

Trace Typing: An Approach for Evaluating Retrofitted Type Systems

Esben Andreasen¹, Colin S. Gordon², Satish Chandra³, Manu Sridharan³, Frank Tip³, and Koushik Sen⁴

- 1 Aarhus University
esbena@cs.au.dk
- 2 Drexel University
csgordon@cs.drexel.edu
- 3 Samsung Research America
{schandra,m.sridharan,ftip}@samsung.com
- 4 UC Berkeley
ksen@berkeley.edu

Abstract

Recent years have seen growing interest in the retrofitting of type systems onto dynamically-typed programming languages, in order to improve type safety, programmer productivity, or performance. In such cases, type system developers must strike a delicate balance between disallowing certain coding patterns to keep the type system simple, or including them at the expense of additional complexity and effort. Thus far, the process for designing retrofitted type systems has been largely ad hoc, because evaluating multiple variations of a type system on large bodies of existing code is a significant undertaking.

We present *trace typing*: a framework for automatically and *quantitatively* evaluating variations of a retrofitted type system on large code bases. The trace typing approach involves gathering traces of program executions, inferring types for instances of variables and expressions occurring in a trace, and merging types according to merge strategies that reflect specific (combinations of) choices in the source-level type system design space.

We evaluated trace typing through several experiments. We compared several variations of type systems retrofitted onto JavaScript, measuring the number of program locations with type errors in each case on a suite of over fifty thousand lines of JavaScript code. We also used trace typing to validate and guide the design of a new retrofitted type system that enforces fixed object layout for JavaScript objects. Finally, we leveraged the types computed by trace typing to automatically identify tag tests — dynamic checks that refine a type — and examined the variety of tests identified.

1998 ACM Subject Classification F.3.3 Studies of Program Constructs

Keywords and phrases Retrofitted type systems, Type system design, trace typing

Digital Object Identifier 10.4230/LIPIcs.ECOOP.2016.

1 Introduction

In recent years, there have been a number of efforts to *retrofit* [21] type systems onto dynamically-typed languages, to aid developer productivity, correctness, and performance. These languages are of increasing importance, primarily due to their common use in web applications on both the client and server side. As more large-scale, complex programs are written in such languages, greater need arises for static types, due to the resulting benefits for static error checking, developer tools, and performance. Recent high-profile projects in



© Esben Andreasen, Colin S. Gordon, Satish Chandra, Manu Sridharan, Frank Tip, Koushik Sen; licensed under Creative Commons License CC-BY

30th European Conference on Object-Oriented Programming (ECOOP 2016).

Editor: John Q. Open and Joan R. Access; Article No. ; pp. 1–25

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

type system retrofitting include Closure [16], TypeScript [5] and Flow for JavaScript [13], and Hack for PHP [32]. Given the proliferation of dynamically-typed languages, there are many retrofitted type systems ahead.

The key questions in the design of these type systems involve selecting which features to include. Richer features may allow more coding patterns to be understood and validated by the type checker. However, richer type systems come at the cost of greater complexity and implementation effort, and it is rarely a priori obvious whether a given feature’s benefit outweighs its implementation burden. Hence, great care must be taken in deciding if a particular type system feature is worthy of inclusion. Such decision points continue to arise as a type system evolves: e.g., TypeScript recently added union types [29], and Flow recently added bounded polymorphism [8].

When retrofitting a type system, there is generally a wealth of existing code to which one would like to introduce types. When considering a type system feature, it would be very helpful to know *how* useful the feature would be for typing an existing body of code. Unfortunately, truly gauging the feature’s usefulness would require implementing type checking (and possibly inference) for the feature and adding the necessary annotations to the code base. This involves a significant effort, which ultimately may not have lasting value if an alternate type system proves to be more suitable. Some previous work [3, 12, 21, 23] reduces the first burden by providing a reusable framework for implementing type systems, with some inference. However, each still requires manual annotation, so large-scale evaluation remains a substantial undertaking — in one case an expert required over 17 hours to annotate part of one program with type qualifiers [17]. For this reason, in current practice, decisions are often made based on intuition and anecdotal evidence, without the aid of carefully collected quantitative results on the impact of certain features. Anecdotes and other qualitative criteria may validate inclusion of features that are rarely used but important, but quantitative results on a large corpus can rapidly guide the high level design.

In this work, we propose a novel framework, called *trace typing*, for *automatic* and *quantitative* evaluation of retrofitted type system features for existing code bases. Our system works by collecting detailed, unrolled traces of dynamic executions of the code in question. Evaluating a type system feature then requires implementing type checking and inference for the unrolled trace, and also defining a *merge strategy* (explained shortly) for combining the types of different occurrences of a variable in the trace into the corresponding static program entity. The crucial observation here is that type checking and effective inference for the unrolled trace is *far* simpler than static type checking / inference for the full source language, and requires no type annotations due to the simplicity of the trace format. Hence, our system dramatically reduces the implementation effort required to perform quantitative comparisons of different type system variants, and allows type system designers to *automatically* gather feedback from large code bases by typing the trace.

While type inference and checking for unrolled traces is simpler to implement than a static type checker, the ultimate goal of the designer is a type system for source programs, not traces. Our key insight is that—given types for the unrolled trace—many source type systems can be expressed via a family of *merge* operations, used to merge the types of runtime entities that the source type system cannot distinguish. For example, a merge operator can control whether assigning an integer and an object into the same variable yields a union type for the variable, or a type error. (We give examples of merging throughout the paper.) The type errors found by using merged types give an indicative lower bound on the set of program locations that would be ill-typed under a static type checker, as long as the merge operator produces useful types. A designer can, therefore, evaluate the usefulness of a

proposed static type system via suitable merge operators in the trace typing framework, and can also compare the relative power of different static type systems.

We have implemented our techniques in a system for evaluating type system variants for JavaScript. We evaluated our system on a diverse suite of programs, and evaluated several different variations on a core object-oriented type system. In particular, our experiments yielded quantitative answers to the following questions for our benchmarks:

- How pervasive is the need for supporting union types on top of a type system that supports subtyping?
- If union types are supported, what kind of “tag tests” usually occur to discriminate between branches of a union?
- How pervasive is the need for supporting parametric polymorphism?
- How pervasive is the need for intersection types for functions?

The results are given in Section 6. We believe trace typing gives a quick and easy way to obtain answers to these (and similar) questions on a given code corpus.

We have also used trace typing to guide the design of a type system to support ahead-of-time compilation of JavaScript extending previous work [9]. Trace typing allowed us to validate some of our design choices *without implementing a type inference engine* that works on real code. Moreover, it helped point to where we need to extend the type system. Section 6.4 details this case study.

Trace typing is designed to present a rapid, aggregate, quantitative picture to help a type system designer decide whether to support certain features and prioritize the development of a type checker. When features are not supported, a programmer needing said features must either ignore the warnings from the type system, or refactor her code to fit the type system restrictions. For example, if trace typing shows that parametric polymorphism is rarely needed in the code of interest, then the designer can defer supporting polymorphism; in the interim, developers must work around the limitation, e.g., by code cloning. Of course, a type system designer will need to consider other factors besides the output of trace typing when prioritizing, such as idioms used by critical modules, or the difficulty of converting code to fit within certain restrictions.

Contributions This paper makes the following contributions:

- We propose *trace typing* as an approach for easily evaluating the usefulness of type system features, allowing large bodies of existing code to inform type system design. Trace typing requires far less effort than implementing various static type checkers, and potentially adding type annotations to large bodies of code.
- Using the trace typing approach, we systematically compare the precision of a number of object type systems for JavaScript on some of the most popular packages for `node.js`, totaling over 50,000 lines of code. We include union types as well as several variants of polymorphism in our study.
- We describe our experience with trace typing to guide and check choices in the design of a type system for use by an optimizing JavaScript compiler.
- We use trace typing to automatically identify tag tests in our large corpus, and analyze the relative frequency of different tests and their structure.

2 Trace Typing by Examples

In this section, we show informally how trace typing can be used to carry out different kinds of quantitative experiments relevant when designing a type system. Section 3 gives a more formal description of the framework.

```

1 function f(a) { a.p = 7; }
2 function g(b) { return b; }
3 var x = { p: 3 };
4 var y = { p: 4, q: "hi" };
5 var z = { q: "bye", r: false };
6 f(x); // f.6: {p: Number} -> void
7 f(y); // f.7: {p: Number, q: String} -> void
8 g(y); // g.8: {p: Number, q: String} -> {p: Number, q: String}
9 var w = g(z); // g.9: {q: String, r: Boolean} -> {q: String, r: Boolean}
10 w.r = true;

```

■ **Figure 1** A JavaScript program to illustrate polymorphism.

2.1 Polymorphism

Figure 1 gives a small JavaScript example. The program allocates new objects at lines 3–5, and then calls functions `f` and `g`. The `f` function accesses the `p` field, while `g` just returns its parameters. All field accesses are well-behaved: they access pre-existing fields, and fields are only updated with a value of a compatible type. Here, we show how trace typing can *quantify* the relative effectiveness of two type systems with different levels of polymorphism for handling this code.

We focus on the types observed in the dynamic trace for functions `f` and `g`. For our purpose, these types correspond to the parameter and return types observed per function call site. So, for function `f` we observe the type $\{p: \text{Number}\} \rightarrow \text{void}$ for the call at line 6, and $\{p: \text{Number}, q: \text{String}\} \rightarrow \text{void}$ for line 7. Similarly, for function `g` we observe type $\{p: \text{Number}, q: \text{String}\} \rightarrow \{p: \text{Number}, q: \text{String}\}$ for the line 8 call, and $\{q: \text{String}, r: \text{Boolean}\} \rightarrow \{q: \text{String}, r: \text{Boolean}\}$ for line 9. Note that the types above are not based on inference on the body of functions `f` and `g`.

