1994) An Efficient Implementation of Multiple Return Values in  
1660 Scheme. *LISP and Functional Programming (LFP)*.
- 1661 Ballantyne, M., King, A. & Felleisen, M. (2020) Macros for Domain-Specific Languages. Object-  
1662 Oriented Programming, Systems, Languages and Applications (OOPSLA).
- 1663 Barendregt, H. P. (1981) *The Lambda Calculus*. North-Holland Publishing Co.
- 1664 Bartocci, E., Falcone, Y., Francalanza, A. & Regeer, G. (2018) Introduction to Runtime Verification.  
1665 In *Lectures on Runtime Verification*. Springer.
- 1666 Beugnard, A., Jézéquel, J.-M. & Plouzeau, N. (2010) Contract Aware Components, 10 Years After.  
1667 International Workshop on Component and Service Interoperability (WCSI).
- 1668 Beugnard, A., Jézéquel, J.-M., Plouzeau, N. & Watkins, D. (1999) Making Components Contract  
1669 Aware. *Computer*.
- 1670 Blume, M. & McAllester, D. (2006) Sound and Complete Models of Contracts. *Journal of Functional*  
1671 *Programming (JFP)*.
- 1672 Bocchi, L., Honda, K., Tuosto, E. & Yoshida, N. (2010) A Theory of Design-by-Contract for  
1673 Distributed Multiparty Interactions. International Conference on Concurrency Theory.
- 1674 Chen, F., d'Amorim, M. & Roşu, G. (2005) Checking and Correcting Behaviors of Java Programs at  
1675 Runtime with Java-MOP. Workshop on Runtime Verification (RV).
- 1676 Chen, F. & Roşu, G. (2007) MOP: An Efficient and Generic Runtime Verification Framework.  
1677 Object-Oriented Programming, Systems, Languages and Applications (OOPSLA).
- 1678 Degen, M., Thiemann, P. & Wehr, S. (2009) True Lies: Lazy Contracts for Lazy Languages  
1679 (Faithfulness is Better than Laziness). Arbeitstagung Programmiersprachen (ATPS).
- 1680 Dimoulas, C. & Felleisen, M. (2011) On Contract Satisfaction in a Higher-Order World. *Transactions*  
1681 *on Programming Languages and Systems (TOPLAS)*.
- 1682 Dimoulas, C., Findler, R. B. & Felleisen, M. (2013) Option Contracts. Object-Oriented  
1683 Programming, Systems, Languages and Applications (OOPSLA).
- 1684 Dimoulas, C., Findler, R. B., Flanagan, C. & Felleisen, M. (2011) Correct Blame for Contracts: No  
1685 More Scapagoating. Principles of Programming Languages (POPL).
- 1686 Dimoulas, C., New, M. S., Findler, R. B. & Felleisen, M. (2016) Oh Lord, Please Don't Let Contracts  
1687 Be Misunderstood (Functional Pearl). International Conference on Functional Programming  
1688 (ICFP).
- 1689 Dimoulas, C., Tobin-Hochstadt, S. & Felleisen, M. (2012) Complete Monitors for Behavioral  
1690 Contracts. European Symposium on Programming (ESOP).
- 1691 Disney, T., Flanagan, C. & McCarthy, J. (2011) Temporal Higher-Order Contracts. International  
1692 Conference on Functional Programming (ICFP).
- 1693 Felleisen, M. (1991) On the Expressive Power of Programming Languages. *Science of Computer*  
1694 *Programming*.
- 1695 Felleisen, M., Findler, R. B., Flatt, M., Krishnamurthi, S., Barzilay, E., McCarthy, J. & Tobin-  
1696 Hochstadt, S. (2018) A Programmable Programming Language. *Communications of the ACM*  
1697 *(CACM)*.
- 1698 Findler, R. B. & Blume, M. (2006) Contracts as Pairs of Projections. Functional and Logic  
1699 Programming (FLP).
- 1700 Findler, R. B. & Felleisen, M. (2002) Contracts for Higher-Order Functions. International  
1701 Conference on Functional Programming (ICFP).
- 1702 Flatt, M. (2002) Composable and Compilable Macros: You Want it When? International Conference  
on Functional Programming (ICFP).
- Flatt, M. & PLT. (2010) Reference: Racket. Technical Report PLT-TR-2010-1. PLT Design Inc.  
<https://racket-lang.org/tr1/>.
- Gay, S. J., Vasconcelos, V. T., Ravara, A., Gesbert, N. & Caldeira, A. Z. (2010) Modular Session  
Types for Distributed Object-Oriented Programming. Principles of Programming Languages  
(POPL).

