KafKa: Gradual Typing for Objects

Benjamin Chung, Paley Li, Francesco Zappa Nardelli, Jan Vitek

▶ To cite this version:

Benjamin Chung, Paley Li, Francesco Zappa Nardelli, Jan Vitek. KafKa: Gradual Typing for Objects. ECOOP 2018 - 2018 European Conference on Object-Oriented Programming, Jul 2018, Amsterdam, Netherlands. hal-01882148

HAL Id: hal-01882148

https://inria.hal.science/hal-01882148

Submitted on 26 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

KafKa: Gradual Typing for Objects

Benjamin Chung
Northeastern University, Boston, MA, USA

Paley Li
Czech Technical University, Prague, Czech Republic, and
Northeastern University, Boston, MA, USA

Francesco Zappa Nardelli
Inria, Paris, France, and
Northeastern University, Boston, MA, USA

Jan Vitek
Czech Technical University, Prague, Czech Republic, and
Northeastern University, Boston, MA, USA

Abstract

A wide range of gradual type systems have been proposed, providing many languages with the ability to mix typed and untyped code. However, hiding under language details, these gradual type systems embody fundamentally different ideas of what it means to be well-typed.

In this paper, we show that four of the most common gradual type systems provide distinct guarantees, and we give a formal framework for comparing gradual type systems for object-oriented languages. First, we show that the different gradual type systems are practically distinguishable via a three-part litmus test. We present a formal framework for defining and comparing gradual type systems. Within this framework, different gradual type systems become translations between a common source and target language, allowing for direct comparison of semantics and guarantees.

2012 ACM Subject Classification Software and its engineering → Semantics

Keywords and phrases Gradual typing, object-orientation, language design, type systems

Digital Object Identifier 10.4230/LIPIcs.ECOOP.2018.12

Supplement Material ECOOP Artifact Evaluation approved artifact available at
`http://dx.doi.org/10.4230/DARTS.4.3.10`

Funding This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 695412), the NSF (award 1544542 and award 1518844) as well as ONR (award 503353).

Acknowledgements The author thank the reviewers of ECOOP, POPL, ESOP, and again ECOOP for comments that gradually improved this paper.
We are grateful to Leif Andersen, Fabian Muelbrock, Éric Tanter, Celeste Hollenbeck, Sam Caldwell, Ming-Ho Yee, Lionel Zoubritzky, Benjamin Greenman and Matthias Felleisen for their feedback.

© Benjamin Chung, Paley Li, Francesco Zappa Nardelli, and Jan Vitek;
licensed under Creative Commons License CC-BY
32nd European Conference on Object-Oriented Programming (ECOOP 2018).
Editor: Todd Millstein; Article No. 12; pp. 12:1–12:25

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

"Because half the problem is seeing the problem"

There never was a single approach to gradual typing. The field was opened by two simultaneously published papers. One, by Siek and Taha, typed individual Scheme terms using a consistency relation, casts being inserted by a type directed translation [19]. The other, by Tobin-Hochtstadt and Felleisen, described a system allowing programmers to add types to individual modules, using constraint solving to determine where contracts are needed to protect typed and untyped code from each other [26]. These two approaches set the tone for a decade of research. Today, gradual type systems rely on a variety of languages, enforcement mechanisms with various guarantees; this linguistic diversity is not without consequence, however, as the very notion of what constitutes an error remains unsettled.

The type system and semantics of a programming language are necessarily tightly coupled; each has to deal with the language's complexity. As a result, the same gradual type system may seem very different when applied to two different languages, an issue that shows up clearly with object-oriented languages. Siek and Taha's first effort [20] presented a gradual type system for an object-oriented programing language. It related objects by generalizing the notion of consistency [19] over structural subtyping. The work had drawbacks, most notably in the handling of mutable state and aliasing – vital features of object-oriented languages. Underlying each subsequent gradual type system are different design choices on how to deal with mutability and aliasing.

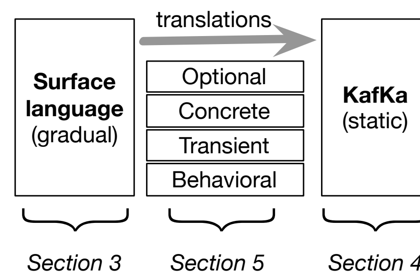
The landscape of gradually typed object-oriented languages is rich and includes:

- Typed Racket: a rich gradual type system based on contracts.
- Gradualtalk: a gradual variant of Smalltalk.
- C#: a statically typed language with a dynamic type.
- Dart: a class-based language with optional types.
- Hack: a statically typed variant of PHP that allows untyped code.
- Thorn: a language with both statically typed and untyped code.
- TypeScript: JavaScript with optional types.
- StrongScript: a variant of TypeScript with nominal types.
- Nom: a language supporting dynamic types and nominal typing.
- Reticulated Python: a family of gradual type systems for Python.

These languages differ in their type systems and associated run-time enforcement strategies. There are four major approaches, labeled here as optional, concrete, behavioral, and transient. The *optional* approach, chosen by TypeScript, Dart, and Hack, amounts to static type checking followed by type erasure. Erroneous values flowing from dynamically typed code to statically typed code will not be caught. The *concrete* approach, used in C# and Nom, uses run-time subtype tests on type constructors to supplement static typing. While statically typed code executes at native speed, values are dynamically checked at typed-untyped boundaries. The *behavioral* approach of Typed Racket and Gradualtalk monitors values to ensure that they behave in accordance to their assigned types. Instead of checking higher-order and mutable values for static type tags like concrete, wrappers ensure enduring conformance of values to their declared type. The *transient* approach, specific to Reticulated Python, lies between concrete and behavioral; it adds type casts but does so only for the top level of data structures. Finally, Thorn and StrongScript combine the optional and concrete approaches, differentiating between erased types and run-time-checked types.

Static type systems for object-oriented languages are designed to prevent dynamic "method not understood" errors. For gradual type systems, however, some method not found errors

cannot be ruled out before execution. In such a gradual type system, untyped code can pass an ill-typed value to typed code, breaking soundness. The meaning of an "error" for a gradual type system, therefore, depends on how type specifications are enforced. In other words, each gradual type system may catch different "errors." We demonstrate this with a litmus test consisting of three simple programs capable of distinguishing the four above-mentioned approaches. The litmus test programs are statically well-typed and "correct" in the sense that they run to completion without error in an untyped language. However, when executed under different gradual typing systems, they produce different errors. For intuition, consider a call, x.m(), where x : C and C has a method m returning a D. In the concrete approach, this call will succeed. With behavioral, the call will go through, but an error may be reported if m returns a value of the wrong type. In transient, the call is similarly guaranteed to go through, but might return the wrong type without reporting an error. Finally, in optional, the call may get stuck, as x may not have a method named m; and, if it succeeds, there is no guarantee that type D will be returned.



We propose to compare approaches to gradual typing for objects by translating a gradually typed surface language to a target language called KafKa. Our surface language is a *gradually* typed class-based object-oriented language similar to Featherweight Java. KafKa is a *statically* typed class-based object calculus with mutable state. The key difference between the two is the sound type system and casts of KafKa. Where the surface language allows implicit coercions, KafKa requires explicit casts to convert types. Casts come in two kinds: *structural casts* check for subtyping, while *behavioral casts* monitor that an object behaves as if it was of some type. Translating from surface to target language involves adding casts, the location and type of which depends on the gradual type system.

This paper makes the following contributions:

- The design of a core calculus for gradual type systems for objects.
- Translations of each gradual approach to the core calculus.
- A litmus test comprised of three programs to tell apart the gradual type systems.
- Supplementary material includes a mechanized proof of soundness of the type system of the core calculus and its proof-of-concept implementation on .Net.

Our work does not address the question of performance of the translations. Each of the semantics for gradual typing has intrinsic performance costs; but these can be mitigated by compiler and run-time optimizations, which we do not perform. KafKa departs from prior work (e.g. [12] as KafKa is statically typed. By translating to a statically typed core, we can clearly see where wrapper-induced dynamic errors can occur. Another design choice is the use of structural subtyping in KafKa. This is motivated by our desire to represent behavioral and transient approaches that require structural subtyping. We do not foresee difficulties either switching to a nominal type system or providing an additional nominal subtype cast.

All of our code and proofs are available from:

`github.com/BenChung/GradualComparisonArtifact`.

2 Background

"If you know the enemy and know yourself..."

The intellectual lineage of gradual typing can be traced back to attempts to add types to Smalltalk and LISP. On the Smalltalk side, work on the Strongtalk optional type system [7] led to Bracha's notion of pluggable types [6]. In Bracha's notion of pluggable types, types exist solely to catch errors at compile-time, never affecting the run-time behavior of programs. An optional type system is *trace preserving*: that is to say, if a term `e` reduces to `a`, then adding type annotations to `e` does not prevent it from reducing to `a` [18]. This property is valuable to developers as it ensures that type annotations will not introduce errors; and thus, adding types does not increase the testing burden! Optional type systems in wide use include Hack [25], TypeScript [3] and Dart [24].