Type system variants can be distinguished by how their merge strategies combine these types into a single function type each for `f` and for `g`. We consider the following merge strategies. \mathbb{T}_{sub} merges, separately the argument and return types, in each case taking the least upper bound in a suitable subtyping lattice. For our example, this strategy yields the type $\{p: \text{Number}\} \rightarrow \text{void}$ for `f` and $\{q: \text{String}\} \rightarrow \{q: \text{String}\}$ for `g`.¹ \mathbb{T}_{poly} merges the function types by introducing type variables, arriving at a signature with parametric polymorphism (details in Section 4.3). The type we obtain for `f` is the same as in the case of \mathbb{T}_{sub} , but for `g`, it is $(E) \rightarrow E$; E is a type parameter.

Next, the trace typing system performs type checking using the merged types, and reports the number of type errors to the user. This type error count identifies program locations that a static checker for the type system would be unable to validate, a valuable statistic for determining the usefulness of the variant in practice (fewer errors² means that the variant can validate more code patterns).

With \mathbb{T}_{sub} , the types are sufficient for type-checking the bodies of the functions, but `g`'s type is still insufficient to typecheck the dereference at line 10, leaving one error remaining. With \mathbb{T}_{poly} , the types are sufficient to ensure there is no type error in the program. The type checking phase also checks conformance of the actual parameters being passed to function calls at lines 6, 7, 8 and 9; these checks pass for both \mathbb{T}_{sub} and \mathbb{T}_{poly} .

\mathbb{T}_{sub} produces one type error on this trace, while \mathbb{T}_{poly} produces 0. These counts are

¹ Note that we did not compute a type for `g` via least-upper-bound with respect to *function* subtyping (with contravariance in argument position); see Section 3.1 for further details.

² Throughout this paper, we use *error* as a synonym for *static type error*, not to refer to actual developer mistakes in a program.

```

1 function f(x){
2   ... x ... // x.2: A | int
3   if(x instanceof A){ // x.3: A | int
4     ... x ... // x.4: A
5   }
6   ... x ... // x.6: A | int
7 }
8 f(new A());
9 f(3);

```

■ **Figure 2** Tag test example.

underestimates with respect to a (hypothetical) static type checker, both due to our reliance on dynamic information, and due to the over-approximations inherent in any static type checking. Nevertheless, across a large code base, such data could be invaluable in finding the extent to which a particular type system feature, e.g. parametric polymorphism, is useful. Section 6.2 presents results from an experiment that evaluated the usefulness of parametric polymorphism and a synthetic type system with “unbounded” polymorphism on a suite of benchmarks, via comparison of error counts.

2.2 Discriminating Unions

Suppose we wish to design a type system for JavaScript that is equipped to reason about tag checks at the points where the program refines a union type. Such a construct is shown in the example of Figure 2. In this example, a union type with more than one case is narrowed to a specific case guarded by a condition. In general, the guard can be an arbitrary expression, but that makes type checking difficult — it would need to track an arbitrary number of predicates in carrying out an analysis. For example, the following “non-local” variation of the type guard in the example in Figure 2 would be more complex to handle:

```

1   var y = x instanceof A
2   ...
3   if (y) { ...

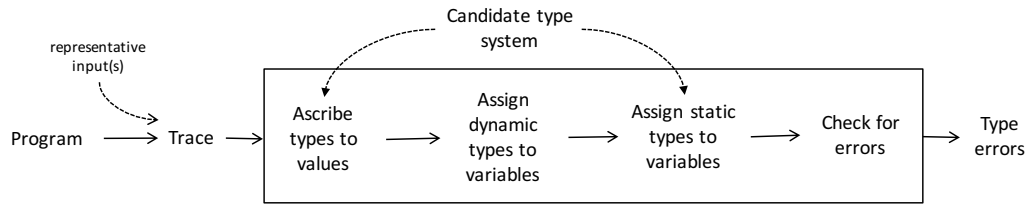
```

It is simpler to support a limited number of syntactic forms.

Can a type system designer obtain *quantitative* feedback on what kind of conditional constructs occur in the code corpora of interest? Are non-local type guards common? What about conjunctions and disjunctions of predicates? Simple syntactic pattern matching is an unreliable way to find this out, because it is fundamentally a flow analysis problem.

Trace typing makes this experiment easy. Trace typing ascribes a type to each occurrence of each variable in the program based on dynamic observation. Given these types, one can find pairs of successive reads of a variable without an intervening write, where the type ascribed to the variable in the second read is different from that of the first. This is a heuristic to identify instances of a variable’s type becoming narrower on an unrolled path, likely to be a tag test.

Consider the example in Figure 2 again. We examine some of the type ascriptions carried out by trace typing during an execution of the program. Here, at line 2, `x` refers to an object of type `A` and an `int`, respectively, in the invocations from line 8 and line 9. Let us label the two dynamic occurrences of `x` as `x.2.1` and `x.2.2`, where the first index refers to the line on which the variable occurs and the second refers to the invocation sequence number. The same types will be observed for `x.3.1` and `x.3.2`. At line 4, `x.4.1` observes `A`; that is the only dynamic occurrence of `x` on this line. Finally at line 6, `x.6.1` and `x.6.2` observe `A` and `int` respectively.



■ **Figure 3** Trace typing framework overview.

Crucially, trace typing can be configured to *merge* the types ascribed to variable occurrences on the same line, but across different invocations. Say that our type system includes union types. In that case, $x.2$ (the result of merging $x.2.1$ and $x.2.2$) gets the type $A \mid \text{int}$, which represents a union type consisting of types A and int . For $x.3$ we get $A \mid \text{int}$ as well. However, for $x.4$, we get just A . Hence, for successive reads $x.3$ and $x.4$ (with no intervening write), we observe the situation that the second type is a *refinement* of the first. This is a clue that the conditional at line 3 (that falls between the first and the second reads) is a type guard.

Using this technique on a large code corpus, we found that while non-local type guards are uncommon, boolean combinations of predicates are commonly used in type guards. Section 6.3 details the experiment we carried out to find the nature of tag tests in real code.

3 Trace typing

Figure 3 gives an overview of the trace typing process. The trace typing framework takes two inputs:

- a type system of interest, specified by defining various parameters in a framework (Section 3.1); and
- a program P of interest, and inputs for exercising P .

Given these inputs, trace typing proceeds in the following four phases:

1. An unrolled *trace* is produced by executing an instrumented version of P on the provided inputs (Section 3.2).
2. The types of all values (including objects and functions) and variables in the trace are inferred (Section 3.3).³
3. The inferred types are *merged*, coarsening the precise types in the trace to a corresponding source typing (Section 3.4).
4. The resulting types are used for type checking, to estimate the ability of the type system to handle the observed behaviors (Section 3.5).

The number of errors reported in the last step is a useful metric for evaluating type system features (see Section 7 for further discussion). Our framework has proven to be sufficiently expressive to model even modern industrial proposals for retrofitted type systems (Section 4.4). Also, the computed types can be used as part of experiments to test richer type system features, such as tag tests (Section 4.2).

3.1 Trace Type Systems

Here, we describe the interface by which a type system is specified to the trace typing framework. A type system \mathbb{T} has five components:

³ In the remainder of the paper, we use the terms “ascribed” and “inferred” interchangeably.

$\mathbb{T}.\alpha: (\mathbf{Value}) \rightarrow \mathbf{Type}$ ascribes a type to a concrete value. For JavaScript, values include both primitives as well as object instances.

$\mathbb{T}.\sqcup: (\mathbf{Type}, \mathbf{Type}) \rightarrow \mathbf{Type}$ computes the least-upper-bound of the input types in the subtyping lattice for the type system. For termination, we require $\mathbb{T}.\sqcup$ to be monotonic.

$\mathbb{T}.\vec{\sqcup}: (\mathbf{Function-type} \dots) \rightarrow \mathbf{Function-type}$ takes a list of types for individual invocations of a function \mathbf{f} and computes a type that can accommodate all invocations. A type for an invocation is computed by applying $\mathbb{T}.\alpha$ to the receiver, argument, and return values, and then combining them into a function type. $\mathbb{T}.\vec{\sqcup}$ often differs from applying $\mathbb{T}.\sqcup$ to the invocation types; further discussion below.

$\mathbb{T}.\text{equiv}: (\mathbf{FI} \mid \mathbf{FS}) \times (\mathbf{CI} \mid \mathbf{CS})$ governs how types for different trace occurrences of the same source variable are merged to approximate source-level typing (see Section 3.4). The first component determines whether to maintain a separate type for a local variable at different program points (flow sensitivity), while the second component determines if separate types should be maintained across different function invocations (context sensitivity). Note that use of \mathbf{CI} vs. \mathbf{CS} is usually dictated by the choice of $\mathbb{T}.\vec{\sqcup}$ (see Section 6.2). Also, our implementation supports more fine-grained flow- and context-sensitivity settings; here we only present two options for simplicity.

$\mathbb{T}.\text{check}: (\mathbf{TypeEnv}, \mathbf{Statement}) \rightarrow \mathbf{number}$ is a type checking function that counts the number of type errors a trace statement causes in a given type environment.

Together, $\mathbb{T}.\sqcup$, $\mathbb{T}.\vec{\sqcup}$, and $\mathbb{T}.\text{equiv}$ comprise the type system's *merge strategy*. **Value** represents the values of our trace language (see Section 3.2), while **Type** and **Function-type** are types defined entirely by \mathbb{T} . Defining $\mathbb{T}.\alpha$ for JavaScript objects is non-trivial, since the shape and property types of the object may change over time; we detail the issues involved in Section 4.1.