- 1703 Gordon, C. S. (2017) A Generic Approach to Flow-Sensitive Polymorphic Effects. European  
1704 Conference on Object-Oriented Programming (ECOOP).
- 1705 Greenberg, M., Pierce, B. C. & Weirich, S. (2010) Contracts Made Manifest. Principles of  
1706 Programming Languages (POPL).
- 1707 Greenman, B., Takikawa, A., New, M. S., Feltey, D., Findler, R. B., Vitek, J. & Felleisen, M.  
1708 (2019) How to Evaluate the Performance of Gradual Typing Systems. *Journal of Functional  
1709 Programming (JFP)*.
- 1710 Havelund, K., Reger, G., Thoma, D. & Zălinescu, E. (2018) Monitoring Events that Carry Data. In  
1711 *Lectures on Runtime Verification*. Springer.
- 1712 Hinze, R., Jeuring, J. & Löh, A. (2006) Typed Contracts for Functional Programming. Functional  
1713 and Logic Programming (FLP).
- 1714 Honda, K., Vasconcelos, V. T. & Kubo, M. (1998) Language Primitives and Type Discipline for  
1715 Structured Communication-Based Programming. European Symposium on Programming (ESOP).
- 1716 Jaspán, C. & Aldrich, J. (2009) Checking Framework Interactions with Relationships. European  
1717 Conference on Object-Oriented Programming (ECOOP).
- 1718 Kiczales, G., Lamping, J., Mendhekar, A., Maeda, C., Lopes, C., Loingtier, J.-M. & Irwin, J.  
1719 (1997) Aspect-Oriented Programming. European Conference on Object-Oriented Programming  
1720 (ECOOP).
- 1721 Koskinen, E. & Terauchi, T. (2014) Local Temporal Reasoning. Logic in Computer Science (LICS).
- 1722 Lazarek, L., King, A., Sundar, S., Findler, R. B. & Dimoulas, C. (2020) Does Blame Shifting Work?  
1723 Principles of Programming Languages (POPL).
- 1724 Leucker, M. & Schallhart, C. (2009) A Brief Account of Runtime Verification. *The Journal of Logic  
1725 and Algebraic Programming*.
- 1726 McCarthy, J. (2011) Automata: Compiling State Machines. [https://docs.racket-lang.org/  
1727 automata/index.html](https://docs.racket-lang.org/automata/index.html).
- 1728 Meredith, P. O., Jin, D., Griffith, D., Chen, F. & Roşu, G. (2011) An Overview of the MOP Runtime  
1729 Verification Framework. *International Journal on Software Tools for Technology Transfer*.
- 1730 Meyer, B. (1988) *Object-Oriented Software Construction*. Prentice Hall.
- 1731 Meyer, B. (1992) Applying “Design by Contract”. *Computer*.
- 1732 Moore, S., Dimoulas, C., Findler, R. B., Flatt, M. & Chong, S. (2016) Extensible Access Control with  
1733 Authorization Contracts. Object-Oriented Programming, Systems, Languages and Applications  
1734 (OOPSLA).
- 1735 Moy, C., Nguyễn, P. C., Tobin-Hochstadt, S. & Van Horn, D. (2021) Corpse Reviver: Sound and  
1736 Efficient Gradual Typing via Contract Verification. Principles of Programming Languages (POPL).
- 1737 Nguyễn, P. C., Gilray, T., Tobin-Hochstadt, S. & Van Horn, D. (2018) Soft Contract Verification for  
1738 Higher-Order Stateful Programs. Principles of Programming Languages (POPL).
- 1739 Nielson, F., Nielson, H. R. & Hankin, C. (2005) *Principles of Program Analysis*. Springer Verlag.
- 1740 Owens, Z. (2012) Contract Monitoring as an Effect. Higher-Order Programming with Effects  
1741 (HOPE).
- 1742 Plotkin, G. (1975) Call-by-name, call-by-value and the  $\lambda$ -calculus. *Theoretical Computer Science*.
- 1743 Pucella, R. & Tov, J. A. (2008) Haskell Session Types with (Almost) No Class. Haskell Symposium.
- 1744 Scholliers, C., Tanter, E. & De Meuter, W. (2015) Computational Contracts. *Science of Computer  
1745 Programming*.
- 1746 Steele, G. L. (1990) *Common Lisp the Language*. Digital Press.
- 1747 Strickland, T. S., Tobin-Hochstadt, S., Findler, R. B. & Flatt, M. (2012) Chaperones and  
1748 Impersonators: Run-Time Support for Reasonable Interposition. Object-Oriented Programming,  
1749 Systems, Languages and Applications (OOPSLA).
- 1750 Strom, R. E. & Yemini, S. (1986) Typestate: A Programming Language Concept for Enhancing  
1751 Software Reliability. *IEEE Transactions on Software Engineering*.
- 1752 Swaine, J., Fetscher, B., St-Amour, V., Findler, R. B. & Flatt, M. (2012) Seeing the Futures: Profiling  
1753 Shared-Memory Parallel Racket. Functional High-Performance Computing (FHPC).
- 1754 Swaine, J., Tew, K., Dinda, P. A., Findler, R. B. & Flatt, M. (2010) Back to the Futures: Incremental  
1755 Parallelization of Existing Sequential Runtime Systems. Object-Oriented Programming, Systems,  
1756

Languages and Applications (OOPSLA).

1749 Swords, C. (2019) *A Unified Characterization of Runtime Verification Systems as Patterns of*  
1750 *Communication*. Ph.D. thesis. Indiana University.

1751 Tate, R. (2013) *The Sequential Semantics of Producer Effect Systems*. Principles of Programming  
1752 Languages (POPL).

1753 Toninho, B., Caires, L. & Pfenning, F. (2011) *Dependent Session Types via Intuitionistic Linear Type*  
Theory. Principles and Practice of Declarative Programming (PPDP).

1754 Toronto, N. & Harsányi, A. (2011) *Plot: Graph Plotting*. [https://docs.racket-lang.org/](https://docs.racket-lang.org/plot/index.html)  
1755 [plot/index.html](https://docs.racket-lang.org/plot/index.html).