Felleisen and his students have contributed substantially to gradual typing. The Typed Scheme [27] design, that later became Typed Racket, is influenced by their earlier work on higher-order contracts and semantic casts [10, 11]. Typed Racket was envisioned as a vehicle for teaching programming, being able to explain the source of errors and avoiding surprises for beginning users were important considerations. For this reason, a value that flowed in a variable of type `t`, was required behave as if it belonged to that type throughout its lifetime. Whenever a higher-order or mutable value crosses a boundary between typed and untyped code, it is wrapped in a contract that monitors the value's behavior. If the value misbehaves, blame can be assigned to the boundary that assigned it the type that was violated. The granularity of typing is the module, thus a module is either entirely typed or entirely untyped. Typed Racket's support for objects was described by Takikawa et al. [23].

Siek and Taha coined the term gradual typing in [19] as "any type system that allows programmers to control the degree of static checking for a program by choosing to annotate function parameters with types, or not." They formalized this idea in the lambda calculus augmented with references. To make the type system a gradual one, they defined the type consistency relation $\mathtt{t} \sim \mathtt{t}'$. If $\mathtt{t} \sim \mathtt{t}'$, then `t` is consistent with $\mathtt{t}'$, and can therefore be used implicitly where a $\mathtt{t}'$ instance is expected. This enables gradual typing, as $\star \sim \mathtt{t}$ for every `t` and vice versa, allowing untyped values to be passed where typed ones are expected. Siek and Taha extended this idea to an object calculus [20]. In order to do so, they combined consistency with structural subtyping, producing consistent subtyping. With consistent subtyping, consistency can be used when checking structural subtyping, allowing typed objects and untyped objects to be mixed. To explore the design space, Reticulated Python [28] was given three modes: the *guarded* mode behaves as Typed Racket with contracts applied to values. The *transient* mode performs shallow subtype checks on reads and method returns, only validating if the value obtained has matching method types. The *monotonic* mode is fundamentally different from any of the previous approaches. Monotonic cast updates the type of values in place by replacing some of the occurrences of $\star$ with more specific types, and these updates propagate recursively through the heap until fix-point.

Other noteworthy systems include Gradualtalk [1], C# 4.0 [4], Thorn [5], Nom [15] and StrongScript [18]. Gradualtalk is a variant of Smalltalk with behavioral casts and mostly nominal type equivalence (structural equivalence can be specified on demand, but it is rarely used). It has an optional mode and a mode in which blame can be turned off.

C# 4.0 adds the type `dynamic` to C# and adds dynamically resolved method invocation. Thus C# has a dynamic sublanguage that allows developers to write unchecked code, working alongside a sound typed sublanguage in which values are always of their declared type. The implementation replaces $\star$ by the type `object` and adds casts where needed.

	Nominal	*Optional*	*Concrete*	*Behavioral*	*Class based*	*First-class Class*	*Soundness claim*	*Unboxed prim.*	*Subtype cast*	*Shallow subtype cast*	*Behavioral cast*	*Blame*	*Pathologies*
Dart	•	•			•				•				-
Hack	•	•			•				•				-
TypeScript		•			•								-
C#	•		•		•		•[2]	•	•			•	-
Thorn	•	•	•		•		•[2]	•	•				-
StrongScript	•	•	•	•	•		•[2]		•		•		1.1x
Nom	•		•		•		•[2]	•	•			•	1.1x
Gradualtalk	•[1]			•	•		•				•	•	5x
Typed Racket				•	•	•	•			•	•	•	121x
Reticulated Python													
Transient		•			•		•			•			10x
Monotonic				•	•		•				•		27x
Guarded				•	•		•				•	•	21x

Figure 1 Gradual type systems. (1) Opt. structural constraints. (2) Typed expressions are sound.

Thorn and StrongScript extend the C# approach with the addition of optional types (called *like types* in Thorn). Thorn is implemented by translation to the JVM. StrongScript is implemented on top of a modified version of the V8 VM. The presence of concrete types means that the compiler can optimize code (unbox data and in-line methods) and programmers are ensured that type errors will not occur within concretely typed code. Nom is similar to Thorn in that it is nominal and follows the concrete approach.

Fig. 1 reviews gradual type systems for objects. All languages are class-based, except TypeScript which has both classes and JavaScript objects. While that choice is not crucial; classes are useful as a source of type declarations. Most languages build subtyping on explicit subtype declarations, nominal typing, rather than on structural similarities. TypeScript uses structural subtyping but does not implement a run-time check for it. Anecdotal evidence suggests that TypeScript could switch to nominal subtyping with little effort, as was done for StrongScript [18]. While nominal subtyping leads to more efficient type casts, Reticulated Python's subtype consistency relation is fundamentally structural; it would be nonsensical to use it in a nominal system.

For Racket, the heavy use of first-class classes and class generation naturally leads to structural subtyping as many of the classes being manipulated have no names and arise during computation.

The optional approach is the default for Dart, Hack, and TypeScript. Transient Reticulated Python allows any value to flow in a field, regardless of type annotations, leading to its "open world" soundness guarantee [28]. Some languages like Dart and Gradualtalk can operate in a checked and an uncheck mode. In Thorn, Nom, and C#, primitives are concretely typed; they can be unboxed without tagging. The choice of casts follows from other design decisions. The concrete approach naturally tends to use subtype tests to establish the type of values. For nominal systems, there are highly optimized algorithms.

Shallow casts are casts that only check the presence of methods but not their signature. They are used by Racket and Reticulated Python to ensure some basic form of type confor-

mance. Behavioral casts are used when information, such as a type or a blame label, must be associated with a reference or an object.

Some of the system provide soundness guarantees. In Typed Racket, Gradualtalk and Guarded Reticulated Python, there is a guarantee that a type error that is exrcised will be caught by a contract. In concrete approaches, any typed expression is guaranteed to be correct, errors occur at boundary crossings. The guarantees provided by Transient and Monotonic are somewhat harder to characterize and out of the scope of this review.

Blame assignment is a topic of investigation in its own right. Anecdotal evidence suggests that the context provided by blame helps developers pinpoint the provenance of errors. In the same way that a Java stack trace identifies the function that went wrong, blame identifies where a type assertion came from. This is especially important in behavioural gradual type systems, as type assertions become wrappers which can propagate through the heap. Blame identifies where a failing wrapper came from, a task that would otherwise require extensive backtracking debugging. Unlike stack traces, which have little run-time cost, blame tracking has a cost due to its meta-data. Blame information has to be stored whenever a wrapper is applied and is believed to cause substantial slowdowns. However, this has not been measured in detail. With the concrete approach, blame is trivially available as evaluation stops at the boundary that causes the failure [15]. We are primarily concerned with where the error arises, rather than what information is reported; thus, we do not consider blame further.

The last column of Fig. 1 lists self-reported performance pathologies. These numbers are not comparable, as they refer to different programs and different configurations of type annotations. They are not worst case scenarios either; most languages lack a sufficient corpus of code to conduct a thorough evaluation. Nevertheless, one can observe that for optional types no overhead is expected, as the type annotations are erased during compilation. Concrete types insert efficient casts and lead to code that can be optimized. The performance of the transient semantics for Reticulated Python is a worst case scenario for concrete types – i.e, there is a cast at almost every call. Finally, languages with behavioral casts are prone to significant slowdowns. Compiler optimizations for reducing these overheads are an active research topic [2, 17]. Languages such as C#, Nom, Thorn, and StrongScript are designed so that the performance of fully typed code is better than untyped code, so that mixed code performs well thanks to the relatively inexpensive nominal subtype tests.

In contemporary work, Greenman and Felleisen describe three approaches to *migratory typing* in the context of a lambda calculus extended with pairs and primitive values [12]. Their *natural embedding* corresponds to the behavioral approach, the *erasure embedding* is the optional approach, and the *locally-defensive embeddeding* is transient. They do not consider the concrete approach – neither objects or mutable state. They give performance results for a non-optimizing implementation of the embeddedings, and the results are as expected: behavioral has extreme worst cases and transient significantly slows down fully-typed programs. While they do not evalute object-oriented programs, these are unlikely to fare better. Our work differs in that we are trying to express a translation between object calculi using features that are readily available in most virtual machines.

3 A Family of Gradually Typed Languages and their Litmus Test

"There is no perfection only life"

There is no single, common notion of what constitutes an erroneous gradually typed program – a consequence of the varied enforcement strategies. The choice of enforcement strategy is reflected in the semantics of the language which, in turn, implies that developers have to understand the details of that strategy to avoid run-time errors. This also means that it is possible to differentiate between approaches by simply observing the run-time errors that each type system produces. We propose a litmus test consisting of three programs whose execution depends on which gradual type system is in use. Each of these programs is statically well-typed and runs without error when executed with a purely dynamic semantics. However, this varies as we use different semantics for gradual typing. We start by presenting a common surface language in which we can express our programs, and then explain why the various approaches to gradual typing yield different run-time errors.

3.1 A Common Surface Language

To normalize our presentation, we use a single surface language for all four of the gradual type systems under study. The surface language is a *gradually* typed object calculus without inheritance, method overloading or explicit type cast operations. Fig. 2 gives its syntax and an extract of its static semantics. The distinctive feature of the calculus is the presence of type $\star$ – the dynamic type. A variable of type $\star$ can hold any value, an invocation of a method with receiver of type $\star$ is always statically well-typed, and an expression of type $\star$ can appear anywhere within a typed program.

This dynamic type lets us convert our otherwise statically typed language to a gradually typed one. If a well-typed program does not use $\star$, then it will not get stuck on method invocation. A program where all variables are annotated as $\star$ is fully dynamic, and any given invocation may get stuck. Gradual typing comes into play when an expression of type $\star$ occurs as an argument to a method that expects some other type C and conversely when an argument of type C is passed to a method that expects $\star$. The static type system of the surface language allows such implicit coercions – using the *convertibility* relation – but run-time checks may be inserted to catch potential type mismatches. We formalize the semantics of this system later; here, we appeal to the readers' intuition.