Note that $\mathbb{T}.\vec{\sqcup}$ should *not* use contravariance when handling function parameters, unlike typical function subtyping. $\mathbb{T}.\vec{\sqcup}$ is used to compute a type τ for function \mathbf{f} that is consistent with all observed invocations of \mathbf{f} . Hence, each parameter type in τ should accommodate *any* of the corresponding parameter types from the individual calls, a covariant handling. For an example, see the handling of calls to \mathbf{g} in Figure 1 (Section 2.1).

Example 1: Simple OO To simulate a basic structural OO type system with width subtyping:

- $\mathbb{T}.\alpha$ gives a type to an object by recursively characterizing the types of its properties (see Section 4.1 for details).
- $\mathbb{T}.\sqcup$ gives the least-upper-bound of two object types (retaining common fields that have identical types) and returns the same type if given two identical types. Otherwise it returns \top .
- $\mathbb{T}.\vec{\sqcup}$ applies $\mathbb{T}.\sqcup$ to the individual argument and return types (*not* using standard function subtyping; see above) to compute a generalized function type.
- $\mathbb{T}.\text{equiv}$ is $(\mathbf{FI}, \mathbf{CI})$.
- $\mathbb{T}.\text{check}$ performs standard assignment compatibility checks for variable and field writes, property type lookup on property reads, and standard checks at function invocations.

```

<trace> ::= <statement>*
<statement> ::= <var> = <expression> | <var>.<name> = <var> | delete <var>.<name> | <meta>
<expression> ::= <var> | <var>.<name> | allocate | null | undefined | <bool> | <number> | <string>
<var> ::= v1 | v2 | v3, ...
<name> ::= a | b | ...
<meta> ::= begin-call<var><var><var>* | end-call<var> | end-initialization<var>

```

■ **Figure 4** TL grammar

```

1 <tmp0@G> = allocate // ... allocation and assignment at line 6
2 <tmp1@G> = 3
3 <tmp0@G>.p = <tmp1@G>
4 end-initialization <tmp0@1>
5 <x@G> = <tmp0@G>
6 begin-call <f@G> <x@G> // ... first call to f, at line 9
7 <a@f_1> = <x@G>
8 <tmp0@f_1> = 7
9 <a@f_1>.p = <tmp0@f_1>
10 end-call
11 begin-call <f@G> <y@G> // ... second call to f, at line 10
12 <a@f_2> = <y@G>
13 <tmp0@f_2> = 7
14 <a@f_2>.p = <tmp0@f_2>
15 end-call

```

■ **Figure 5** Excerpts from the trace for the program in Figure 1. The name of variables include their scope, e.g. `<x@G>` is the `x` variable in the global scope, and `<a@f_2>` is the `a` variable during the second call to `f`.

Example 2: Union Types To extend Example 1 with union types, we modify \mathbb{T}_{\sqcup} to introduce unions when merging different sorts of types (e.g., a number and object type). For the code ‘`var x = 3; x = { f : 4 };`’, `x` would be given the union type `number | {f : number}`: Section 4.1 discusses how we perform type checking in the presence of such union types.

Example 3: \mathbb{T}_{sub} and \mathbb{T}_{poly} revisited We referred to \mathbb{T}_{sub} and \mathbb{T}_{poly} in Section 2.1. \mathbb{T}_{sub} is precisely the Simple OO of Example 1. \mathbb{T}_{poly} replaces $\mathbb{T}_{sub}^{\vec{\square}}$ of \mathbb{T}_{sub} to introduce parametric polymorphism (described further in Section 4.3).

3.2 Traces

The execution of a source code program can be recorded as a finite sequence of primitive instructions that reproduces the dataflow of the execution. We call such a sequence a trace. A trace is a simplified version of the original source program:

- Every dynamic read or write of a variable or subexpression is given a unique identifier.
- Control flow structures are erased: conditionals are reduced to the instructions of the conditional expression and chosen branch, loops are unrolled, and calls are inlined.
- Native operations are recorded as their results; e.g. a call to the runtime’s `cosine` function is recorded as the computed value.

We express our JavaScript traces as programs in a simple language TL. TL is an imperative, object-oriented language without control flow structures or nested expressions. TL has reads and writes for variables and fields, deletion for fields, and values (objects and primitive values). The syntax of TL can be seen in Figure 4.

Figure 5 shows excerpts from the trace for executing the code in Figure 1. Many JavaScript details are omitted here for clearer exposition but handled by our implementation. The variable names in the trace are generated from source variable names where present, and

$$\begin{aligned}
\hat{\Gamma}(x) &= \hat{\Gamma}_m(x) \mathbb{T}.\sqcup \hat{\Gamma}_p(x) & \hat{\Gamma}_m(x) &= \mathbb{T}.\bigsqcup_{x_i \in \hat{x}} \hat{\Gamma}(x_i) \\
\hat{\Gamma}_p(x) &= \mathbb{T}.\bigsqcup_{x=rhs' \in stmts} \begin{cases} \Gamma_0(x), & rhs = \text{allocate} \vee rhs \in \text{primitives} \\ \hat{\Gamma}(y), & rhs = y \\ \hat{\Gamma}(b).p, & rhs = b.p \wedge \hat{\Gamma}(b) \text{ is object with property } p \\ \Gamma_0(b).p, & rhs = b.p \wedge (\hat{\Gamma}(b) \text{ not an object} \vee \hat{\Gamma}(b).p \text{ not present}) \end{cases}
\end{aligned}$$

■ **Figure 6** Equations for type merging and propagation.

otherwise fresh temporaries are generated. (Note that fresh temporaries are used within each call to `f`.) A source mapping for each trace statement is maintained separately.

The trace contains some meta-statements to aid in later analyses. Each call is demarcated by `begin-call` and `end-call` statements, to aid in recovering the relevant types for the invocation. Data flow from parameter passing and return values is represented explicitly. The `end-initialization` statement marks the end of a constructor or initialization of an object literal.

3.3 Initial Type Ascription

Given a trace and a type system as specified in Section 3.1, trace typing first ascribes precise types to variables and values in the trace, without consideration for mapping types back to the source code. First, types are ascribed to all values. Primitive values and objects are handled directly using $\mathbb{T}.\alpha$, while functions are handled by combining the type for each invocation with $\mathbb{T}.\vec{\alpha}$, as described in Section 3.1.

Once types for values have been ascribed, each trace variable is given the type of the value it holds (recall that each trace variable is only written once). For a given trace T , we define the initial type environment Γ_0 as: $\forall v \in \text{Variables}(T). \Gamma_0(v) = \mathbb{T}.\alpha(\text{ValueOf}(T, v))$. This typing for trace variables is very precise—every variable access is given its own type, based solely on the value it holds at runtime. The next phase (Section 3.4) generalizes these initial types to more closely mimic the desired source-level treatment of variables.

Example Consider the variable `<tmp1@G>` at line 2 in Figure 5; it gets the value `3`, for which $\mathbb{T}_{sub}.\alpha$ yields the type `Number`. Similarly the type of `<x@G>` at line 5 is `{p: Number}` because that is the type ascribed to the value in `<tmp0@G>` at line 5.

3.4 Type Merging and Propagation

To model realistic type systems, we construct a less precise type environment $\hat{\Gamma}$ by merging the types of “equivalent” trace variables in Γ_0 . Variable equivalence classes are determined by the $\mathbb{T}.\text{equiv}$ operator provided by the type system (see Section 3.1). If the first component of $\mathbb{T}.\text{equiv}$ is `FI`, we employ a flow-insensitive treatment of variables: all variable occurrences `v_i` within a single dynamic call corresponding to the same source variable `v` are made equivalent. No such equivalences are introduced if the first $\mathbb{T}.\text{equiv}$ component is `FS`. If the second component of $\mathbb{T}.\text{equiv}$ is `CI`, we impose a context-insensitive treatment of variables: each matching `v_i` across function invocations (i.e., the same read or write of source variable `v`) is placed in the same equivalence class.

Given variable equivalence classes and Γ_0 , Figure 6 defines $\hat{\Gamma}$ in terms of two components: $\hat{\Gamma}_m$ for *merging* equivalent variables, and $\hat{\Gamma}_p$ for *propagating* across assignments. (As $\hat{\Gamma}_m$

```

1: f :: <({p: Number}) -> Undefined>
2: g :: <({q: String}) -> {q: String}>>
6: x :: {p: Number}
7: y :: {p: Number, q: String}
8: z :: {q: String, r: Boolean}
12: w :: {q: String}

```

■ **Figure 7** Types ascribed for variable writes in Figure 1 with \mathbb{T}_{sub} .

and $\hat{\Gamma}_p$ are themselves defined in terms of $\hat{\Gamma}$, the equations must be solved by computing a fixed point.) Given trace variable x , $\hat{\Gamma}_m(x)$ computes the least-upper bound ($\mathbb{T}.\sqcup$) of all variables in its equivalence class \hat{x} . However, this merging alone is insufficient for mimicking source typing, as it does not consider relationships between variables: if the program contains statement $x = y$, then the type of y influences the type of x in a source level type system.

$\hat{\Gamma}_p$ in Figure 6 handles type propagation across assignments. The first case handles assignments of values, using the baseline environment Γ_0 . For the second case, $x = y$, we generalize $\hat{\Gamma}(x)$ to include $\hat{\Gamma}(y)$. This step is important to mimic source-level handling of the assignment. The final two cases handle object field reads $b.p$. If $\hat{\Gamma}(b)$ is an object type with a property p , the handling is straightforward. However, this may not hold, due to other approximations introduced in computing $\hat{\Gamma}(b)$ (e.g., if p were dropped when merging object types due to width subtyping). If $\hat{\Gamma}(b).p$ does not exist, we fall back to $\Gamma_0(b).p$, i.e., we use the *precise* type of b . Without this treatment, type errors could propagate throughout the remainder of the trace, misleadingly inflating the type error count. Our fallback to the precise type of b helps to localize type errors, thereby allowing for gathering more useful information from the rest of the trace than the alternative of producing \top . For example, if x holds a number and later an object with property `foo`, FI merging will give it type \top . A read of `x.foo` would produce a type error. Rather than tainting all subsequent uses of the read result, we ascribe the type of the value observed as the result of that read.