1756 Tov, J. A. & Pucella, R. (2010) *Stateful Contracts for Affine Types*. European Symposium on  
1757 Programming (ESOP).

1758 Waye, L., Chong, S. & Dimoulas, C. (2017) *Whip: Higher-Order Contracts for Modern Services*.  
International Conference on Functional Programming (ICFP).

1759 Wolff, R., Garcia, R., Tanter, E. & Aldrich, J. (2011) *Gradual Typestate*. European Conference on  
1760 Object-Oriented Programming (ECOOP).

1761 Xiang, C., Qi, Z. & Binder, W. (2015) *Flexible and Extensible Runtime Verification for Java*.  
1762 *International Journal of Software Engineering and Knowledge Engineering*.

1763

1764

1765

1766

1767

1768

1769

1770

1771

1772

1773

1774

1775

1776

1777

1778

1779

1780

1781

1782

1783

1784

1785

1786

1787

1788

1789

1790

1791

1792

1793

1794

## A Proof syntax and judgments

The proofs in the sections that follow require some additional syntax and judgments. In particular, certain sets of expressions that exist implicitly in the semantics must be named explicitly. Additionally, a judgment identifying valid expressions is needed.

$\Lambda$  Proof Syntax

$$\begin{aligned} a \in \text{Ans} &= t \mid \text{opaque} \\ t \in \text{Ter} &= v \mid \text{err}_j^k \\ r \in \text{Redex} &= \text{if } v e e \mid v v \mid o v \mid \text{queue} \mid \text{add! } v v \mid \text{err}_j^k \end{aligned}$$

$\Lambda_B$  Proof Syntax extends  $\Lambda$

$$r \in \text{Redex} = \dots \mid \text{mon}_j^{k,l} e e$$

$\Lambda_C$  Proof Syntax extends  $\Lambda$

$$r \in \text{Redex} = \dots \mid \text{mon}_j^{k,l} e e \mid \text{mon}_j^k v v \mid (\text{grd}_j^k \omega v) \cdot l$$

Fig. 14. Proof Syntax of  $\Lambda$ ,  $\Lambda_B$ , and  $\Lambda_C$

Figure 14 defines three sets of terms. An answer is the result of  $\text{eval}_{\mathcal{L}}$  (for a language  $\mathcal{L}$ ) and is either a terminal expression or the `opaque` token. A terminal expression is either a value or an error token. Finally, a reducible expression (redex) is an expression that inhabits the hole of the evaluation context on the left-hand side of a reduction rule.

$$\begin{array}{c} \text{fv}(e) = \emptyset \\ \forall \alpha \in \text{addrs}(e) [\sigma \Vdash \sigma(\alpha)] \\ \hline \sigma \Vdash e \end{array} \qquad \frac{}{\sigma \Vdash \text{null}} \qquad \frac{\sigma \vdash v \quad \sigma \Vdash \sigma(\alpha)}{\sigma \Vdash \text{cons } v \alpha}$$

Fig. 15. Valid Expression Judgment

Figure 15 defines a judgment that identifies valid expressions from the too-liberal grammar for the evaluation syntax. A valid expression is closed and contains only addresses that map to valid queues. A valid queue contains only valid values.



## B Functional evaluator proof

The theorems in this section hold for all languages presented in section 4.

**Theorem 5.1 (Functional Evaluator).**  $\text{eval}_{\mathcal{L}}$  is a partial function.

*Proof.* A straightforward consequence of lemma B.1.  $\square$

**Lemma B.1 (Deterministic Evaluator).** If  $\langle e, \sigma \rangle \mapsto^* \langle t_1, \sigma_1 \rangle$  and  $\langle e, \sigma \rangle \mapsto^* \langle t_2, \sigma_2 \rangle$ , then  $t_1 = t_2$  and  $\sigma_1 = \sigma_2$ .

*Proof.* By lemma B.2, every expression can be decomposed into a unique evaluation context and a unique redex. For each redex, there is only one reduction rule that could apply. Thus evaluation is deterministic.  $\square$

**Lemma B.2 (Unique Decomposition).** For all  $e \in \text{Expr}$  either  $e \in \text{Ter}$  or there exists a unique evaluation context  $E$  and unique redex  $r$  such that  $e = E[r]$ .

*Proof.* By induction on the structure of  $e$ .

**Case  $e = t$ .**

Trivial.

**Case  $e = o e_a$ .**

Applying the inductive hypothesis to  $e_a$ , it follows that either (1)  $e_a \in \text{Ter}$  or (2) there exists unique  $E_a$  and  $r$  such that  $e_a = E_a[r]$ . For (1),  $e_a$  could be a value, in which case  $E = \square, r = o e_a$ . Otherwise,  $e_a = \text{err}_j^k$ , in which case  $E = o \square, r = \text{err}_j^k$ . For (2),  $E = o E_a$  since  $E[r] = (o E_a)[r] = o E_a[r] = o e_a = e$ . This decomposition is unique since  $E_a$  is unique.

**Case  $e = \text{mon}_j^k e_{\kappa} e_v$ .**

Apply induction to  $e_{\kappa}$ . Either (1)  $e_{\kappa} \in \text{Ter}$  or (2) there exists a unique  $E_{\kappa}$  and  $r$  such that  $e_{\kappa} = E_{\kappa}[r]$ . For (1), there are two subcases.