Before presenting the litmus tests, some details about the type system of the surface language may prove helpful. The subtyping relation is structural with the Amber rule [8] to enable recursion. $\mathsf{M}\ \mathsf{K} \vdash \mathsf{C} <: \mathsf{D}$ holds if class C has (at least) all the methods of class D and the arguments and return types are related by subtyping in the usual contra- and co-variant way; the class table K holds definitions of all classes, and M is a helper for recursion that records the subtype relations encountered so far. One noteworthy feature of subtyping is that the fields of objects do not play a role in deciding if classes are subtypes. Following languages like Smalltalk, fields are encapsulated and can only be accessed from within their defining object. Syntactically, field reads and writes are limited to the self-reference this.

The static type checking rules, $\Gamma\,\mathsf{K} \vdash_s \mathsf{e} : \mathsf{t}$ where Γ is a type environment and K is a class table, are standard with two exceptions: method invocation and convertibility. Method invocation is always allowed when the receiver e is of type $\star$; therefore, $\mathsf{e.m(e')}$ has type $\star$ if the argument can have type $\star$. Convertibility is used when statically typed and dynamically typed terms interact. The convertibility relation, written $\mathsf{K} \vdash_s \mathsf{t} \Rrightarrow \mathsf{t'}$, states that type t is convertible to type $\mathsf{t'}$ in class table K. It is used both for up-casting and for conversions of $\star$ to non-$\star$ types. $\mathsf{K} \vdash_s \mathsf{t} \Rrightarrow \mathsf{t'}$ holds when $\mathsf{t} <: \mathsf{t'}$, this allows up-casts. The remaining

Syntax:

$$k ::= \textbf{class}\ C\,\{\, fd_1 ..\ md_1 ..\,\} \qquad md ::= m(x\colon t)\colon t\ \{e\} \qquad fd ::= f\colon t \qquad t ::= \star \mid C$$

$$K ::= k\ K \mid \cdot \qquad \Gamma ::= x\colon t\ \ \Gamma \mid \cdot \qquad M ::= C <: D\ M \mid \cdot$$

$$e ::= x \mid this \mid this.f \mid this.f = e \mid e.m(e) \mid \textbf{new}\ C(e_1..)$$

Typing expressions:

$$\frac{\Gamma(x) = t}{\Gamma\,K \vdash_s x : t} \qquad \frac{\Gamma(this) = C \quad f\colon t \in K(C)}{\Gamma\,K \vdash_s this.f : t} \qquad \frac{\Gamma(this) = C \quad f\colon t \in K(C) \quad \Gamma\,K \vdash_s e : t' \quad K \vdash_s t' \Mapsto t}{\Gamma\,K \vdash_s this.f = e : t} \qquad \frac{\Gamma\,K \vdash_s e : \star \quad \Gamma\,K \vdash_s e' : t}{\Gamma\,K \vdash_s e.m(e') : \star}$$

$$\frac{\Gamma\,K \vdash_s e : C \quad \Gamma\,K \vdash_s e' : t \quad m(t_1)\colon t_2 \in K(C) \quad K \vdash_s t \Mapsto t_1}{\Gamma\,K \vdash_s e.m(e') : t_2} \qquad \frac{f_1\colon t_1 .. \in K(C) \quad \Gamma\,K \vdash_s e_1 : t_1' .. \quad K \vdash_s t_1' \Mapsto t_1 ..}{\Gamma\,K \vdash_s \textbf{new}\ C(e_1..) : C}$$

Convertibility:

$$\frac{\cdot\ K \vdash_s t <: t'}{K \vdash_s t \Mapsto t'} \qquad \frac{}{K \vdash_s t \Mapsto \star} \qquad \frac{}{K \vdash_s \star \Mapsto t}$$

Subtyping:

$$\frac{}{M\ K \vdash \star <: \star} \qquad \frac{C <: D \in M}{M\ K \vdash C <: D} \qquad \frac{M' = C <: D\ M \quad md \in K(D) \implies md' \in K(C)\ .\ M'\ K \vdash md <: md'}{M\ K \vdash C <: D}$$

$$\frac{M\ K \vdash t_1' <: t_1 \qquad M\ K \vdash t_2 <: t_2'}{M\ K \vdash m(x\colon t_1)\colon t_2\ \{e\} <: m(x\colon t_1')\colon t_2'\ \{e'\}}$$

Figure 2 Surface language syntax and type system (extract).

two rules allow implicit conversion to and from the dynamic type. To avoid collapsing the type hierarchy, convertibility is not transitive. It is through convertibility that our surface language becomes gradual.

3.2 Litmus

Using three different programs, we can differentiate between four gradual type systems. The litmus test, shown in Fig. 3, and its constituent programs are written in our surface language. Each of these programs consists of a class table and an expression whose evaluation in the context of the class table determines if the litmus test succeeds or fails.

The programs in the litmus test are designed to induce errors. This is done by arranging for values to cross typed/untyped boundaries in a way that will cause some type systems to report an error but not others. At heart, these programs can be summarized by the type boundaries that are crossed by an object. The notation $C\,|_t$ denotes an object of class C passing through a boundary that expects it to be of type t. For example, if method m expects an argument of type t, a method call $e.m(e')$ would induce the boundary $e'\,|_t$. In program **L1**, we have:

$$A\,|_\star\,|_I$$

	L1	L2	L3
Concrete			
Behavioral		✓	
Transient		✓	✓
Optional	✓	✓	✓

L1

```
class A {
 m(x:A):A {this}}

class I {
 n(x:I):I {this}}

class T {
 s(x:I):T {this}
 t(x:*):* {this.s(x)}}
```

```
new T().t(new A())
```

L2

```
class A {
 m(x:A):A {this}}

class Q {
 n(x:Q):Q {this}}

class I {
 m(x:Q):I {this}}

class T {
 s(x:I):T {this}
 t(x:*):* {this.s(x)}}
```

```
new T().t(new A())
```

L3

```
class C {
 a(x:C):C {x}}

class D {
 b(x:D):D {x}}

class E {
 a(x:D):D {x}}

class F {
 m(x:E):E {x}
 n(x:*):* {this.m(x)}}
```

```
new F().n(new C())
       .a(new C())
```

Figure 3 Litmus test. Each program consists of a class table (top) and an expression (bottom). Top left table indicates successful executions.

An instance of class A is first implicitly converted to $\star$ and then to I; in this program classes A and I are unrelated by subtyping. In **L2**, the same sequence of conversions is applied:

$$\mathsf{A}\,|_{\star}\,|_{\mathsf{I}}$$

This time A and I both have a method m, but the methods have incompatible argument types. Lastly, in **L3** we start by converting a C to $\star$ and then to E and finally back to $\star$:

$$\mathsf{C}\,|_{\star}\,|_{\mathsf{E}}\,|_{\star}$$

The resulting value is then used to call method m with an argument of class C. This correct as method m in C does expect an argument of that type. If the object was an instance of E instead, the call would not be legal because E's method m expects a class D as argument.

Optional. An optional gradual type system simply erases all of the type annotations at run-time; all three programs run to completion without error.

Concrete. The concrete approach ensures that a variable of some class C always refers to an object of that class or of a subtype of it. To ensure this is the case, all implicit conversions imply a run-time subtype check. This causes all three programs to fail. **L1** and **L2** fail on a subtype test $K \vdash A <: I$. **L3** fails on the subtype test $K \vdash C <: E$.

Behavioral. The behavioral approach allows conversion from $\star$ to C if the value is compatible to C and if, after that, it behaves as if it was an instance of C. The former is checked by a shallow cast that only looks at method names, and the latter by a wrapper that monitors further interactions. **L1** fails at because A does not have the method n expected by I. **L2**, however, executes without error because A has the method m expected by I. **L3** fails, since the instance of C has been applied a wrapper for E. When method a is called with a C, the wrapper notices that E's method a expects a D and that C and D are not compatible.

Transient. The transient approach is weaker than behavioral. It retains the shallow structural checks at casts of the behavioral approach, but does not wrap values. Transient fails **L1**, for the same reason as the other two type systems, and passes **L2**, for the same reason as behavioral. **L3** succeeds because transient forgets that the C object was cast to E.

3.3 Discussion

These programs capture the behavior of implementations. The three have been expressed in TypeScript (optional), StrongScript (concrete), Typed Racket (behavioral) and Reticulated Python (transient), with the same errors.[1] What the litmus test shows is that a precise understanding of the semantics of gradually typed languages, and their run-time enforcement machinery, is crucial for developers to know if a program is "correct." Here, all errors are false positives, since none of these programs performs an invalid operation. This underlines the fact that in gradually typed language, type specifications can lead to run-time errors just as faulty code can. Thus, type annotations must be audited and tested just like code.