Example Consider type propagation for the `a` variable in function `f` of Figure 1, using type system \mathbb{T}_{sub} from Section 2. The corresponding variables in Figure 5 are `<a@f_1>` and `<a@f_2>`, whose respective types are $\{p: \text{Number}\}$ and $\{p: \text{Number}, q: \text{String}\}$ in Γ_0 (`<y@G>`'s initialization (line 4 in Figure 1) is elided in Figure 5). Hence, with the width-subtyping-based merge mechanism $\mathbb{T}_{sub}.\sqcup$, both `<a@f_1>` and `<a@f_2>` are assigned type $\{p: \text{Number}\}$ in $\hat{\Gamma}$. The types in $\hat{\Gamma}$ for all variables in Figure 1 (using \mathbb{T}_{sub}) appear in Figure 7.

3.5 Type Checking

Once type propagation is done, type checking can be performed applying $\mathbb{T}.\text{check}$ to every statement in the trace, using $\hat{\Gamma}$ as the context. $\mathbb{T}.\text{check}$ should return, for each statement, the number of falsified antecedents for the corresponding type rule (returning 0 implies a well-typed statement).

Note that our type propagation biases the location of type errors to occur more often at reads instead of writes. Type propagation uses the $\mathbb{T}.\sqcup$ operator to find a variable type that handles all writes to that variable. This process may produce a type higher in the subtyping lattice than what the programmer may have intended, causing type errors to be pushed to uses of that variable.

Consider the example: `'x = {name: "Bob"}; x = {nam: "Bill"}; print(x.name);'` Most likely, the programmer intended the type of `x` to include a `name` field. If this type were declared, the type checker would produce an error at statement two, due to the misspelled `nam` field. With trace typing under the \mathbb{T}_{sub} type system, we will instead compute a type for

$$\begin{aligned}
\tau &::= \top \mid \perp \mid \text{number} \mid \text{boolean} \mid \text{string} \mid \mathbb{O} \mid \mathbb{F} \mid \bigcup \bar{\tau} \mid \mu X.\tau \\
\mathbb{O} &::= \text{PropName} \rightarrow \tau \\
\mathbb{F} &::= [\tau](\bar{\tau}) \rightarrow \tau \mid \bigwedge \bar{\mathbb{F}}
\end{aligned}$$

■ **Figure 8** Types used in the type systems we model.

$$\begin{aligned}
\text{shapeMap}(o) &\equiv \{(p_i, v_i) \mid o.p_i = v_i \text{ at some trace point}\} \\
\mathbb{T}.\alpha(o) &\equiv \{(p_i, \tau_i) \mid M = \text{shapeMap}(o) \wedge p_i \in \text{dom}(M) \wedge \tau_i = \mathbb{T}.\perp \mid \mathbb{T}.\alpha(v)\}_{v \in M[p_i]}
\end{aligned}$$

■ **Figure 9** Type ascription for objects.

x that has *no* fields by merging the types of the two objects — essentially over-generalizing due to the presence of what is well described as a static type error. Hence, the type error is pushed to the access of `name` at statement three, which is no longer in the type of x .

Intuitively, we want that when mimicking two comparable source-level type systems, trace typing ends up reporting fewer type errors for the more powerful one. Indeed, our results show that this is generally the case. However, this is not *guaranteed* to be the case because our ascription operation ($\mathbb{T}.\alpha$) can sometimes over-generalize, leading to unpredictable error counts. See Section 7 for further discussion of this matter.

4 Instantiations

In this section, we detail several instantiations of the trace typing framework of Section 3, showing its flexibility. We first describe a core type system that allows many of JavaScript’s dynamic object behaviors. We then show how this system can be used to detect possible tag tests in programs. Then, we present an extension of this system to test the usefulness of parametric polymorphism. Finally, we show an instantiation with a different, stricter type system for objects based on recent work [9], and how to evaluate various features of that system using trace typing.

4.1 Core Type System

The types we use for most of our experiments appear in Figure 8. The top (\top), bottom (\perp), primitive, union (\bigcup), and recursive types (μ) are standard. An object type (\mathbb{O}) is a map from property names to types. Function types (\mathbb{F}) are either standard (with receiver, argument, and return types), or an intersection type over function types. We use intersection types as a very precise model of call-return behavior, for a limit study of the usefulness of polymorphism (Section 6.2).

In our implementation, our object types are significantly more complex, mirroring the many roles objects play in JavaScript. In JavaScript, functions are themselves objects, and hence can have properties in addition to being invocable. Arrays are also objects with numeric properties and possibly non-numeric as well. Our implemented object types model these semantics, but we elide them here for simplicity. JavaScript also has prototype inheritance, where a read of property p from object o is delegated to o ’s prototype if p is not present on o . For the core type system, object types collapse this prototype chain, so unshadowed properties from the prototype parent are collapsed into the child object type (more discussion shortly). The type system of Section 4.4 deals with prototype inheritance more precisely.

Ascribing Object Types As noted in Section 3.1, defining the $\mathbb{T}.\alpha$ operation for object values is non-trivial. Since object properties may point to other objects, types for object values are inter-dependent. Further, JavaScript objects may be mutated in many ways: beyond changing property values, properties can also be added and deleted. Even the prototype of an object may be mutated.

Trace typing is currently limited to object types that do not change over time, i.e., only one type can be associated with each object instance during the entire execution. So, for an object o , $\mathbb{T}.\alpha(o)$ must merge together the different observed properties and property types for o into a single object type. Nearly all practical type systems use object types in this manner, as state-dependent object types require complex reasoning and restrictions around pointer aliasing.

Figure 9 gives our definition of $\mathbb{T}.\alpha$ for objects. We first define $shapeMap(o)$, a map from property names to sets of values (shown as a set of pairs). $shapeMap(o)$ captures all the property values that could be read from o at any point in the trace. (Note that we do not require the property read to actually exist in the trace.) Since JavaScript property reads may be delegated to the prototype object, $shapeMap(o)$ gives a “flattened” view of o that includes inherited properties. Consider the following example:

```

1 var y = {};
2 var x = { a: 4 } proto y; // prototype inheritance shorthand from [9]
3 x.b = { c: false };
4 y.d = "hello";
5 x.a = 10;

```

For this example, the shape map for the object allocated on line 2 is: $[a \mapsto \{4, 10\}, b \mapsto \{\{c: false\}\}, d \mapsto \{\text{"hello"}\}]$.

The object type $\mathbb{T}.\alpha(o)$ is written as a set of pairs in Figure 9. The properties in $\mathbb{T}.\alpha(o)$ are exactly those in the domain of $shapeMap(o)$. The type for property p_i is obtained by applying $\mathbb{T}.\alpha$ to each value of p_i in $shapeMap(o)$, and then combining the resulting types using $\mathbb{T}.\sqcup$. When computing $\mathbb{T}.\alpha$, cyclic data structures must be detected and handled by introducing a recursive type.

Type Checking The type rules we checked in our implemented type systems are standard. Property accesses can be checked normally, since the accessed property name is always evident in the trace. Function invocations are checked normally with one caveat: passing too few arguments is permitted, as all JavaScript functions are implicitly variadic.

For type systems with union types, we optimistically allow most operations that work for one case of the union (e.g., property access on a union typed expression) without validating that the code has checked for the appropriate case, but treat assignment of union-typed expressions to storage locations soundly (TypeScript and Flow have very similar rules for unions). Our trace language does not currently include conditionals, making it non-trivial to use tag tests to eliminate union types [20, 28]. However, our types do allow for discovering likely tag tests in the program, as we show below.

In the case when $\mathbb{T}.\sqcup$ causes a type to be \top , we count every use of and assignment to a variable of that type to be an error.

4.2 Detecting Tag Tests

To soundly eliminate union types, a type system must support narrowing the union under a conditional that checks some tag information of a value. While the theory for such narrowing is rich and well-developed [20, 28], industrial type systems for JavaScript (Flow and TypeScript) treat union elimination unsoundly in the same “optimistic” manner we do (Section 4.1). A

key issue is that tag tests can vary in complexity, from simple type checks (e.g., JavaScript’s `typeof` operator) to complex combinations of user-defined predicates, and it is unclear what level of complexity must be handled to provide adequate support for tag tests.

Using trace typing, we designed a technique to observe which tag test constructs are being used most frequently. As described previously in Section 2.2, the technique works by observing when there are two consecutive reads of a variable v (with no intervening write), and v ’s type is narrower at the second read than the first. This technique requires using the FS setting in $\mathbb{T}.\text{equiv}$, to keep separate types for each variable occurrence.⁴ For function types, we used $\mathbb{T}.\vec{\sqcup}$ from \mathbb{T}_{sub} , thereby generalizing using only subtype polymorphism. In Section 6.3, we discuss the kinds of tag tests discovered in our benchmarks with this technique and their prevalence.

4.3 Parametric Polymorphism

Here, we define $\mathbb{T}.\vec{\sqcup}_{poly}$, an instantiation of the $\mathbb{T}.\vec{\sqcup}$ operator that discovers types with parametric polymorphism. With this approach, one can evaluate the usefulness of adding parametric polymorphism to a retrofitted type system.