**Case  $e_{\kappa} = v_{\kappa}$ .**

Apply induction to  $e_v$ . If  $e_v = v$  then  $E = \square, r = \text{mon}_j^k v_{\kappa} v = e$ . If  $e_v = E_v[r]$  then  $E = \text{mon}_j^k v_{\kappa} E_v$ .

**Case  $e_{\kappa} = \text{err}_j^k$ .**

$E = \text{mon}_j^k \square e_v, r = \text{err}_j^k$ .

For (2),  $E = \text{mon}_j^k E_{\kappa} e_v$ .

**Otherwise.**

The remaining cases are similar to one of the above.  $\square$

## C Uniform evaluator proof

The proofs in this section hold for all languages presented in section 4.

**Theorem 5.2 (Uniform Evaluator).** Either  $\text{eval}_{\mathcal{L}}(e)$  is defined or the reduction sequence starting with  $\langle e, \emptyset \rangle$  is unbounded.

*Proof.* By interleaved application of lemma C.1 and lemma C.2.  $\square$

**Lemma C.1 (Progress).** If  $\sigma \vdash e$  then either  $e \in \text{Ter}$  or  $\langle e, \sigma \rangle \mapsto \langle e', \sigma' \rangle$ .

*Proof.* By lemma B.2 either  $e \in \text{Ter}$  or  $e = E[r]$ . By cases on  $r$ .

**Case**  $r = \text{if } v e_t e_f$ .

Either IF-TRUE or IF-FALSE apply.

**Case**  $r = \text{add! } v_\alpha v_a$ .

Suppose  $v_\alpha$  is an address. Since  $\sigma \vdash e$ ,  $\text{add}(\sigma, v_\alpha, v)$  is defined, so ADD! applies. If  $v_\alpha$  is not an address then ERR-ADD! applies.

**Case**  $r = \text{mon}_j^k v_\kappa v$ .

By cases on  $v_\kappa$ .

**Case**  $v_\kappa \notin \text{Con}$ .

ERR-MON applies.

**Case**  $v_\kappa = b$ .

Either MON-TRUE or MON-FALSE applies.

**Case**  $v_\kappa = \lambda x.e$ .

MON-FLAT applies.

**Case**  $v_\kappa = v_d \rightarrow_i v_c$ .

MON-FUN applies.

**Case**  $v_\kappa = \text{tr } v_b v_p$ .

MON-TRACE applies.

**Case**  $v_\kappa = \text{co } \alpha v_p$ .

MON-COL applies.

**Otherwise.**

The remaining cases are similar to one of the above.  $\square$

**Lemma C.2 (Preservation).** If  $\sigma \vdash e$  and  $\langle e, \sigma \rangle \mapsto \langle e', \sigma' \rangle$  then  $\sigma' \vdash e'$ .

*Proof.* By cases on the reduction relation.

**Case**  $\langle E[\text{if } v e_t e_f], \sigma \rangle \mapsto \langle E[e_t], \sigma \rangle, v \neq \text{false}$ .

Since  $\sigma \vdash \text{if } v e_t e_f$  we know  $\sigma \vdash e_t$ .

**Case**  $\langle E[(\lambda x.e_b) v], \sigma \rangle \mapsto \langle E[e_b[v/x]], \sigma \rangle$ .

This follows from lemma C.3.

**Case**  $\langle E[\text{add! } \alpha v], \sigma \rangle \mapsto \langle E[\alpha], \text{add}(\sigma, \alpha, v) \rangle$ .

This follows from lemma C.4.

**Case**  $\langle E[\text{mon}_j^k(\text{tr } v_b v_p) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^k(v_b(\text{co } \alpha v_p)) v], \sigma[\alpha \mapsto \text{null}] \rangle$ .

The contractum is closed since no new variables are introduced. A new address  $\alpha$  is introduced. For the expression to remain valid,  $\sigma \Vdash \sigma(\alpha)$  must hold which it does since  $\sigma(\alpha) = \text{null}$ .

**Case**  $\langle E[\text{mon}_j^k(\text{co } \alpha v_p) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^k(v_p(\text{add! } \alpha v)) v], \sigma \rangle$ .

No variables are introduced, no addresses are introduced, and the store is maintained. Therefore, the contractum remains closed with addresses still to valid queues.