These approaches induce usability trade-offs. One way to contextualize this is with the gradual guarantee of Siek et al [21]. Informally, it states that if there exists a static type assignment to an untyped program that allows the program to run to completion, any partial assignment of those types will do too. The optional approach trivially fulfills this guarantee. Transient likewise satisfies the gradual guarantee [29], as it only checks top-level structure of values at type boundaries. Unlike optional, transient does ensure that typed calls succeed; however, a call may produce a dynamic error if the receiver is of the wrong type (even in a typed context), or the return is an ill-typed value. The behavioral approach also fulfills the guarantee, as it only checks arguments and return types when wrappers are invoked. However, typed function calls can fail if they call an untyped function that returns the wrong value. Finally, the concrete approach ensures that every typed call is successful. However, this comes at the expense of the gradual guarantee – partially typed classes are not compatible with more-or-less typed ones. The guarantee is incompatible with subtyping. Suppose a program were to rely on the judgment $\{m(\mathsf{C}) : \mathsf{D}\} <: \{m(\mathsf{C}) : \mathsf{D}\}$. Relaxing the argument type to a dynamic type, $\{m(\star) : \mathsf{D}\} <: \{m(\mathsf{C}) : \mathsf{D}\}$, violates subtyping. To overcome this, Reticulated Python augments subtyping with the aforementioned consistency relation. This increases the number of programs accepted by the static type system. However, it is not used by the run-time semantics and is fundamentally incompatible with the concrete approach (because any use of consistent subtyping will fail). We omit consistent subtyping.

As an alternative consider the approaches taken by Thorn or StrongScript. They have three kinds of types: $\star$ (dynamic), C (concrete), and like C (optional), combining the concrete and optional approach into the same language. This design allows for a different kind of migration; once a program is fully annotated with optional types, they all can be converted to concrete types without introducing any run-time errors [18]. We do not model this combination directly, as the underlying details are no different from the concrete approach.

The motivation for making fields private is to simplify the system. With private fields, errors are limited to method invocation. Field accesses can be trivially checked; as they are always accessing this. Moreover, interposing on method invocation can easily be achieved by wrappers, whereas interposing on field access would require modifying the code of clients. This would make the formal development more cumbersome without adding insight.

[1] github.com/BenChung/GradualComparisonArtifact/examples

4 KafKa: A Core Calculus

"Aux chenilles du monde entier et aux papillons qu'elles renferment"

Even without gradual typing, comparing languages is difficult. Small differences in syntax and features can make even the most similar languages appear different. As a result, the nuances of gradual type systems are often hidden amongst irrelevant details. To enable direct comparison, we propose to translate gradually typed languages down to a common calculus designed to highlight the distinctions between designs. The target for this translation is our core language, KafKa. KafKa is a statically typed language similar to the surface language but with several features added to enable its use as a common target language.

These additions make an explicit distinction between static and dynamic operations and replace implicit conversions with explicit casts. KafKa's first additions consist of two casts which are used at boundaries between typed and untyped code. The structural subtype cast, written $\langle t \rangle\, e$, ensures that expression e evaluates to a subtype of t. The behavioral cast, written $\blacktriangleleft t \blacktriangleright e$, creates a wrapper around the value of e that monitors e to ensure that it behaves as if it was of type t. Additionally, a syntactic distinction is made between static method invocation, written $e.m_{t \to t'}(e')$, dynamic method invocation, $e@m_{\star\to\star}(e')$. Static invocations does not get stuck, whereas dynamic invocations can. This makes explicit which function calls can fail.

KafKa was designed to align with common statically-typed class-based object-oriented compilation targets like .NET or the JVM. It maps to the intermediate language supported by these platforms. KafKa also requires the ability to generate new classes at run-time, a feature supported by these environments but typically not present in class-based calculi.

4.1 Syntax and Semantics

We had two requirements when designing KafKa: first, to be expressive enough to capture the dynamic semantics implied by each gradual type system; second, to have a type system that can express when code is known to be error free. KafKa's syntax and semantics, loosely inspired by Featherweight Java [13], are shown in Fig. 4. At the top level, classes are notated as **class** $C\,\{\, fd_1 ..\; md_1 .. \}$; methods, ranged over by md, are denoted as $m(x\colon t) : t\ \{e\}$; and fields $f : t$. Expressions consist of:

- variables, x;
- the self-reference this; and wrapper target, that;
- field accesses, this.f, and field writes, this.f = e;
- object creation, **new** $C(e_1..)$;
- static and dynamic method invocations;
- subtype and behavioral casts;
- and heap addresses.

The static semantics holds only a few surprises; key typing rules appear in Fig. 4. The subtyping relation is inherited from the surface language. The program typing relation (not shown here), $e\ K\ \ \checkmark$, indicates that expression e is well-formed with respect to class table K. The expression typing judgment $\Gamma\,\S\,K \vdash e : t$, indicates that against Γ with heap typing §and class table K, e has type t. Unlike the surface language, KafKa does not rely on a convertibility relation from $\star$ to C and back; instead, explicit casts are required.

Evaluation is mostly standard with an evaluation context consisting of a class table K, the expression being evaluated e, and a heap σ mapping from addresses a to objects, denoted $C\{a\ldots\}$. Due to the need for dynamic code generation, the class table is part of the state.

Syntax:

$$
\begin{array}{lllll}
\mathsf{e} & ::= & \mathsf{x} & \mid \mathsf{this} & \mid \mathsf{that} \quad \mid \\
 & & \mathsf{this.f} & \mid \mathsf{this.f} = \mathsf{e} & \mid \mathbf{new}\ \mathsf{C}(\mathsf{e}_1..) \mid \\
 & & \mathsf{e.m}_{\mathsf{t}\to\mathsf{t}}(\mathsf{e}) & \mid \mathsf{e@m}_{\star\to\star}(\mathsf{e}) & \mid \\
 & & \langle \mathsf{t} \rangle\ \mathsf{e} & \mid \blacktriangleleft \mathsf{t} \blacktriangleright \mathsf{e} & \mid \\
 & & \mathsf{a} & \mid \mathsf{a.f} & \mid \mathsf{a.f} = \mathsf{e}
\end{array}
\qquad
\begin{array}{lll}
\mathsf{k} & ::= & \mathbf{class}\ \mathsf{C}\,\{\,\mathsf{fd}_1..\ \mathsf{md}_1..\,\} \\
\mathsf{md} & ::= & \mathsf{m}(\mathsf{x}:\mathsf{t}):\mathsf{t}\ \{\mathsf{e}\} \\
\mathsf{fd} & ::= & \mathsf{f}:\mathsf{t} \\
\mathsf{t} & ::= & \star \mid \mathsf{C} \\
\mathsf{K} & ::= & \mathsf{k}\ \mathsf{K} \mid \cdot \\
\Sigma & ::= & \mathsf{a}:\mathsf{t}\ \Sigma \mid \cdot
\end{array}
$$

Static semantics:

$$
\frac{\Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{e} : \mathsf{C} \qquad \mathsf{m}(\mathsf{t}):\mathsf{t}' \in \mathsf{K}(\mathsf{C}) \qquad \Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{e}' : \mathsf{t}}{\Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{e.m}_{\mathsf{t}\to\mathsf{t}'}(\mathsf{e}') : \mathsf{t}'}
\qquad
\frac{\Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{e} : \star \qquad \Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{e}' : \star}{\Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{e@m}_{\star\to\star}(\mathsf{e}') : \star}
$$

$$
\frac{\Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{e} : \mathsf{t}'}{\Gamma\,\Sigma\,\mathsf{K} \vdash \langle \mathsf{t} \rangle\,\mathsf{e} : \mathsf{t}}
\qquad
\frac{\Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{e} : \mathsf{t}'}{\Gamma\,\Sigma\,\mathsf{K} \vdash \blacktriangleleft \mathsf{t} \blacktriangleright \mathsf{e} : \mathsf{t}}
\qquad
\frac{\Sigma(\mathsf{a}) = \mathsf{C}}{\Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{a} : \mathsf{C}}
\qquad
\frac{}{\Gamma\,\Sigma\,\mathsf{K} \vdash \mathsf{a} : \star}
$$

Execution contexts:

$$
\begin{array}{lllllll}
\mathsf{E} & ::= & \mathsf{a.f} = \mathsf{E} & \mid \mathsf{E.m}_{\mathsf{t}\to\mathsf{t}}(\mathsf{e}) & \mid \mathsf{a.m}_{\mathsf{t}\to\mathsf{t}}(\mathsf{E}) & \mid \mathsf{E@m}_{\star\to\star}(\mathsf{e}) & \mid \\
 & & \mathsf{a@m}_{\star\to\star}(\mathsf{E}) & \mid \langle \mathsf{t} \rangle\ \mathsf{E} & \mid \blacktriangleleft \mathsf{t} \blacktriangleright \mathsf{E} & \mid \mathbf{new}\ \mathsf{C}(\mathsf{a}_1..\,\mathsf{E}\,\mathsf{e}_1..) & \mid \square
\end{array}
$$