$\mathbb{T}.\vec{\sqcup}_{poly}$ works by enumerating all possible type parameter replacements in each invocation type for a function, and finally choosing the most general parameterized type matching all invocations. The enumeration proceeds by replacing all subsets of occurrences of a concrete type by a type variable. Consider the identity function, `function id(x){ return x; }`, and two invocations, `id(3)` and `id(true)`. The observed invocation types for `id` are `Number -> Number` and `Boolean -> Boolean`. (Applying $\mathbb{T}.\vec{\sqcup}$ from \mathbb{T}_{sub} to these types would yield the unhelpful type $(\top) \rightarrow \top$ for `id`.) For the first type, the enumeration would yield types `X -> Number`, `Number -> X`, and `X -> X` (along with the original type). After proceeding similarly for the second type, $\mathbb{T}.\vec{\sqcup}_{poly}$ returns the type `X -> X` for `id`, as it is the most general type matching both invocations.

In general, $\mathbb{T}.\vec{\sqcup}_{poly}$ also generates signatures with multiple type parameters. It also attempts to replace types nested one level into each argument or return type, to discover types operating over generic data structures. While this enumeration is exponential in the arity of the function, we did not observe an appreciable slowdown in practice.

After discovering polymorphic types, we must check the invocations against those types in $\mathbb{T}.\text{check}$. The checking can be done by treating each type variable as a direct subtype of \top , distinct from other type variables. Our implementation currently uses an alternate strategy of modifying the context-sensitivity policy in $\mathbb{T}.\text{equiv}$, so that variables in function invocations with distinct instantiations of the type parameters are not merged. This strategy corresponds to checking a more powerful form of bounded polymorphism, which may yield a reduced error count compared to directly checking unbounded type parameters.

4.4 Fixed Object Layout

We now show how to enhance the object types presented in Section 4.1 to support *fixed object layout*, as described in recent work by Choi et al. [9]. In general, JavaScript objects is used as dictionaries, with arbitrary addition and deletion of properties (keys). However, implementing all objects in this manner would lead to poor performance. Real-world programs *often* do not use this dictionary-like behavior. Objects often behave like records, in which the set of

⁴ Our implementation also disabled the assignment merging shown in Figure 6, to discover more tag tests.

properties is stable: after initialization, new properties are not added or removed. Modern just-in-time compilers detect the record usage pattern at run time, and use it to optimize the object representation where possible.

Choi et al. [9] show how to enforce the record usage pattern in a type system, enabling efficient static compilation of a JavaScript subset. Prototype inheritance complicates the design of such a type system, since a write of an inherited property creates a shadowing property in the child object. Consider the following example:

```

1 var o1 = {a:3, m:function(x) {this.a = x;}}
2 var o2 = {b:5} proto o1 // prototype inheritance shorthand from [9]
3 o2.a = o2.b; // adds field a to o2

```

The write at line 3 adds a local `a` property to `o2`, rather than updating the inherited property `o1.a`. The type system for fixed layout handles this case by distinguishing between *read-write* properties local to an object and *read-only* properties inherited from the prototype chain. With this distinction, the type system would reject the statement on line 3 above, as it is a write to a read-only property.⁵

To track read-only and read-write properties in object types, we extend $\mathbb{T}.\alpha$ from Section 4.1 as follows. Instead of tracking a single shape map for each object, we keep a read-write shape map for locally-declared properties, and a read-only shape map for inherited properties. In Section 4.1, the domain of $shapeMap(o)$ includes any p_i present on o at any point in the execution. Here, to enforce the fixed layout property, we restrict the domain of each shape map to properties present before the end of o 's initialization. (Recall from Section 3.2 that our traces include `end-initialization` statements to mark these points.) Given these two shape maps, computation of read-only and read-write properties in each object type proceeds as in Figure 9. Type checking is enhanced to ensure that property writes are only performed on read-write properties.

We also check for two additional properties from the Choi et al. type system [9]. The type system requires any object used as a prototype parent to have a *precise* type, i.e., no properties can have been erased from the type using width subtyping. (This is required to handle subtle cases with inheritance [9].) In trace typing, we check this property by only flagging types of object literals as precise, so if width subtyping is ever applied, the resulting type is not flagged. Finally, the type system requires that if a property from a prototype is shadowed, the parent and child properties have the same type. This condition is easily checked in $\mathbb{T}.\alpha$ once read-only and read-write properties are computed.

5 Implementation

We implemented the trace typing process described in Section 3 as a tool for typing JavaScript programs. This section reports on some of the more noteworthy aspects of the implementation.

Tool Architecture Our tool consists of two components: Trace Collector and Trace Typer. Trace Collector is a Jalangi [26] analysis that obtains a trace of a program of interest by monitoring the execution using source-level instrumentation; it consists of about 2500 lines of JavaScript code. Trace Typer implements the core trace typing framework. The framework and the type systems described in Section 4 are implemented in about 4000 lines of TypeScript, of which about 1000 lines are the type system implementations.

⁵ The type system of Choi et al. also tracks which properties are written on method receivers, to locally check validity of inheritance [9]. With trace typing, this information need not be tracked explicitly, since receiver types are propagated into method calls.

Modeling JavaScript Semantics In our implementation, great care is taken to accurately model JavaScript’s complex primitive operators. Much of the complexity in operator semantics lies in the implicit conversions performed on operands. For example, the binary `&&` operator first coerces each of its operands to a boolean value; this coercion can be applied to a value of any type. However, the value of the `&&` expression is always one of the *uncoerced* operands. The implementation creates trace statements that models these conversions explicitly, enabling accurate data flow tracking and type ascription.

Modeling the Native Environment To handle interactions with the native environment (e.g., built-in JavaScript library routines), we require a model of the native behavior. We express native models using trace statements, avoiding the need for a separate modeling language. In many cases, we can actually infer native models from the concrete states of the execution, thereby avoiding the significant work of writing models by hand. Model inference works by associating a new global *escape variable* with each object that can escape to the native environment. Passing an object to the native environment is modeled as a write to the corresponding escape variable, while retrieving the object from the environment is modeled as a read. With this handling, a model can be automatically inferred for much of the native environment, including complex functions such as the overloaded Array constructor. Our technical report [4] provides more details.

6 Experiments

We report on several experiments we conducted using trace typing. First, we studied the trends in error counts for six type system variants, generated by varying handling of subtyping and function types (Section 6.2), yielding useful insights on the relative importance of these variants. Second, we used trace typing to discover tag tests in our benchmark and characterize their relative frequency and complexity (Section 6.3). Finally, we studied several questions around the restrictiveness in practice of a type system for fixed object layout (Section 6.4). Together, these experiments show the usefulness and versatility of the trace typing framework. We first describe our benchmarks, and then present the experiments.

6.1 Benchmarks

We use nine popular packages from `npm`, the node.js package manager, as benchmark programs. Figure 10 measures the static source code size of each program,⁶ as well as the number of statements in the traces we generate and the number of static source lines covered by our executions. `underscore` and `lazy.js` are utility libraries. `esprima` and `escodegen` are a JavaScript parser and code generator, respectively. `typescript` is the compiler for the TypeScript language.⁷ `minimist`, `optparse` and `qs` are parsers for command line options and query strings. `xml2js` is an XML serializer/deserializer.

These benchmarks capture a range of interesting JavaScript styles. The parsers, code generators, and compilers manipulate OO-style AST data structures. In contrast, the utility libraries and option parsers are highly reflective (e.g., constructing or extending objects using dynamically-computed property names), but are otherwise written in a largely functional style. Hence, the suite as a whole exercises a variety of features of trace typing.

⁶ Non-comment, non-blank lines as counted by `cloc`, excluding build scripts, tests and test harness code.

⁷ Figure 10 gives lines of JavaScript source (the compiler is bootstrapped).

Benchmark	Benchmark sizes			Fixed-object-layout error rates		
	LOC	LOC exec	Length	ro/rw	prototypal	inheritance
<code>escodegen</code>	2132	723	375325	0/53	0/0	0/217
<code>esprima</code>	4610	1052	23986	0/44	0/2	0/298
<code>lazy.js</code>	2557	1016	25439	114/387	0/17	2/470
<code>minimist</code>	186	151	140812	0/14	0/0	0/81
<code>optparse</code>	222	141	15246	0/15	0/1	0/51
<code>qs</code>	726	256	102637	0/41	0/0	0/62
<code>typescript</code>	35456	8139	167730	1/2085	2/3	0/3286
<code>underscore</code>	1098	727	120081	2/175	0/0	1/294
<code>xml2js</code>	840	275	73672	0/42	1/2	0/91
Total	52193	14541	1243857	117/2856	3/25	3/4850

■ **Figure 10** Benchmarks sizes and error rates for the fixed-object-layout type system. ‘LOC’ is the lines of code in the program. ‘LOC exec’ is the lines covered in our execution of the program. ‘Length’ is the total number of trace statements. ‘ro/rw’, ‘prototypal’ and ‘inheritance’ are the error-rates for different each kind of type check.

To exercise the programs for trace generation, we wrote a separate driver for each benchmark. For the option parsers and `escodegen`, this driver comprised extracted tool invocations from their test suites. We exercised the utility libraries with code snippets from their official tutorials.

6.2 Comparing Type System Variants

For our first experiment, we compared six type systems, generated by varying the $\mathbb{T}.\sqcup$ and $\mathbb{T}.\vec{\sqcup}$ operators over the core types of Section 4.1. For $\mathbb{T}.\sqcup$, we used two options:

- **subtyping**: This uses standard structural (width) subtyping for objects, as in \mathbb{T}_{sub} from Section 2.1 (see also Example 1 in Section 3.1).
- **unions**: $\mathbb{T}.\sqcup$ treats two object types as in **Subtyping**, but also introduces a union type when applied to an object type and a primitive (see Example 2 in Section 3.1).