**Otherwise.**

The remaining cases are similar to one of the above.  $\square$

$$\begin{array}{l}
1933 \quad \lambda x. \text{let } x_j = \text{mon}_j^{l,j} v_d x \text{ in } \sim \lambda x. \text{let } x_g = \text{mon}_j^l \tilde{v}_d x \text{ in} \quad \square \sim \square \\
1934 \quad \text{let } x_k = \text{mon}_j^{l,k} v_d x \text{ in} \quad \text{let } x_j = x_g \cdot j \text{ in} \quad \text{mon}_j^{k,l} E e \sim (\text{mon}_j^k \tilde{E} \tilde{e}) \cdot l \\
1935 \quad \text{mon}_j^{k,l} (v_c x_j) (v x_k) \quad \text{let } x_k = x_g \cdot kn \text{ in} \quad \text{mon}_j^{k,l} v E \sim (\text{mon}_j^k \tilde{v} \tilde{E}) \cdot l \\
1936 \quad \text{mon}_j^{k,l} (\tilde{v}_c x_j) (\tilde{v} x_k) \quad \dots \\
1937 \quad \text{let } x_j = e_j \text{ in} \quad \sim \text{let } x_j = \tilde{e}_j \text{ in} \\
1938 \quad \text{let } x_k = e_k \text{ in} \quad \text{let } x_k = \tilde{e}_k \text{ in} \\
1939 \quad \text{mon}_j^{k,l} (v_c x_j) (v x_k) \quad \text{mon}_j^{k,l} (\tilde{v}_c x_j) (\tilde{v} x_k) \\
1940 \quad \text{let } x_k = e_k \text{ in} \quad \sim \text{let } x_k = \tilde{e}_k \text{ in} \\
1941 \quad \text{mon}_j^{k,l} (v_c v_j) (v x_k) \quad \text{mon}_j^{k,l} (\tilde{v}_c \tilde{v}_j) (\tilde{v} x_k) \\
1942 \quad \text{mon}_j^{k,l} e_\kappa e \sim (\text{mon}_j^k \tilde{e}_\kappa \tilde{e}) \cdot l \\
1943 \quad \dots \\
1944 \\
1945 \\
1946 \\
1947 \\
1948 \\
1949 \\
1950 \\
1951 \\
1952 \\
1953 \\
1954 \\
1955 \\
1956 \\
1957 \\
1958 \\
1959 \\
1960 \\
1961 \\
1962 \\
1963 \\
1964 \\
1965 \\
1966 \\
1967 \\
1968 \\
1969 \\
1970 \\
1971 \\
1972 \\
1973 \\
1974 \\
1975 \\
1976 \\
1977 \\
1978
\end{array}$$

Fig. 16. Expression and Evaluation Context Simulation Relation

**Lemma C.3 (Substitution Preservation).** If  $\sigma \vdash \lambda x. e_b$  and  $\sigma \vdash v$  then  $\sigma \vdash e_b[v/x]$ .

*Proof.* By induction on  $e_b$ . □

**Lemma C.4 (Store Preservation).** If  $\sigma \vdash \alpha$  and  $\sigma \vdash v$  then  $\text{add}(\sigma, \alpha, v) \vdash \alpha$ .

*Proof.* By induction on  $|\text{dom}(\sigma)| - \alpha$ . □

## D Evaluator equivalence proof

This section shows the equivalence of  $\Lambda_B$  and  $\Lambda_C$  in the absence of queue mutations. Because no mutation occurs, the store is irrelevant to reduction calculations and is thus omitted. The proof proceeds by a simulation argument. Figure 16 relates  $\Lambda_B$  expressions and evaluation contexts to equivalent ones in  $\Lambda_C$ .

**Lemma D.1 (Mutation Freedom).** If expression  $e$  contains no queue subexpression, then it is mutation free.

*Proof.* Assume to the contrary that  $\langle e, \emptyset \rangle \mapsto^* \langle e', \emptyset \rangle \mapsto \langle e'', \sigma \rangle$  for  $\sigma \neq \emptyset$ . The latter reduction must be QUEUE because the only other store-manipulating rule, ADD!, presupposes a non-empty store. However, this is a contradiction since QUEUE only applies if the initial program  $e$  contains a queue subexpression. □

**Theorem 5.3 (Evaluator Equivalence).** If  $e$  is mutation free, then  $\text{eval}_{\Lambda_B}(e) = \text{eval}_{\Lambda_C}(e)$ .

**Note.** By design, trace contracts use mutation and the existing behavior of dependent function contracts is inappropriate for this case. Conversely, queue-mutating programs are excluded because it is the purpose of  $\Lambda_C$  to specify a behavior for  $\rightarrow_i$  that is appropriate when contracts perform mutation.

*Proof.* There are two directions to prove. First, that  $\text{eval}_{\Lambda_B} \subseteq \text{eval}_{\Lambda_C}$  on the restricted domain of mutation-free expressions. By cases on  $\text{eval}_{\Lambda_B}(e)$ .

**Case  $\text{eval}_{\Lambda_B}(e) = b$ .**

Thus  $e \mapsto_{\Lambda_B}^* b$ . Because  $e \sim e$ , lemma D.2 yields  $e \mapsto_{\Lambda_C}^* b_f$  and there exists  $\tilde{b}$  such that  $b_f \simeq_{\text{obs}} \tilde{b}$  and  $b \sim \tilde{b}$ . Observational equivalence and the simulation both preserve Booleans, therefore  $b_f = \tilde{b} = b$ . Hence,  $e \mapsto_{\Lambda_C}^* b$  and  $\text{eval}_{\Lambda_C}(e) = b$ .

**Case  $\text{eval}_{\Lambda_B}(e) = \text{opaque}$ .**

Similar to the prior case since preserving Booleans also implies preserving non-Booleaness.

The inverse direction states that  $\text{eval}_{\Lambda_C} \subseteq \text{eval}_{\Lambda_B}$ . There is only one interesting case, namely showing that the situation where  $e \mapsto_{\Lambda_C}^* t$  but  $\text{eval}_{\Lambda_B}(e)$  is undefined is impossible.