Dynamic semantics:

$$
\begin{array}{llllllll}
\mathsf{K} & \mathbf{new}\ \mathsf{C}(\mathsf{a}_1..) & \sigma & \to & \mathsf{K} & \mathsf{a}' & \sigma' & \mathbf{where}\ \ \mathsf{a}'\ \text{fresh} \qquad \sigma' = \sigma[\mathsf{a}' \mapsto \mathsf{C}\{\mathsf{a}_1..\}] \\
\mathsf{K} & \mathsf{a.f}_i & \sigma & \to & \mathsf{K} & \mathsf{a}_i & \sigma & \mathbf{where}\ \ \sigma(\mathsf{a}) = \mathsf{C}\{\mathsf{a}_1, \ldots \mathsf{a}_i, \mathsf{a}_n \ldots\} \\
\mathsf{K} & \mathsf{a.f}_i = \mathsf{a}' & \sigma & \to & \mathsf{K} & \mathsf{a}' & \sigma' & \mathbf{where}\ \ \sigma(\mathsf{a}) = \mathsf{C}\{\mathsf{a}_1, \ldots \mathsf{a}_i, \mathsf{a}_n \ldots\} \\
 & & & & & & & \qquad\quad \sigma' = \sigma[\mathsf{a} \mapsto \mathsf{C}\{\mathsf{a}_1, \ldots \mathsf{a}', \mathsf{a}_n \ldots\}] \\
\mathsf{K} & \mathsf{a.m}_{\mathsf{t}\to\mathsf{t}'}(\mathsf{a}') & \sigma & \to & \mathsf{K} & \mathsf{e}' & \sigma & \mathbf{where}\ \ \mathsf{e}' = [\mathsf{a}/\mathsf{this}\ \mathsf{a}'/\mathsf{x}]\mathsf{e} \\
 & & & & & & & \qquad\quad \mathsf{m}(\mathsf{x}:\mathsf{t}_1):\mathsf{t}_2\ \{\mathsf{e}\} \in \mathsf{K}(\mathsf{C}) \\
 & & & & & & & \qquad\quad \sigma(\mathsf{a}) = \mathsf{C}\{\mathsf{a}_1..\} \quad \emptyset\ \mathsf{K} \vdash \mathsf{t} <: \mathsf{t}_1 \\
 & & & & & & & \qquad\quad \emptyset\ \mathsf{K} \vdash \mathsf{t}_2 <: \mathsf{t}' \\
\mathsf{K} & \mathsf{a@m}_{\star\to\star}(\mathsf{a}') & \sigma & \to & \mathsf{K} & \mathsf{e}' & \sigma & \mathbf{where}\ \ \mathsf{e}' = [\mathsf{a}/\mathsf{this}\ \mathsf{a}'/\mathsf{x}]\mathsf{e} \\
 & & & & & & & \qquad\quad \mathsf{m}(\mathsf{x}:\star):\star\ \{\mathsf{e}\} \in \mathsf{K}(\mathsf{C}) \\
 & & & & & & & \qquad\quad \sigma(\mathsf{a}) = \mathsf{C}\{\mathsf{a}_1..\} \\
\mathsf{K} & \langle \star \rangle\ \mathsf{a} & \sigma & \to & \mathsf{K} & \mathsf{a} & \sigma & \\
\mathsf{K} & \langle \mathsf{D} \rangle\ \mathsf{a} & \sigma & \to & \mathsf{K} & \mathsf{a} & \sigma & \mathbf{where}\ \ \emptyset\ \mathsf{K} \vdash \mathsf{C} <: \mathsf{D} \qquad \sigma(\mathsf{a}) = \mathsf{C}\{\mathsf{a}_1..\} \\
\mathsf{K} & \blacktriangleleft \mathsf{t} \blacktriangleright \mathsf{a} & \sigma & \to & \mathsf{K}' & \mathsf{a}' & \sigma' & \mathbf{where}\ \ \mathsf{K}'\ \mathsf{a}'\ \sigma' = \mathsf{bcast}(\mathsf{a}, \mathsf{t}, \sigma, \mathsf{K}) \\
\mathsf{K} & \mathsf{E}[\mathsf{e}] & \sigma & \to & \mathsf{K}' & \mathsf{E}[\mathsf{e}'] & \sigma' & \mathbf{where}\ \ \mathsf{K}\ \mathsf{e}\ \sigma \to \mathsf{K}'\ \mathsf{e}'\ \sigma'
\end{array}
$$

Figure 4 KafKa dynamic semantics and static semantics (extract).

4.1.1 Method Invocation

KafKa has two invocation forms, the dynamic $\mathsf{e@m}_{\star\to\star}(\mathsf{e}')$ and the static $\mathsf{e.m}_{\mathsf{t}\to\mathsf{t}'}(\mathsf{e}')$, both denoting a call to method m with argument e'. There are several design issues worth discussing. First, as our calculus is a translation target, it is acceptable to require some explicit preparation for objects to be used in a dynamic context. A dynamic call is only successful if the receiver has a method of the expected name and with argument and return types of $\star$. This is a design choice of KafKa, even dynamic invocation has to be well-typed

(even if this typing is trivial). Secondly, it is possible for a static invocation to call an untyped method (of type $\star$ to $\star$). Consider the following class definition

$$\begin{array}{l}\textbf{class}\ \mathsf{C}\ \{\\ \qquad \mathsf{m(x\colon Int)\colon Int}\ \{\ \mathsf{x}+2\ \}\\ \qquad \mathsf{m\ (x\colon \star)\ :\ \star}\ \{\ \langle\star\rangle\,\mathsf{this.m_{Int\to Int}}(\langle\mathsf{Int}\rangle\,\mathsf{x})\ \}\\ \}\end{array}$$

assume that class table K holds a definition for Int, and that we have integer constants and addition. The class demonstrates several features of KafKa. Its class well-formedness rules (not shown here) allow a limited form of method overloading. A class may have at most two occurrences of any method m: one "untyped" with $\star$ as argument and return type; and one, which we call "typed", where either its argument or return type differ from $\star$. The static type system enforces a single means of invoking a typed method m:

$$\textbf{new}\ \mathsf{C}().\mathsf{m_{Int\to Int}}(2)$$

Here, the receiver is obviously of type C and the argument is Int; thus, the call is statically well-typed. The expression will therefore evaluate the body of m. For an untyped method, there are two invocation modes:

$$\textbf{new}\ \mathsf{C}()@\mathsf{m}_{\star\to\star}(2)$$

The call executes correctly here, as C has an untyped m. However, in the general case, there is no guarantee that the receiver of an untyped invocation has the requested method; and therefore, a dynamic invocation can get stuck. The receiver of a dynamic invocation has type $\star$. When the receiver type is known to be some C and that class has the requested method, then the static invocation can be used:

$$\textbf{new}\ \mathsf{C}().\mathsf{m}_{\star\to\star}(2)$$

The invocation will succeed. All nuances of invocation will come in handy when translating the surface language to KafKa.

4.1.2 Run-time Casts

KafKa has two cast operations: the structural subtype cast $\langle \mathsf{t}\rangle\,\mathsf{e}$ and the behavioral cast $\blacktriangleleft \mathsf{t} \blacktriangleright\, \mathsf{e}$, both indicating the desire for the result of evaluating e to be of type t. Where the casts differ is what is meant by "has type t". The subtype cast checks that the result of evaluating e is an object whose class is compatible with t. If t is a class, then it will check for a subtyping relation; otherwise, if $\mathsf{t} = \star$, the cast will succeeds. As every value in the heap is tagged by its type constructor, it is always possible to perform this check. The behavioral cast is more complex; we will describe it in the remainder of this section.

The objective of the behavioral cast is to ensure that the wrapped object behaves as the target type dictates. When given the address a of some object, this cast creates a wrapper object, say a', that enforces the invariant that a *behaves* as a value of type t. Function $\mathsf{bcast}(\mathsf{a}, \mathsf{t}, \sigma, \mathsf{K}) = \mathsf{K}'\,\mathsf{a}'\,\sigma'$ specifies its semantics, shown in Fig. 5. There are two cases to consider: either the target type is a class C', or it is $\star$.

If the target type is C', then $\mathsf{bcast}(\mathsf{a}, \mathsf{C}', \sigma, \mathsf{K})$ will return an updated class table K', a reference to the wrapped object a', and an updated heap σ'. As long as a has every method name specified by C', the cast itself will succeed. If a is missing a method, it is

Behavioral cast: $\mathsf{bcast}(a, t, \sigma, K) = K'\, a'\, \sigma'$

a	Reference to wrap		a'	Wrapped reference
t	Target type to enforce		K'	Class table with wrapper
σ	Original heap		σ'	New heap
K	Original class table			

$$\mathsf{bcast}(a, C', \sigma, K) = K'\, a'\, \sigma' \quad \textbf{where} \quad \begin{cases} \sigma(a) = C\{a_1..\} \quad D, a' \text{ fresh} \quad \sigma' = \sigma[a' \mapsto D\{a\}] \\ md_1.. \in K(C) \quad \mathsf{names}(md'_1..) \subseteq \mathsf{names}(md_1..) \\ md'_1.. \in K(C') \quad \mathsf{nodups}(md_1..) \quad \mathsf{nodups}(md'_1..) \\ K' = K\ W(C, md_1.., md'_1.., D) \end{cases}$$

$$\mathsf{bcast}(a, \star, \sigma, K) = K'\, a'\, \sigma' \quad \textbf{where} \quad \begin{cases} \sigma(a) = C\{a_1..\} \quad md_1.. \in K(C) \quad D, a' \text{ fresh} \\ \mathsf{nodups}(md_1..) \quad K' = K\ W\star(C, md_1.., D) \\ \sigma' = \sigma[a' \mapsto D\{a\}] \end{cases}$$

$$W(C, md_1.., md'_1.., D) = \textbf{class}\ D\,\{\,\mathsf{that}\colon C\ md''_1..\,\}$$
$$\textbf{where}\quad m(x\colon t_1)\colon t_2\ \{e\} \in md_1..$$
$$md''_1 = m(x\colon t'_1)\colon t'_2\ \{\ \blacktriangleleft t'_2 \blacktriangleright \mathsf{this.that}.m_{t_1\to t_2}(\blacktriangleleft t_1 \blacktriangleright x)\}\ ..\quad \textbf{if}\quad m(x\colon t'_1)\colon t'_2\ \{e'\} \in md'_1..$$
$$m(x\colon t_1)\colon t_2\ \{\ \mathsf{this.that}.m_{t_1\to t_2}(x)\}\ ..\quad \textbf{otherwise}$$

$$W\star(C, md_1.., D) = \textbf{class}\ D\,\{\,\mathsf{that}\colon C\ md'_1..\,\}$$
$$\textbf{where}\quad md'_1 = m(x\colon \star)\colon \star\ \{\ \blacktriangleleft \star \blacktriangleright \mathsf{this.that}.m_{t\to t'}(\blacktriangleleft t \blacktriangleright x)\}\ ..\quad \textbf{if}\quad m(x\colon t)\colon t'\ \{e\} \in md_1..$$