For $\mathbb{T}.\vec{\sqcup}$, we use three options:

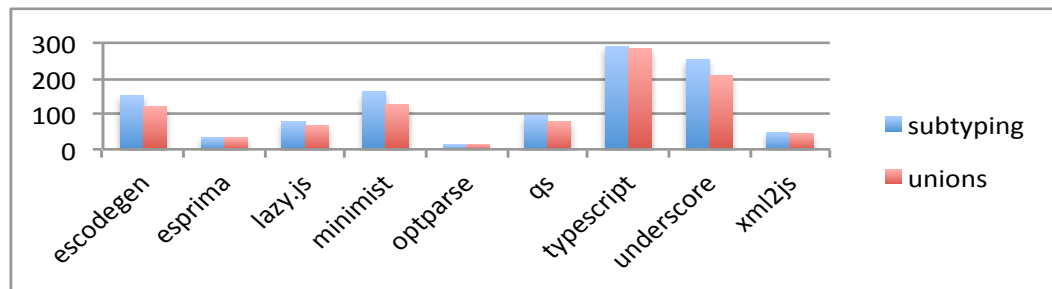
- **base**: $\mathbb{T}.\sqcup$ is applied individually to argument and return types, as discussed in Section 3.1.
- **poly**: This uses $\mathbb{T}.\vec{\sqcup}_{poly}$ to discover parametric polymorphism, as discussed in Section 4.3.
- **intersect**: $\mathbb{T}.\sqcup$ simply combines two function types into an intersection of the two types. This strategy yields more precise types than any practical type system, and hence is only used as a limit study of the usefulness of polymorphic function types.

$\mathbb{T}.\text{equiv}$ uses FI for flow-insensitive variable types in all configurations. Context sensitivity is varied in accordance with $\mathbb{T}.\vec{\sqcup}$: for **base**, we use CI, while for **intersect** we use CS (since each invocation may have its own case in the intersection type). Context sensitivity for **poly** was discussed in Section 4.3.

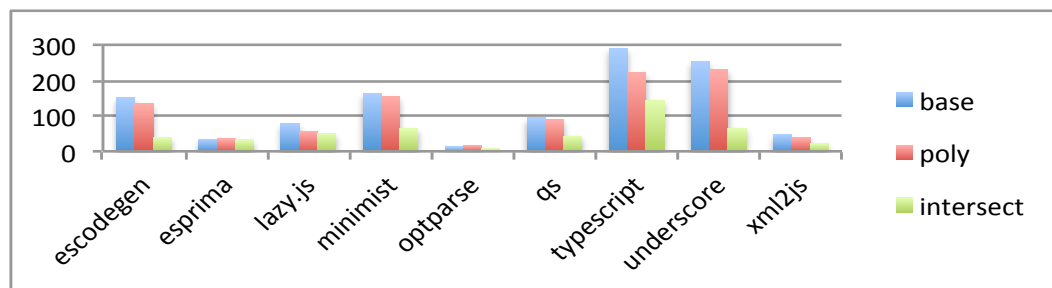
After collecting traces, we ran our trace typer with each of our type system variants, counting the number of static (source location) type errors that arose. We include only type errors that occur in the subject program, not in the libraries they depend on.

Experiments were run on a quad-core Intel Core i7-3520M@2.90GHz with 16GB RAM, running Ubuntu 15.10 and node.js 5.0. Collecting traces and running all six configurations required less than 15 minutes.

Our full data set is available [4]. We present two select slices of data here. Figure 11 presents the number of syntactic locations with errors for **subtyping** and **unions**, where $\mathbb{T}.\vec{\sqcup}$



■ **Figure 11** Error counts for **subtyping** and **unions**, selecting **base** for $\mathbb{T}.\vec{\square}$.



■ **Figure 12** Error counts for different $\mathbb{T}.\vec{\square}$ settings, selecting **subtyping** for $\mathbb{T}.\square$.

is **base** in either case. Figure 12 shows error counts for the three $\mathbb{T}.\vec{\square}$ settings, where $\mathbb{T}.\square$ is **subtyping** in all three cases. The trends shown here are similar across other configurations.

Overall we see trends we would expect, with more powerful type systems yielding fewer errors. There are two small inversions in the results, where a more precise configuration (**poly**) produces more errors than a less precise configuration (**base**) discussed shortly.

This experiment let us find *quantitative* answers to the kinds of questions we raised in Section 1 and Section 2. We remind the reader that the objective of these experiments is not to pass verdict on the usefulness of certain type system features in an absolute sense. Rather the objective is to help designers of retrofitted type systems prioritize features based on empirical evidence rather than by intuition alone. If a type system feature ends up unsupported, then users will need to work around that limitation in their code (or live with type errors).

How prevalent is the need for union types? While Figure 11 shows that union types do reduce error counts over just using subtyping (as expected, since unions are more powerful), the degree of reduction is relatively small. While TypeScript and Flow now both include unions, TypeScript did not add them until two years after its initial release. The small error count reduction due to union types in our experiments could help to explain why they were not needed in TypeScript from the very beginning.⁸

How prevalent is the need for parametric polymorphism? For intersection types? As shown in Figure 12, the **intersection** configuration always reduced errors compared to **poly**, while **poly** almost always reduces errors compared to **base**⁹. While the error count reductions for **poly** are not dramatic in general (**typescript** is discussed below), parametric polymorphism is crucial for ascribing types to core routines manipulating data structures

⁸ TypeScript's issue tracker has further discussion of their motivation for adding unions; see <https://github.com/Microsoft/TypeScript/issues/805>.

⁹ The inversions in **esprima** and **optparse** are due mostly to implementation quirks, such as an incomplete merge operator on recursive types that returns \top more than necessary.

(e.g., array routines), and hence its inclusion in TypeScript and Flow is unsurprising. The dramatic drop in error count for several programs under the **intersection** configuration gives strong evidence that JavaScript requires types that can express non-uniform polymorphism, such as intersection types. In fact, TypeScript has long supported overloaded functions for expressing intersection types.¹⁰

The `typescript` benchmark stands out from the others as it was originally implemented in TypeScript, and hence we would expect it to be mostly well typed. Our results confirm this: the error count is only 290 in 8139 (JavaScript) lines of code for the least precise configuration, a much lower error rate than the other programs. Also, the significant reduction in errors going from **base** to **poly** for the benchmark (see Figure 12) corresponds nicely to the use of parametric polymorphism in the TypeScript code base; see Section 7 for an example.

While in the above experiment we used trace typing to retroactively find evidence of the need of type system features, the next two experiments are concerned with more exploratory questions.

6.3 Finding Tag Tests

Here, we provide the first empirical analysis of the use of tag tests in JavaScript which identifies tag tests based on observed narrowing (as described in Sections 2.2 and 4.2), rather than syntactic criteria. We post-process the ascription results to automatically identify occurrences of guard predicates using the techniques of Section 4.2. We then manually analyze the predicates to identify non-local guards (where the boolean result of a check is computed outside a syntactic conditional statement), non-atomic guards (i.e., conjunction and disjunction of guards), and predicate functions (other functions whose result indicates a certain refinement).

Methodology As mentioned in Section 4.2, we detect tag tests by observing consecutive reads of a variable where the type is narrowed at the second read. Usually, the first read of this pair will be the use in the type guard. In cases where the guard expression is saved to a local variable before being checked, this technique detected spurious additional guards, but they were easy to recognize and discard manually.

We ran the set of benchmark programs from Section 6.1 with the union-producing `T.⊔` operator of Section 6.2, and a flow sensitive `T.equiv` policy to keep types of variable occurrences under conditionals separated.

Results It took 90 minutes to manually classify all of the discovered type guards, the results of which can be seen in Figure 13. The figure contains rows for different kinds of type guards, if a type is non-atomic, it can be classified as having multiple kinds. Due to their prevalence, extra rows have been added for type guards that are applied on the property of the guarded value.

Overall there are 6 major classes of narrowing checks: `typeof`, `instanceof`, checking an explicit tag field in a data structure¹¹ (mostly in the AST manipulations in our corpus), a general predicate function, explicit property presence checks (`hasOwnProperty`), and checks of objects' `prototype` or `constructor` fields. The most prevalent type guards are `typeof`, use of general predicate functions, and (a variety of) property presence tests. Checking an explicit tag field, used to mimic algebraic datatypes in JavaScript, are common but highly

¹⁰ See <http://www.typescriptlang.org/Handbook#functions-overloads>.

¹¹ A popular requested TypeScript feature: <https://github.com/Microsoft/TypeScript/issues/186>

Measure	Count	Example
Type guards	164	-
Non-local	6	<code>var isObj = typeof x == 'object'; if(isObj)</code>
Non-atomic	46	<code>if(typeof x == 'string' typeof x == 'object')</code>
typeof	30	<code>typeof x === 'function'</code>
instanceof	9	<code>x instanceof Function</code>
tag field	16	<code>x.kind = Kinds.Template</code>
predicate function	30	<code>Array.isArray(x)</code>
property presence	35	<code>x.hasOwnProperty('prop'), x.prop === undefined</code>
.prototype/.constructor	3	<code>x.constructor === Function</code>
.. test on property	37	<code>typeof x.prop === 'function'</code>
other	23	<code>className === toString.call(x)</code>

■ **Figure 13** Tag test classification results.

program-dependent. Non-local type guards are rare, and would be easy to rewrite if needed. Boolean combination of type guards is common enough that they probably should not be ignored by a sound union elimination check.

6.4 Evaluating Fixed Object Layout

We present a case study of evaluating Choi et al.’s type system [9] to enforce *fixed object layout* in JavaScript, as described in Section 4.4. We checked the following salient features in the trace typing framework:

1. *Property writes* We checked that property writes were only performed on properties deemed to be “read-write” properties in the type system.
2. *Precise types* We checked that only objects with a “precise” type (see Section 4.4) were used as a prototype parent for other objects.
3. *Consistent shadowing* We checked that when properties were shadowed in the prototype chain, the parent and child properties had the same type.