Assume the contrary. Using lemma D.3 yields a contradiction. By theorem 5.2, the reduction sequence in  $\Lambda_B$  is unbounded. Let  $e \mapsto_{\Lambda_B}^* e'$  and  $e \mapsto_{\Lambda_C}^* \tilde{e}'$  where  $e' \sim \tilde{e}'$  are the last pair of expressions related under  $\sim$ . This choice is possible since the reduction sequence in  $\Lambda_C$  is finite. Because  $e'$  can take a step, lemma D.3 applies and generates a later pair of related expressions, contradicting the choice of  $e' \sim \tilde{e}'$ .  $\square$

**Lemma D.2 (Transitive Simulation).** Let  $e$  be mutation free. If  $e \mapsto_{\Lambda_B}^* t$  and  $e \sim \tilde{e}$ , then there exists  $t_f$  and  $\tilde{t}$  such that  $\tilde{e} \mapsto_{\Lambda_C}^* t_f$ ,  $t_f \simeq_{\text{obs}} \tilde{t}$ , and  $t \sim \tilde{t}$ .

*Proof.* By induction on the number of steps  $n$  in  $e \mapsto_{\Lambda_B}^* t$ .

**Case  $n = 0$ .**

Trivial.

**Case  $n > 0$ .**

By lemma D.3,  $e \mapsto_{\Lambda_B}^+ e''$ ,  $\tilde{e} \mapsto_{\Lambda_C}^+ e_i$ ,  $e_i \simeq_{\text{obs}} \tilde{e}''$ , and  $e'' \sim \tilde{e}''$ . From lemma B.1,  $e'' \mapsto_{\Lambda_B}^* t$ . Applying the inductive hypothesis yields  $\tilde{e}'' \mapsto_{\Lambda_C}^* t_i$  where  $t_i \simeq_{\text{obs}} \tilde{t}$  and  $t \sim \tilde{t}$ . In summary,  $\tilde{e} \mapsto_{\Lambda_C}^+ e_i \simeq_{\text{obs}} \tilde{e}'' \mapsto_{\Lambda_C}^* t_i \simeq_{\text{obs}} \tilde{t}$ , which suffices.  $\square$

**Lemma D.3 (Simulation).** Let  $e$  be mutation free and  $e \sim \tilde{e}$ . If  $e \mapsto_{\Lambda_B} e'$ , then there exists  $e''$ ,  $e_i$ ,  $\tilde{e}''$  such that  $e \mapsto_{\Lambda_B}^+ e''$ ,  $\tilde{e} \mapsto_{\Lambda_C}^+ e_i$ ,  $e_i \simeq_{\text{obs}} \tilde{e}''$ , and  $e'' \sim \tilde{e}''$ .

*Proof.* By cases on  $e \mapsto_{\Lambda_B} e'$ . Each case relies on lemma D.4 followed by lemma D.5.

**Case  $E[\text{if } v \ e_t \ e_f] \mapsto E[e_t]$ ,  $v \neq \text{false}$ .**

Let  $\tilde{e} = \tilde{E}[\text{if } \tilde{v} \ \tilde{e}_t \ \tilde{e}_f]$ . The simulation preserves non-Booleans, so  $\tilde{v} \neq \text{false}$ . Thus,  $\tilde{E}[\text{if } \tilde{v} \ \tilde{e}_t \ \tilde{e}_f] \mapsto \tilde{E}[\tilde{e}_t]$ .

**Case  $E[(\lambda x. \text{let } x_j \ \text{---}) \ v] \mapsto E[\text{let } x_j \ \text{---}]$ .**

This reduction implies that  $\tilde{E}[(\lambda x. \text{let } x_g \ \text{---}) \ \tilde{v}] \mapsto \tilde{E}[\tilde{e}_i]$  where

$$\begin{aligned} e_i &= \text{let } x_g = \text{mon}_j^l \tilde{v}_d \tilde{v} \text{ in} \\ &\quad \text{let } x_j = x_g \cdot j \text{ in} \\ &\quad \text{let } x_k = x_g \cdot k \text{ in} \\ &\quad \text{mon}_j^{k,l} (\tilde{v}_c x_j) (\tilde{v} x_k). \end{aligned}$$

Because  $e$  is mutation free,  $e_i \simeq_{\text{obs}} \tilde{e}'$  where

$$\begin{aligned} \tilde{e}' &= \text{let } x_j = (\text{mon}_j^l \tilde{v}_d \tilde{v}) \cdot j \text{ in} \\ &\quad \text{let } x_k = (\text{mon}_j^l \tilde{v}_d \tilde{v}) \cdot k \text{ in} \\ &\quad \text{mon}_j^{k,l} (\tilde{v}_c x_j) (\tilde{v} x_k). \end{aligned}$$

Thus,  $\tilde{E}[e_i] \simeq_{\text{obs}} \tilde{E}[\tilde{e}']$ . Note that  $e' \sim \tilde{e}'$ , therefore  $E[\text{let } x_j \text{ ---}] \sim \tilde{E}[\tilde{e}']$ .