Figure 5 Behavioral cast semantics.

impossible for a to implement C' correctly, and early failure is indicated. Otherwise, the metafunction continues to generate a type wrapper. Class generation itself is delegated to the W metafunction. A W invocation, $W(C, md_1.., md'_1.., D)$, takes a class C, a fresh name D, and two method lists $md_1..$ and md'_1, respectively the method of C and the methods of the type to enforce. The class generated by W will have adapter methods for each method m occurring in both $md_1..$ and $md'_1...$ Type mismatches between the wrapped object and the wrapping type are resolved with more behavioral casts. For methods that do not need to be adapted (methods only in $md_1..$), a simple pass-through method is generated. This method simply calls the wrapped object; itself referred to by the distinguished variable that.

If the target type is $\star$, the wrapper class is simpler. It only needs to check that method arguments match the types expected by the wrapped object. This is done by another behavioral cast. Return values are cast to $\star$.

For example, consider the following program which has two classes C and D. Even though C and D both have method a, they are not subtypes because the arguments to their m implementations are not related.

```
(◀C▶ (◀D▶ new C())).b(2)      where  K  =  class C {
                                              m(x: ⋆): ⋆ {  x  }
                                              n(x: ⋆): ⋆ {  x  }
                                            }
                                            class D {  m(x: Int): Int {  x  }   }
```

The program starts with a C, casts it to D, and then back to C. The reason we generate pass-through methods (the wrapper that enforces type D has a method n) is that without them, the method n would be "lost". Without the pass-through method, it would be impossible to cast back to C, as the wrapper would only have an implementation for m. Wrappers with this semantics are referred to as opaque, as it is not possible to see methods of the underlying object though them. In contrast, KafKa uses transparent wrappers. To illustrate what this looks like, the following class E is generated by the cast from C to D:

```
class E {
        that: C
        m(x: Int): Int {  ◀Int▶ that.m⋆→⋆(◀⋆▶ x)  }
        n(x: ⋆): ⋆ {  that.n⋆→⋆(x)  }
}
```

By keeping n present, it is possible to return an instance of E to C again. If we were to remove n, then E would no longer be convertible back to C again.

4.2 Type soundness

The KafKa type soundness theorem ensures that a well-formed program can only get stuck at a dynamic invocation, a subtype cast, or a behavioral cast, and only there when justified.

▶ **Theorem 1** (KafKa type soundness). *For every well-formed state* $\mathsf{K}\ \mathsf{e}\ \sigma\ \checkmark$ *and well-typed expression* $\emptyset\,\Sigma\,\mathsf{K} \vdash \mathsf{e} : \mathsf{t}$, *where heap typing* Σ *is obtained by mapping the class of each object to the corresponding address, one of the following holds:*

- *There exists some reference* a *such that* $\mathsf{e} = \mathsf{a}$.
- $\mathsf{K}\ \mathsf{e}\ \sigma \to \mathsf{K}'\ \mathsf{e}'\ \sigma'$, *where* $\mathsf{K}'\ \mathsf{e}'\ \sigma'\ \checkmark$, $\emptyset\,\Sigma'\,\mathsf{K}' \vdash \mathsf{e}' : \mathsf{t}$, σ' *has all of the values of* σ, Σ' *has all of the types of* σ', *and* K' *has all of the classes of* K.
- $\mathsf{e} = E[\mathsf{a}@\mathsf{m}_{\star\to\star}(\mathsf{a}')]$ *and* a *refers to an object without a method* m.
- $\mathsf{e} = E[\langle \mathsf{C} \rangle\ \mathsf{a}]$, *and* a *refers to an object whose class is not a subtype of* C.
- $\mathsf{e} = E[◀\mathsf{C}▶\ \mathsf{a}]$, *and* C *contains a method that* a *does not.*

The proof is mostly straightforward, with one unusual case, centered around the bcast metafunction. When the bcast metafunction is used to generate a wrapper class, which is then instantiated, producing a new class table and heap, we must then show that the new class table is well formed, that the new heap is also well formed, and that the new wrapper is a subtype of the given type C. Proving these properties is relatively easy. Class table well-formedness follows by construction of the wrapper class and by well-formedness of the old class table. Heap well-formedness follows by well-formedness of the class table, construction of the new heap, and well-formedness of the old heap. Proving that the type of the wrapper is a subtype of the required type proceeds by structural induction over the required type. The proof of soundness has been formalized in Coq and is available in the supplementary material. The proof relies on two axioms dealing with recursive structural subtyping. We did not prove these as they have been shown in prior work [14].

4.3 Discussion

The design of KafKa's two invocation forms bears discussion. In some previous works, dynamic invocation has been implemented by a combination of a cast and a statically typed call. In our case, following this approach would require creating a type for each invocation (as was done in [30]). Instead, providing a dynamic invocation form seemed more natural. The use of explicitly typed invocation is a result of our desire to be able to rule out more errors statically. Whenever a translation can generate a static invocation, the soundness result ensure that the call will succeed. But, some methods need to be called from both a typed and untyped context, to achieve this we generate two versions of the method and leverage the difference between typed and untyped calls to express invovations occuring in each context.

One of our requirements for KafKa was that it support transparent wrappers as these are needed for the behavioral approach. The combination of structural subtyping and dynamic class generation allows to generate subtypes on the fly. These subtypes have all the methods of the target type plus some new ones. The choice of having fields be private means that the wrappers do not have to special case field access. If fields were accessible from outside the object, some more complex rewritting would have to be used.

KafKa was intended to match the intermediate languages of commercial VMs. To validate this, we implemented a compiler from KafKa to C#.[2] The only challenge was due to subtyping. KafKa uses structural typing, while C# is nominal, and KafKa allows methods in subtypes to be contra-variant in argument and co-variant in return type, while C# requires invariance.

Implementing structural subtyping on top of a nominally typed language is tricky. Structural types create implicit subtyping relationships, which the nominal type system expects to be explicit. Prior work used reflection and complex run-time code generation [9], but this is needlessly complex for a proof of concept. Instead, we reify the implicit relationships introduced by structural subtyping into explicit nominal relationships by generating interfaces. Given two classes C and D, where $\mathsf{K} \vdash \mathsf{C} <: \mathsf{D}$ holds, we generate two interfaces CI and DI, where CI is declared to extend DI. As a result, if two types are subtypes, their corresponding C# interfaces will be as well. The next problem is that KafKa allows subtype methods to be contra-variant in argument and co-variant in return types. As a result, a single method in CI may not be sufficient to implement DI. We solve this by having every class's C# equivalent implement every interface explicitly, with each explicit implementation delegating to the real, most general, implementation.

Despite these issues, we were able to accurately translate KafKa types. We translate static and dynamic invocations into corresponding C# invocations since C# also has a dynamic type. The underlying run time can then use the translated KafKa types to perform method dispatch, while inserting dynamic checks wherever the KafKa code calls for an untyped invocation. This prototype shows that KafKa primitives are close to those of intermediate languages. As a result, the translation of gradual type systems to KafKa provides insight as to how they might be implemented.

---

[2] `github.com/BenChung/GradualComparisonArtifact/netImpl`

5 Translating Gradual Type Systems

“Was ist mit mir geschehen? dachte er. Es war kein Traum”

Equipped with the source and target languages, we can describe the gradual-to-statically typed translation from source to KafKa. Each semantics is translated through a function mapping well-typed surface programs into well-typed KafKa terms. The translation explicitly determines which type casts need to be inserted and the invocation forms to use. A type-driven translation will insert casts where the surface language used consistency.

5.1 Class Translation

The translations for surface level classes are shown in Fig. 6. Each class in the surface language translates to a homonymous KafKa class, retaining type names through translation. The grey background denotes code generated in the translation. The notation $e; e'$ denotes sequencing.

Optional. The optional approach provides no correctness guarantees. Retaining the surface type annotations through translation would not preserve this semantics, so we erase them. The resulting class has all fields, all method arguments, and all return values typed as $\star$.