We ran trace typing with this type system with the benchmark suite of Section 6.1 (with the subtyping `T.equiv` and `base` policy from Section 6.2, merging all occurrences of a variable in a dynamic function invocation). Figure 10 shows the errors reported by our trace typing model. The column *ro/rw* shows, in absolute numbers, how many of the property writes were to fields that were inherited from prototype chain (and not shadowed by a local redefinition), of the total number of property writes. The column *prototypal* shows, how many of the assignments to some object’s prototype property was in violation of the restriction mentioned (second bullet) above, vs total number of such assignments. The column *inheritance* shows in how many cases, a property was shadowed with an inconsistent type.

The results add significant confidence to our initial beliefs that some of the restrictions the type system imposes are not onerous in practice: the numerator numbers are generally (though not universally) quite small compare to the denominator. We manually found that most of the *ro/rw* type errors for `lazy.js` were due to the way prototype hierarchies were initialized, and that the type errors could be eliminated by a simple refactoring.

Additionally, from the trace typing data (not presented here) we found that support for *optional properties* in the type system could have helped eliminate a large number of type errors, and so would be a very useful feature to add to the type system. Trace typing

```
// TypeScript/src/compiler/parser.ts:39
export function forEachChild<T>(node:Node, ...):T {
  ... switch (node.kind) {
    ... case SyntaxKind.PropertySignature:
      ... return visitNode(cbNode, (<VarLikeDecl>node).propName);
  }
}
// TypeScript/src/compiler/type.ts:504
export interface VarLikeDecl extends Declaration {
  propName?: Identifier; \n dotDotDotToken?: Node;
  name: DeclarationName; \n questionToken?: Node;
  type?: TypeNode; \n initializer?: Expression;
}
```

■ **Figure 14** Code fragment from `typescript`.

finds this information without the need for implementing an actual static type inferencer for JavaScript that is robust enough to work on large code bases.

7 Discussion

In this section, we discuss the quality of types inferred by trace typing, present limitations and threats to validity, and then mention possible future generalizations of our approach.

Quality of Types Inferred Because our type inference approach is non-standard and our error counts are sensitive to over-generalization (Section 3.5), it is worth asking whether the types we infer are sensible — whether a human developer would write similar types. While there is no systematic way to verify this, our experience says that the types are sensible. In the course of developing the framework and type system plugins, we spent a significant amount of time examining the types inferred. Object (property) types rarely deviated from what we would have written ourselves.

We discuss one specific example of types inferred by trace typing for the `typescript` benchmark, for which the original TypeScript code contains declared types (trace typing analyzes the transpiled JavaScript version). Figure 14 contains an excerpt from the TypeScript code. At the `return` statement the `node` variable always has type `VarLikeDecl` (also shown in the figure). Trace typing finds the relevant properties that we see mentioned in the interface, as well as the inherited properties. Moreover, it also correctly computes the structural type of each property of `VarLikeDecl` such as `initializer`. Due to space reasons, we do not show all the inferred types; but we found them reasonable on manual inspection.

Our ascription of parametric polymorphic types (Section 4.3) also works well in practice. For example, it computes these polymorphic signatures for array methods (from the native environment):

```
Array<E>.prototype.indexOf: (E) -> Number
Array<E>.prototype.concat: (Array<E>) -> Array<E>
Array<E>.prototype.push: (E) -> Number
Array<E>.prototype.pop: () -> E
```

These types are similar to the types used for native environment by TypeScript programs.¹²

The exceptions to the above arise mostly where type errors occur, and in those cases the types are not bizarre but simply a different choice among several incompatible options (e.g., when a method is overridden with an incompatible type). For the `intersect` configuration

¹²<https://github.com/Microsoft/TypeScript/blob/v1.7.3/lib/lib.core.d.ts#L1008>

of Section 6.2, with unbounded intersection types, the gap between inferred types and what one would write at the source level is larger, due to the synthetic nature of the type system. Designing a trace typing system that infers a mix of parametric polymorphism and intersection types closer to what a person would write is future work.

Limitations and Threats to Validity Trace typing works under the assumption that the input programs are mostly type correct; the error locations generated by our framework lose accuracy when this assumption is violated. Moreover, we assume that most parts of a program *have a correct type* in the system being modeled, which may not always hold. Our manual inspection of inferred types suggests that the natural formulations for $\mathbb{T}.\alpha$ and $\mathbb{T}.\sqcup$ generally behave well, but it is possible other pathological cases exist.

A result that type system design *A* yields fewer type errors than design *B* in our framework does *not* strictly imply that the same will hold for complete source-level implementations, due to inevitable over-approximations in a static type checker. Constructs inducing such approximations include reflective constructs like dynamic property accesses (which are resolved in our trace) and loops. Type system feature interactions can exacerbate this effect. The best use of trace typing is to compare type systems that differ in only one or two parameters, and to give less weight to very small differences in the number of errors detected. Consider this example:

```

1 function f(x) { return x.p; }
2 f({p:3, q:4});
3 f({p:3, r:true});
4 var y = {p:3};
5 y = {p:3, q:5};
6 f(y);

```

Here, type ascription's over-generalizing in the presence of what would be static type errors can lead to some unexpected results. For the example above, ascription in the **base** polymorphism configuration (see Section 6.2) with no union types would produce the type $\{p:\text{Number}\} \rightarrow \text{Number}$ for *f*, which would allow the trace of the program to type check. Ascription with the more precise **intersect** configuration would give the intersection type $\{p:\text{Number}, q:\text{Number}\} \rightarrow \text{Number} \wedge \{p:\text{Number}, r:\text{Boolean}\} \rightarrow \text{Number}$. But, the two writes to *y* ensures that it is ascribed type $\{p:\text{Number}\}$, which is not a subtype of either argument type in the intersection! In such cases, a more powerful trace type system can produce more errors than a less powerful one. We did not observe this issue in experimental results we inspected manually, but it may have occurred elsewhere.

Our work is an empirical study using dynamic analysis, and we inherit the normal threats to validity from such an approach, namely possible sampling bias in the subject programs and sampling bias from the inputs used to gather traces. For our study, we took inputs from each program's often-substantial test suite, which appeared from manual inspection to give reasonable path and value coverage.

Potential Extensions We have focused thus far on type systems for JavaScript, but the core ideas of our technique apply to other languages as well. The generalization to other dynamically typed languages (e.g., Ruby, Python) would require instrumentation of the other language. The type system implementations on top of the new framework would need to be tailored to the primitives and idioms of the new target language.

Our existing trace format could be used for additional experiments that analyze the types ascribed to specific parts of the program. For example, this could inform strategies for typing challenging constructs like computed property accesses [24].

More metadata could be added to our traces to enable experimentation with other type-system features, with no changes to our core approach. For example, more detailed information about the evolving types and layouts of objects could be maintained, to enable experimentation with alias types [10, 27]. Our hope is that trace typing will be a useful base for such future extensions.

8 Related work

We focus here primarily on approaches to retrofitting type systems onto existing languages and on other means of evaluating type systems on large bodies of code, with a secondary focus on type systems for JavaScript.

Retrofitting Type Systems Retrofitting type systems onto existing languages is not new,¹³ nor is it restricted to dynamic languages. The notion of *soft typing* for dynamically typed languages is now well-established [7, 28, 31]. The notion of *pluggable type systems* — including extensions to existing type systems — is also an established idea [6], which continues to produce useful tools [12, 23]. Both bodies of work suffer the same difficulty in evaluating lost precision: it requires tremendous manual effort to evaluate on non-trivial amounts of code, leading to design decisions based on intuition and a handful of carefully-chosen examples.

More recently, two groups of researchers arrived at means to induce a gradual type system variant of a given base type system, one as a methodology to produce a gradual type system by careful transformation of a static type system [14] and another as a tool to compute a gradual version of a type system expressed as a logic program [14]. Both deterministically produce one gradual type system from one static type system, aiding with the introduction of gradual typing, but not of overall design.

Evaluating Type Systems The implementation effort for prototyping similar but slightly different type systems and applying them is a significant barrier to evaluating completely new type systems for existing programming languages. TEJAS [21] attempts to remedy this for JavaScript type systems by carefully architecting a modular framework for implementing type systems. It relies on bidirectional typing to reduce annotation burden, but this still requires significant manual annotation.

A smaller but closely-related body of work is evaluating changes to existing type systems. Wright studied approximately 250,000 lines of SML code [30], cataloguing the changes necessary to existing code in order to type-check under the value-restriction to type generalization. The study in this case required less implementation effort than the general approach we propose; SML implementations existed, and could be modified to test the single alternative type system change. Nonetheless, his careful analysis of the impact of the proposed change justified a type system change that has withstood the test of time. SML still uses the value restriction, and OCaml uses a mild relaxation of it [15]. More recently Greenman et al. [18] performed a similar evaluation of an alternative proposal for F-bounded polymorphism in Java.

Dynamic Type Inference Type inference for dynamic languages is another rich area with strong connections to our trace-based type inference. The closest such work to ours is dynamic type inference for Ruby [2]. They use Ruby’s reflection capabilities to pass virtualized values through a program, with each use of the virtualized objects gathering constraints on the

¹³The use of the term *retrofitting* is relatively recent, due to Lerner et al. [21].

type of the object. Their system has a guarantee that the inferred types will be sound if every path within each procedure is executed at least once. Like trace typing, they rely on dynamic information to infer types for program fragments, including for complex source-level constructs. Unlike our approach, they make no attempt to simulate the effect of type *checking* the program: they infer types that hold at method entry and exit, but do not infer types for local variables, whose types may change arbitrarily. Also, they cannot *infer* method polymorphism (though they support annotations), require manual annotation for native code, and evaluate only a single type system design.