**Case**  $E[\text{let } x_j = v_j \text{ in ---}] \mapsto E[\text{let } x_k = e_k \text{ in ---}]$ .

$$\tilde{E}[\text{let } x_j = \tilde{v}_j \text{ in ---}] \mapsto \tilde{E}[\text{let } x_k = \tilde{e}_k \text{ in ---}]$$

**Case**  $E[\text{let } x_k = v_k \text{ in ---}] \mapsto E[\text{mon}_j^{k,l} (v_c v_j) (v v_k)]$ .

$$\tilde{E}[\text{let } x_k = \tilde{v}_k \text{ in ---}] \mapsto \tilde{E}[\text{mon}_j^{k,l} (\tilde{v}_c \tilde{v}_j) (\tilde{v} \tilde{v}_k)]$$

**Case**  $E[\text{mon}_j^{k,l} \text{ true } v] \mapsto E[v]$ .

$$\tilde{E}[(\text{mon}_j^{k,l} \text{ true } \tilde{v}) \cdot l] \mapsto \tilde{E}[(\text{grd}_j^{k,l} \text{ true } \tilde{v}) \cdot l] \mapsto \tilde{E}[\tilde{v}]$$

**Case**  $E[\text{mon}_j^{k,l} \text{ false } v] \mapsto E[\text{err}_j^k]$ .

$$\tilde{E}[(\text{mon}_j^{k,l} \text{ false } \tilde{v}) \cdot l] \mapsto \tilde{E}[\text{err}_j^k \cdot l] \simeq_{\text{obs}} \tilde{E}[\text{err}_j^k]$$

**Case**  $E[\text{mon}_j^{k,l} (\lambda x. e) v] \mapsto E[\text{mon}_j^{k,l} ((\lambda x. e) v) v]$ .

$$\tilde{E}[(\text{mon}_j^{k,l} (\lambda x. \tilde{e}) \tilde{v}) \cdot l] \mapsto \tilde{E}[(\text{mon}_j^{k,l} ((\lambda x. \tilde{e}) \tilde{v}) \tilde{v}) \cdot l]$$

**Case**  $E[\text{mon}_j^{k,l} (v_d \rightarrow_i v_c) v] \mapsto E[\lambda x. \text{let } x_j \text{ ---}]$ .

$$\tilde{E}[(\text{mon}_j^{k,l} (\tilde{v}_d \rightarrow_i \tilde{v}_c) \tilde{v}) \cdot l] \mapsto \tilde{E}[(\text{grd}_j^{k,l} (\tilde{v}_d \rightarrow_i \tilde{v}_c) \tilde{v}) \cdot l] \mapsto \tilde{E}[\lambda x. \text{let } x_g \text{ ---}]$$

**Otherwise.**

The remaining cases are similar to one of the above or are standard.  $\square$

**Lemma D.4 (Simulation Decomposition).** If  $e \sim \tilde{e}$  and  $e = E[e_s]$ , then exists  $\tilde{E}$  and  $\tilde{e}_s$  such that  $\tilde{e} = \tilde{E}[\tilde{e}_s]$  where  $E \sim \tilde{E}$  and  $e_s \sim \tilde{e}_s$ .

*Proof.* By induction on  $e \sim \tilde{e}$ .  $\square$

**Lemma D.5 (Simulation Composition).** If  $E \sim \tilde{E}$  and  $e \sim \tilde{e}$ , then  $E[e] \sim \tilde{E}[\tilde{e}]$ .

*Proof.* By induction on  $E \sim \tilde{E}$ .  $\square$

## E Compiler correctness proof

This section proves that the compiler is correct. Like appendix D, the proof follows from a simulation argument. However, the simulation relation is the compiler function  $\mathcal{C}$  itself extended to the evaluation syntax. Since the evaluation syntax contains collectors,  $\mathcal{C}$  defines the compilation of collectors following the description in section 6.1. Figure 17 defines the relevant extension of  $\mathcal{C}$ .

$$\begin{array}{l}
2071 \quad \mathcal{C}(e \cdot l) = \mathcal{C}(e) \cdot l \qquad \qquad \qquad \mathcal{C}(\square) = \square \\
2072 \quad \mathcal{C}(\text{tr } e_b e_p) = \begin{cases} \text{let } x_b = \mathcal{C}(e_b) \text{ in} \\ \text{let } x_p = \mathcal{C}(e_p) \text{ in} \\ \lambda\_ \cdot \text{let } x_\alpha = \text{queue in} \\ \quad x_b \mathcal{C}(\text{co } x_\alpha x_p) \end{cases} \qquad \begin{array}{l} \mathcal{C}(\text{tr } E e) = \text{tr } \mathcal{C}(E) \mathcal{C}(e) \\ \mathcal{C}(\text{tr } v E) = \text{tr } \mathcal{C}(v) \mathcal{C}(E) \\ \dots \end{array} \\
2073 \\
2074 \\
2075 \\
2076 \quad \mathcal{C}(\text{co } \alpha v_p) = \lambda y. v_p (\text{add! } \alpha y) \\
2077 \\
2078 \quad \dots
\end{array}$$

Fig. 17. Expression and Evaluation Context Compiler

**Theorem 6.1 (Compiler Correctness).**  $\text{eval}_{\Lambda_T} = \text{eval}_{\Lambda_C} \circ \mathcal{C}$

*Proof.* Similar to the proof of theorem 5.3. Let  $e \in \Lambda_T$ . It suffices to show that if  $\sigma \vdash e$  and  $\langle e, \sigma \rangle \mapsto \langle e', \sigma' \rangle$ , then there exists  $e''$  and  $\sigma''$  such that  $\langle e, \sigma \rangle \mapsto^* \langle e'', \sigma'' \rangle$  and  $\langle \mathcal{C}(e), \mathcal{C} \circ \sigma \rangle \mapsto^* \langle \mathcal{C}(e''), \mathcal{C} \circ \sigma'' \rangle$ . By cases on  $\langle e, \sigma \rangle \mapsto \langle e', \sigma' \rangle$ .