Transient. In the transient approach, ensure that for any method call, the receiver does have a method with the corresponding name. To encode this within the KafKa type system

Optional:

$$
\begin{array}{lll}
\mathcal{O}[\![\textbf{class } C\,\{\, fd_1 ..\ md_1 ..\,\}]\!] & = & \textbf{class } C\,\{\, fd'_1 ..\ md'_1 ..\,\} \\
& \textbf{where} & fd'_1 = f:\star \,..\quad fd_1 = f:t.. \\
& & md'_1 = m(x:\star):\star\ \{e'\}\ .. \\
& & md_1 = m(x:t_1):t_2\ \{e\} \quad e' = \mathcal{O}[\![e]\!]
\end{array}
$$

Transient:

$$
\begin{array}{lll}
\mathcal{T}[\![\textbf{class } C\,\{\, fd_1 ..\ md_1 ..\,\}]\!] & = & \textbf{class } C\,\{\, fd'_1 ..\ md'_1 ..\,\} \\
& \textbf{where} & fd'_1 = f:\star \,..\quad fd_1 = f:t.. \\
& & md'_1 = m(x:\star):\star\ \{\langle t\rangle\, x\ ;\ e'_1\}\ .. \\
& & md_1 = m(x:t):t'\ \{e\}.. \quad e'_1 = \mathcal{T}(\!|e|\!)^{\star}_{x:t\ \mathsf{this}:C}\ ..
\end{array}
$$

Behavioral:

$$
\begin{array}{lll}
\mathcal{B}[\![\textbf{class } C\,\{\, fd_1 ..\ md_1 ..\,\}]\!] & = & \textbf{class } C\,\{\, fd_1 ..\ md'_1 ..\,\} \\
& \textbf{where} & md'_1 = m(x:t):t'\ \{e'_1\}\ .. \\
& & md_1 = m(x:t):t'\ \{e_1\}\ .. \quad e'_1 = \mathcal{B}[\![e_1]\!]_{x:t\ \mathsf{this}:C}
\end{array}
$$

Concrete:

$$
\begin{array}{lll}
\mathcal{C}[\![\textbf{class } C\,\{\, fd_1 ..\ md_1 ..\,\}]\!] & = & \textbf{class } C\,\{\, fd_1 ..\ md'_1 .. md''_1 ..\,\} \\
& \textbf{where} & md'_1 = m(x:t_1):t_2\ \{e'\}\ .. \\
& & md_1 = m(x:t_1):t_2\ \{e\}.. \quad e' = \mathcal{C}(\!|e|\!)^{t_2}_{\mathsf{this}:C\ x:t_1}\ .. \\
& & md''_1 = m(x:\star):\star\ \{\langle\star\rangle\, \mathsf{this}.m_{t_1\to t_2}(\langle t_1\rangle\, x)\} \quad \textbf{if}\ \ t_1 \neq \star \\
& & \qquad \textit{empty} \quad \textbf{otherwise}\ ..
\end{array}
$$

Figure 6 Translations for classes.

requires replacing all of the argument and return types with $\star$. This translation allows functions to be called under any type and to return values of any type. Casts to the erased types are then effectively shallow structural checks only. As there is no guarantee that fields contain values of their type, the translation sets their type to $\star$.

Behavioral. The behavioral approach guarantees soundness by wrapping values that cross type-untyped boundaries. Methods are preserved by the translation but bodies are translated.

Concrete. The concrete approach ensures that variables of non-$\star$ types refer to subtypes of the given type. Each method appearing in the original class is retained as such with its body translated. Moreover, all typed methods could be called from an untyped context, so untyped variants are generated that guard the typed functions. These variants perform subtype casts on their arguments to ensure that they were given the right types, then call the guarded typed function.

5.2 Expression Translation

To accommodate differences between the gradual typing semantics, we use two different expression translation schemes. The first is a type-agnostic one, used for the optional approach. $\mathcal{O}[\![\mathsf{e}]\!]$ denotes optional translation, where e is the target expression, and the result is a KafKa term. We use this form to simplify the optional translation, as it is ambivalent about the types of the expressions it is translating; the optional semantics simply eliminates all of them.

The second translation form is type-aware, used for the three other approaches. The type-aware translation has two forms, $\mathcal{S}[\![\mathsf{e}]\!]_\Gamma$ and $\mathcal{S}(\!|\mathsf{e}|\!)^{\mathsf{t}}_\Gamma$, inspired by work on bidirectional type-checking [16]. The first form, $\mathcal{S}[\![\mathsf{e}]\!]_\Gamma$, is analogous to the synthetic case in bidirectional type-checking. It is used for expressions without any specific required type. The second form, $\mathcal{S}(\!|\mathsf{e}|\!)^{\mathsf{t}}_\Gamma$, is used when e must have some type t. Analogous to the analytic case of bidirectional type-checking, this form applies when some enclosing expression has an expectation of the type of e. For example, it is used in translation of method arguments, which must conform to the types of the arguments to the method. We refer to this as assertive translation. These two forms allow us to identify where consistency was needed to conclude the surface level typing judgment.

Transient:

$$\mathcal{T}(\!|\mathsf{e}|\!)^{\mathsf{t}}_\Gamma = \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{e} : \mathsf{t}' \quad \mathsf{K} \vdash \mathsf{t}' <: \mathsf{t} \quad \mathsf{e}' = \mathcal{T}[\![\mathsf{e}]\!]_\Gamma$$
$$\mathcal{T}(\!|\mathsf{e}|\!)^{\mathsf{t}}_\Gamma = \langle \mathsf{t} \rangle\, \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{e} : \mathsf{t}' \quad \mathsf{K} \vdash \mathsf{t}' \not<: \mathsf{t} \quad \mathsf{e}' = \mathcal{T}[\![\mathsf{e}]\!]_\Gamma$$

Behavioral:

$$\mathcal{B}(\!|\mathsf{e}|\!)^{\mathsf{t}}_\Gamma = \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{e} : \mathsf{t}' \quad \mathsf{K} \vdash \mathsf{t}' <: \mathsf{t} \quad \mathsf{e}' = \mathcal{B}[\![\mathsf{e}]\!]_\Gamma$$
$$\mathcal{B}(\!|\mathsf{e}|\!)^{\mathsf{t}}_\Gamma = \blacktriangleleft \mathsf{t} \blacktriangleright \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{e} : \mathsf{t}' \quad \mathsf{K} \vdash \mathsf{t}' \not<: \mathsf{t} \quad \mathsf{e}' = \mathcal{B}[\![\mathsf{e}]\!]_\Gamma$$

Concrete:

$$\mathcal{C}(\!|\mathsf{e}|\!)^{\mathsf{t}}_\Gamma = \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{e} : \mathsf{t}' \quad \mathsf{K} \vdash \mathsf{t}' <: \mathsf{t} \quad \mathsf{e}' = \mathcal{C}[\![\mathsf{e}]\!]_\Gamma$$
$$\mathcal{C}(\!|\mathsf{e}|\!)^{\mathsf{t}}_\Gamma = \langle \mathsf{t} \rangle\, \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{e} : \mathsf{t}' \quad \mathsf{K} \vdash \mathsf{t}' \not<: \mathsf{t} \quad \mathsf{e}' = \mathcal{C}[\![\mathsf{e}]\!]_\Gamma$$

Figure 7 Assertive translation.

The assertive translation of Fig. 7 is responsible for producing well-typed terms by adding casts into expressions where static types differ. The rules closely track the convertibility

relation of the surface language. Every type-driven translation has two cases. The first case is used when the required type happens to be a supertype of the expression's actual type, in which cast upcasting can happen implicitly and no further action is required. The second case handles typed-untyped boundaries, conversions to or from $\star$. The concrete and transient approaches both use the subtype cast operator to protect these boundaries, though the effective semantics are different; concrete retains the types that transient erases, so subtype casts in transient check structural compatibility (e.g. are all the needed methods present) alone whereas concrete subtype casts check the entire object's types. The behavioral approach instead inserts behavioral casts at boundaries.

The translation of field access appears in Fig. 8. The optional translation only inserts a cast to $\star$ in front of uses this as a technicality required for statically typing terms. The transient translation casts variables and fields to their statically expected type, as their values may be of any type. In transient, however, subtype casts only check the structure of the types. The behavioral translation and the concrete translation leave both types of access intact.

Optional:

$$\mathcal{O}[\![\mathsf{x}]\!] = \mathsf{x}$$
$$\mathcal{O}[\![\mathsf{this}]\!] = \langle\star\rangle\,\mathsf{this}$$
$$\mathcal{O}[\![\mathsf{this.f}]\!] = \mathsf{this.f}$$

Transient:

$$\mathcal{T}[\![\mathsf{x}]\!]_\Gamma = \langle\mathsf{t}\rangle\,\mathsf{x} \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{x} : \mathsf{t}$$
$$\mathcal{T}[\![\mathsf{this}]\!]_\Gamma = \mathsf{this}$$
$$\mathcal{T}[\![\mathsf{this.f}]\!]_\Gamma = \langle\mathsf{t}\rangle\,\mathsf{this.f} \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{this} : \mathsf{C} \quad \mathsf{f} : \mathsf{t} \in \mathsf{K}(\mathsf{C})$$

Behavioral:

$$\mathcal{B}[\![\mathsf{x}]\!]_\Gamma = \mathsf{x}$$
$$\mathcal{B}[\![\mathsf{this}]\!]_\Gamma = \mathsf{this}$$
$$\mathcal{B}[\![\mathsf{this.f}]\!]_\Gamma = \mathsf{this.f}$$

Concrete:

$$\mathcal{C}[\![\mathsf{x}]\!]_\Gamma = \mathsf{x}$$
$$\mathcal{C}[\![\mathsf{this}]\!]_\Gamma = \mathsf{this}$$
$$\mathcal{C}[\![\mathsf{this.f}]\!]_\Gamma = \mathsf{this.f}$$

Figure 8 Translations variables and field access.

The translation for assignment is shown in Fig. 9. All the approaches translate the value only differing in the expected type. Behavioral and concrete require that the result has the statically known type, transient expects $\star$, and the optional semantics imposes no type requirement whatsoever.