Saftoiu et al. [25] implement dynamic type inference for JavaScript, generating types for a particular JavaScript type system based on dynamic observation, and find that it is useful for converting code to their typed JavaScript dialect. Their type system is sound with respect to λ_{JS} [19], and they can type-check the source code to verify the inferred types. We cannot do this, because our goal is experimenting with type system design, so our type systems are not adequate to type check source programs. Like us they instrument code via source-to-source translation. We evaluate significantly more code: their examples total 3799 lines of code.

Other Uses of Traces Coughlin et al. [11] use dynamic traces to identify the spans between program safety checks and data uses guarded by those checks, e.g., a null check followed by a dereference. These measurements can inform the design of static analyses, e.g., by indicating whether or not an analysis should be interprocedural. While similar in philosophy to our work, we differ substantially in technical details. First, their implementation is specific to null dereference errors. Second, their goal is to *measure* aspects of program behavior quantitatively, and interpret that measurement when designing an analysis; we instead advocate directly comparing approximations of a real analysis.

Åkerblom and Wrigstad [1] study the receiver polymorphism of a very large base of open source Python programs, primarily to examine how adequate nominal subtyping might be for Python (as opposed to structural subtyping). They classify static call sites by whether they were observed to have multiple receiver types, consider whether call sites might be dispatched on parametrically-polymorphic targets by clustering call sites in different ways (roughly similar to our context-sensitivity policies), and do some reasoning about cases where different static calls on the same syntactic receiver may have different qualities (a coarse approximation of typing local variables). They do not reason about actual method types, argument types (except implicitly if methods are invoked on arguments), and do not attempt to simulate source type checking as we do. ECMAScript 5 — the previous version of JavaScript — lacks classes, so even code written to ECMAScript 6 — which added classes — will generally require structural subtyping. Our type ascription machinery could be used in a style similar to Section 6.3 to implement most of their instrumentation.

Type Systems for JavaScript Most previous work on static typing for JavaScript focuses on novel approaches to typing its most expressive features; we describe particularly relevant examples here. Guha et al.’s flow typing [20] allows for refining types in code guarded by runtime tag checks. DJS [10] implements a sophisticated dependent type system that includes a type-based analogue [27] of separation logic to describe heap changes, and an SMT-based refinement type system for characterizing the shape of objects. TeJaS [21] builds an expressive framework for experimenting with JavaScript type systems, whose core is an extension to $F_{<}^{\omega}$.

Throughout we have discussed Flow [13] and TypeScript [5, 22]. They are quite similar (objects, unions with unsound elimination rules, intersections) with a few small points of

divergence (singleton string types and intentional unsound subtyping in TypeScript, non-nullable fields in Flow). They have very different approaches to type inference and checking: TypeScript uses limited local type inference, while Flow performs a global (per-module) data flow analysis. We have already modeled key features — objects with structural subtyping, union and intersection types — in our framework.

TypeScript and Flow include forms of parametric polymorphism, based on programmer annotations. Trace typing can approximate some forms of parametric polymorphism (Section 3.4). Both TypeScript and Flow also allow some narrowing of union types based on tag tests (unsound in general due to aliasing and mutation). Trace typing could be extended to handle such features in future work, as discussed in Section 7.

Both TypeScript and Flow have added type system features based on user feedback, with great success. Trace typing provides an alternate, complementary source of information for deciding which type system features to include, which can be employed without having a large user base.

9 Conclusions

We presented a framework for quantitatively evaluating variations of a retrofitted type system on large code bases. Our approach involves gathering traces of program executions, inferring types for instances of variables and expressions occurring in a trace, merging types according to merge strategies that reflect options in the source-level type system design space, and type-checking the trace. By leveraging the simple structure of traces, we dramatically reduce the effort required to get useful feedback on the relative benefit of type system features.

To evaluate the framework, we considered six variations of a type system retrofitted onto JavaScript. In each case, we measured the number of type errors reported for a set of traces gathered from nine JavaScript applications (over 50KLOC) written in a variety of programming styles. The data offers quantitative validation for some of the design choices in two popular retrofitted type systems for JavaScript (Typescript and Flow). In a different experiment, we used the results of trace typing to automatically identify places where type narrowing occurred dynamically, to gather empirical results on the frequency and variety of tag tests in JavaScript. In yet another experiment, we evaluated how onerous the restrictions for a new retrofitted JavaScript type system [9] were, validating intuitions that our restrictions were mostly reasonable, and identifying priorities for future extensions.

The feasibility of carrying out these experiments is a strong validation of the trace typing approach.

References

- 1 Beatrice Åkerblom and Tobias Wrigstad. Measuring polymorphism in Python programs. In *Proceedings of the 11th Symposium on Dynamic Languages (DLS)*, 2015.
- 2 Jong-hoon David An, Avik Chaudhuri, Jeffrey S Foster, and Michael Hicks. Dynamic Inference of Static Types for Ruby. In *POPL*, 2011.
- 3 Chris Andreae, James Noble, Shane Markstrum, and Todd Millstein. A Framework for Implementing Pluggable Type Systems. In *OOPSLA*, 2006.
- 4 Esben Andreasen, Colin S. Gordon, Satish Chandra, Manu Sridharan, Frank Tip, and Koushik Sen. Trace Typing: An Approach for Evaluating Retrofitted Type Systems (Extended Version). Technical Report SRA-CSIC-2016-001, Samsung Research America, 2016. URL: <https://arxiv.org/abs/1605.01362>.
- 5 Gavin Bierman, Martín Abadi, and Mads Torgersen. Understanding TypeScript. In *ECOOP*, 2014.

- 6 Gilad Bracha. Pluggable Type Systems. In *OOPSLA Workshop on Revival of Dynamic Languages*, 2004.
- 7 Robert Cartwright and Mike Fagan. Soft Typing. In *PLDI*, 1991.
- 8 Avik Chaudhuri. Bounded Polymorphism. URL: <http://flowtype.org/blog/2015/03/12/Bounded-Polymorphism.html>.
- 9 Philip Wontae Choi, Satish Chandra, George Necula, and Koushik Sen. SJS: A Type System for JavaScript with Fixed Object Layout. In *SAS*, 2015.
- 10 Ravi Chugh, David Herman, and Ranjit Jhala. Dependent Types for JavaScript. In *OOPSLA*, 2012.
- 11 Devin Coughlin, Bor-Yuh Evan Chang, Amer Diwan, and Jeremy G. Siek. Measuring enforcement windows with symbolic trace interpretation: What well-behaved programs say. In *ISSTA*, 2012.
- 12 Werner Dietl, Stephanie Dietzel, Michael D Ernst, Kivanç Muşlu, and Todd W Schiller. Building and Using Pluggable Type-Checkers. In *ICSE*, 2011.
- 13 Facebook. Flow: A Static Type Checker for JavaScript. <http://flowtype.org/>.
- 14 Ronald Garcia, Alison M. Clark, and Éric Tanter. Abstracting Gradual Typing. In *POPL*, 2016.
- 15 Jacques Garrigue. Relaxing the Value Restriction. In *Functional and Logic Programming*, 2004.
- 16 Google. Closure. <https://developers.google.com/closure/>.
- 17 Colin S Gordon, Werner Dietl, Michael D Ernst, and Dan Grossman. JavaUI: Effects for Controlling UI Object Access. In *ECOOP*, 2013.
- 18 Ben Greenman, Fabian Muehlboeck, and Ross Tate. Getting F-bounded Polymorphism into Shape. In *PLDI*, 2014.
- 19 Arjun Guha, Claudiu Saftoiu, and Shriram Krishnamurthi. The Essence of JavaScript. In *ECOOP*, 2010.
- 20 Arjun Guha, Claudiu Saftoiu, and Shriram Krishnamurthi. Typing Local Control and State Using Flow Analysis. In *ESOP*, 2011.
- 21 Benjamin S Lerner, Joe Gibbs Politz, Arjun Guha, and Shriram Krishnamurthi. TeJaS: Retrofitting Type Systems for JavaScript. In *DLS*, 2013.
- 22 Microsoft. TypeScript Handbook. <http://www.typescriptlang.org/Handbook>.
- 23 Matthew M Papi, Mahmood Ali, Telmo Luis Correa Jr, Jeff H Perkins, and Michael D Ernst. Practical Pluggable Types for Java. In *ISSTA*, 2008.
- 24 Joe Gibbs Politz, Arjun Guha, and Shriram Krishnamurthi. Semantics and Types for Objects with First-Class Member Names. In *FOOL*, 2012.
- 25 Claudiu Saftoiu, Arjun Guha, and Shriram Krishnamurthi. Runtime Type-Discovery for JavaScript. Technical Report CS-10-05, Brown University, 2010.
- 26 Koushik Sen, Swaroop Kalasapur, Tasneem Brutch, and Simon Gibbs. Jalangi: A selective record-replay and dynamic analysis framework for JavaScript. In *FSE*, 2013.
- 27 Frederick Smith, David Walker, and Greg Morrisett. Alias Types. In *ESOP*, 2000.
- 28 Sam Tobin-Hochstadt and Matthias Felleisen. Logical Types for Untyped Languages. In *ICFP*, 2010.
- 29 Jonathan Turner. Announcing TypeScript 1.4. URL: <http://blogs.msdn.com/b/typescript/archive/2015/01/16/announcing-typescript-1-4.aspx>.
- 30 Andrew K. Wright. Simple Imperative Polymorphism. *LISP and Symbolic Computation*, 8(4):343–355, 1995.
- 31 Andrew K. Wright and Robert Cartwright. A Practical Soft Type System for Scheme. *ACM TOPLAS*, 19(1):87–152, January 1997. doi:10.1145/239912.239917.
- 32 Haiping Zhao, Iain Proctor, Minghui Yang, Xin Qi, Mark Williams, Qi Gao, Guilherme Ottoni, Andrew Paroski, Scott MacVicar, Jason Evans, and Stephen Tu. The HipHop Compiler for PHP. In *OOPSLA*, 2012.