**Case**  $\langle E[\text{mon}_j^k(\text{tr } v_b v_p) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^k(v_b(\text{co } \alpha v_p)) v], \sigma[\alpha \mapsto \text{null}] \rangle$ .

The compiled reduction sequence mirrors this step:

$$\begin{array}{l}
2090 \quad \langle \mathcal{C}(E[\text{mon}_j^k(\text{tr } v_b v_p) v]), \mathcal{C} \circ \sigma \rangle \\
2091 \quad = \langle \mathcal{C}(E)[\text{mon}_j^k \mathcal{C}(\text{tr } v_b v_p) \mathcal{C}(v)], \mathcal{C} \circ \sigma \rangle \\
2092 \\
2093 \quad = \langle \mathcal{C}(E)[\text{mon}_j^k(\text{let } x_b = \mathcal{C}(v_b) \text{ in} \\
2094 \quad \quad \text{let } x_p = \mathcal{C}(v_p) \text{ in} \\
2095 \quad \quad \lambda\_ \cdot \text{let } x_\alpha \text{ — } ) \mathcal{C}(v)], \mathcal{C} \circ \sigma \rangle \\
2096 \\
2097 \quad \mapsto^+ \langle \mathcal{C}(E)[\text{mon}_j^k(\text{let } x_\alpha = \text{queue in} \\
2098 \quad \quad \mathcal{C}(v_b) \mathcal{C}(\text{co } \mathcal{C}(v_p) x_\alpha)) \mathcal{C}(v)], \mathcal{C} \circ \sigma \rangle \\
2099 \\
2100 \quad \mapsto^+ \langle \text{mon}_j^k \mathcal{C}(v_b) \mathcal{C}(\text{co } \alpha \mathcal{C}(v_p)) \mathcal{C}(v), \mathcal{C} \circ \sigma' \rangle
\end{array}$$

**Case**  $\langle E[\text{mon}_j^k(\text{co } \alpha v_p) v], \sigma \rangle \mapsto \langle E[\text{mon}_j^k(v_p(\text{add! } \alpha v)) v], \sigma \rangle$ .

$$\begin{array}{l}
2104 \quad \langle \mathcal{C}(E[\text{mon}_j^k(\text{co } \alpha v_p) v]), \mathcal{C} \circ \sigma \rangle \\
2105 \quad = \langle \mathcal{C}(E)[\text{mon}_j^k \mathcal{C}(\text{co } \alpha v_p) \mathcal{C}(v)], \mathcal{C} \circ \sigma \rangle \\
2106 \\
2107 \quad = \langle \mathcal{C}(E)[\text{mon}_j^k(\lambda y. \text{ — } ) \mathcal{C}(v)], \mathcal{C} \circ \sigma \rangle \\
2108 \\
2109 \quad \mapsto^+ \langle \mathcal{C}(E)[\text{mon}_j^k(\mathcal{C}(v_p) (\text{add! } \alpha \mathcal{C}(v))) \mathcal{C}(v)], \mathcal{C} \circ \sigma \rangle
\end{array}$$

**Otherwise.**

The remaining cases are straightforward.

The inverse direction follows from an argument similar to the one made in the proof of theorem 5.3.  $\square$

**Lemma E.1 (Simulation Decomposition).**  $\mathcal{C}(E[e]) = \mathcal{C}(E)[\mathcal{C}(e)]$

*Proof.* By induction on  $E$ . □

### F Trace contracts for racket/draw

The following items describe the properties that racket/draw<sup>12</sup> either maintains through defensive-programming checks or documents but does not check:

1. A call to `get-data-from-file` must return `false` unless the bitmap is created with `save-data-from-file` and the image is loaded successfully.
2. The `load-file` method of `bitmap%` cannot be called with bitmaps created by `make-platform-bitmap`, `make-screen-bitmap`, or `make-bitmap` in `canvas%`.
3. The methods `get-text-extent`, `get-char-height`, and `get-char-width` can be called before a bitmap is installed. All others must be called after a bitmap is installed.
4. The method `set-argb-pixels` cannot be called if the given bitmap is produced by `make-screen-bitmap` or `make-bitmap` in `canvas%`.
5. A bitmap can be installed into at most one bitmap drawing context and only when it is not used by a control (as a label), a `pen%`, or a `brush%`.
6. A brush cannot be modified while it is installed into a drawing context.
7. A brush cannot be modified if it is obtained from a `brush-list%`.
8. A color cannot be modified if it is created by passing a string to `make-object` or by retrieving a color from the color database.
9. The methods `start-doc`, `start-page`, `end-page`, and `end-doc` from `dc<%>` must be called in the correct order.
10. Some methods of `dc-path%` extend an open sub-path, some close an open sub-path, and some add closed sub-paths to an existing path. Those must all be kept consistent, e.g. if a method can only extend an open sub-path, then it cannot be called on an object where no sub-path is open.
11. A pen cannot be modified if it is obtained from a `pen-list%`.
12. A pen cannot be modified while it is installed into a drawing context.
13. If `as-eps` is set in a `post-script-dc%` object, then only one page can be created.
14. The `is-empty?` method of `region%` can only be called when associated with a drawing context.
15. There are no restrictions on the sequence of `start-doc`, `start-page`, `end-page`, and `end-doc` for `record-dc%`.

The revision of racket/draw enforces all of these properties with trace contracts.

<sup>12</sup> <https://docs.racket-lang.org/draw/>