Optional:

$$\mathcal{O}[\![\mathsf{this.f} = \mathsf{e}]\!] = \mathsf{this.f} = \mathsf{e}' \quad \textbf{where} \quad \mathsf{e}' = \mathcal{O}[\![\mathsf{e}]\!]$$

Transient:

$$\mathcal{T}[\![\mathsf{this.f} = \mathsf{e}]\!]_\Gamma = \mathsf{this.f} = \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{this} : \mathsf{C} \quad \mathsf{f} : \mathsf{t} \in \mathsf{K}(\mathsf{C}) \quad \mathsf{e}' = \mathcal{T}(\!|\mathsf{e}|\!)^{\star}_{\Gamma}$$

Behavioral:

$$\mathcal{B}[\![\mathsf{this.f} = \mathsf{e}]\!]_\Gamma = \mathsf{this.f} = \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{this} : \mathsf{C} \quad \mathsf{f} : \mathsf{t} \in \mathsf{K}(\mathsf{C}) \quad \mathsf{e}' = \mathcal{B}(\!|\mathsf{e}|\!)^{\mathsf{t}}_{\Gamma}$$

Concrete:

$$\mathcal{C}[\![\mathsf{this.f} = \mathsf{e}]\!]_\Gamma = \mathsf{this.f} = \mathsf{e}' \quad \textbf{where} \quad \mathsf{K},\Gamma \vdash \mathsf{this} : \mathsf{C} \quad \mathsf{f} : \mathsf{t} \in \mathsf{K}(\mathsf{C}) \quad \mathsf{e}' = \mathcal{C}(\!|\mathsf{e}|\!)^{\mathsf{t}}_{\Gamma}$$

Figure 9 Translations for assignment.

The translation for object creation, shown in Fig. 10, follows the same reasoning. It inserts casts for each argument to be the required type according to class translation.

The translations for invocation are shown in Fig. 11. The optional approach translates all invocations to dynamic invocation, as it cannot provide any static guarantee. In the concrete and behavioral approaches, since the static types are retained, arguments must be asserted to have the statically known type. In the transient semantics, the argument type is ignored, so the argument to a statically typed method call is only required to be of type $\star$, but the return type is checked. In all of the systems, if the type of the receiver is $\star$, dynamic invocation is used.

Optional:

$$\mathcal{O}[\![\textbf{new } \mathsf{C}(\mathsf{e}_1..)]\!] = \langle\star\rangle\,\textbf{new } \mathsf{C}(\mathsf{e}'_1..) \quad \textbf{where } \mathsf{e}'_1 = \mathcal{O}[\![\mathsf{e}_1]\!] \; ..$$

Transient:

$$\mathcal{T}[\![\textbf{new } \mathsf{C}(\mathsf{e}_1..)]\!]_\Gamma = \textbf{new } \mathsf{C}(\mathsf{e}'_1..) \quad \textbf{where } \mathsf{f}_1 : \mathsf{t}_1 \in \mathsf{K}(\mathsf{C}) \;\; \mathsf{e}'_1 = \mathcal{T}(\!|\mathsf{e}_1|\!)^{\star}_{\Gamma} \; ..$$

Behavioral:

$$\mathcal{B}[\![\textbf{new } \mathsf{C}(\mathsf{e}_1..)]\!]_\Gamma = \textbf{new } \mathsf{C}(\mathsf{e}'_1..) \quad \textbf{where } \mathsf{f}_1 : \mathsf{t}_1 \in \mathsf{K}(\mathsf{C}) \;\; \mathsf{e}'_1 = \mathcal{B}(\!|\mathsf{e}_1|\!)^{\mathsf{t}_1}_{\Gamma} \; ..$$

Concrete:

$$\mathcal{C}[\![\textbf{new } \mathsf{C}(\mathsf{e}_1..)]\!]_\Gamma = \textbf{new } \mathsf{C}(\mathsf{e}'_1..) \quad \textbf{where } \mathsf{f}_1 : \mathsf{t}_1 \in \mathsf{K}(\mathsf{C}) \;\; \mathsf{e}'_1 = \mathcal{C}(\!|\mathsf{e}_1|\!)^{\mathsf{t}_1}_{\Gamma} \; ..$$

Figure 10 Translations for object creation.

Optional:

$$\mathcal{O}[\![\mathsf{e}_1.\mathsf{m}(\mathsf{e}_2)]\!] = \mathsf{e}'_1@\mathsf{m}_{\star\to\star}(\mathsf{e}'_2) \quad \textbf{where } \mathsf{e}'_1 = \mathcal{O}[\![\mathsf{e}_1]\!] \;\; \mathsf{e}'_2 = \mathcal{O}[\![\mathsf{e}_2]\!]$$

Transient:

$$\mathcal{T}[\![\mathsf{e}_1.\mathsf{m}(\mathsf{e}_2)]\!]_\Gamma = \mathsf{e}'_1@\mathsf{m}_{\star\to\star}(\mathsf{e}'_2) \quad \textbf{where } \mathsf{K},\Gamma \vdash \mathsf{e}_1 : \star \;\; \mathsf{e}'_1 = \mathcal{T}[\![\mathsf{e}_1]\!]_\Gamma \;\; \mathsf{e}'_2 = \mathcal{T}(\!|\mathsf{e}_2|\!)^{\star}_{\Gamma}$$

$$\mathcal{T}[\![\mathsf{e}_1.\mathsf{m}(\mathsf{e}_2)]\!]_\Gamma = \langle\mathsf{D}_2\rangle\,\mathsf{e}'_1.\mathsf{m}_{\star\to\star}(\mathsf{e}'_2) \quad \textbf{where } \mathsf{K},\Gamma \vdash \mathsf{e}_1 : \mathsf{C} \;\; \mathsf{e}'_1 = \mathcal{T}[\![\mathsf{e}_1]\!]_\Gamma \;\; \mathsf{e}'_2 = \mathcal{T}(\!|\mathsf{e}_2|\!)^{\star}_{\Gamma} \;\; \mathsf{m}(\mathsf{D}_1) : \mathsf{D}_2 \in \mathsf{K}(\mathsf{C})$$

Behavioral:

$$\mathcal{B}[\![\mathsf{e}_1.\mathsf{m}(\mathsf{e}_2)]\!]_\Gamma = \mathsf{e}'_1@\mathsf{m}_{\star\to\star}(\mathsf{e}'_2) \quad \textbf{where } \mathsf{K},\Gamma \vdash \mathsf{e}_1 : \star \;\; \mathsf{e}'_1 = \mathcal{B}[\![\mathsf{e}_1]\!]_\Gamma \;\; \mathsf{e}'_2 = \mathcal{B}(\!|\mathsf{e}_2|\!)^{\star}_{\Gamma}$$

$$\mathcal{B}[\![\mathsf{e}_1.\mathsf{m}(\mathsf{e}_2)]\!]_\Gamma = \mathsf{e}'_1.\mathsf{m}_{\mathsf{D}_1\to\mathsf{D}_2}(\mathsf{e}'_2) \quad \textbf{where } \mathsf{K},\Gamma \vdash \mathsf{e}_1 : \mathsf{C} \;\; \mathsf{e}'_1 = \mathcal{B}[\![\mathsf{e}_1]\!]_\Gamma \;\; \mathsf{e}'_2 = \mathcal{B}(\!|\mathsf{e}_2|\!)^{\mathsf{D}_1}_{\Gamma} \;\; \mathsf{m}(\mathsf{D}_1) : \mathsf{D}_2 \in \mathsf{K}(\mathsf{C})$$

Concrete:

$$\mathcal{C}[\![\mathsf{e}_1.\mathsf{m}(\mathsf{e}_2)]\!]_\Gamma = \mathsf{e}'_1@\mathsf{m}_{\star\to\star}(\mathsf{e}'_2) \quad \textbf{where } \mathsf{K},\Gamma \vdash \mathsf{e}_1 : \star \;\; \mathsf{e}'_1 = \mathcal{C}[\![\mathsf{e}_1]\!]_\Gamma \;\; \mathsf{e}'_2 = \mathcal{C}(\!|\mathsf{e}_2|\!)^{\star}_{\Gamma}$$

$$\mathcal{C}[\![\mathsf{e}_1.\mathsf{m}(\mathsf{e}_2)]\!]_\Gamma = \mathsf{e}'_1.\mathsf{m}_{\mathsf{D}_1\to\mathsf{D}_2}(\mathsf{e}'_2) \quad \textbf{where } \mathsf{K},\Gamma \vdash \mathsf{e}_1 : \mathsf{C} \;\; \mathsf{e}'_1 = \mathcal{C}[\![\mathsf{e}_1]\!]_\Gamma \;\; \mathsf{e}'_2 = \mathcal{C}(\!|\mathsf{e}_2|\!)^{\mathsf{D}_1}_{\Gamma} \;\; \mathsf{m}(\mathsf{D}_1) : \mathsf{D}_2 \in \mathsf{K}(\mathsf{C})$$

Figure 11 Translations for function invocation.

5.3 Example

We illustrate the translation with the behavior of litmus program **L3**. The operational principle of **L3** is that it creates a new object (an instance of C), then uses an untyped intermediate to represent it as type E. Type E ascribes the wrong type for argument x, substituting D for the correct type C.

Source:

$\textbf{class}\ \mathsf{F}\,\{\ \mathsf{m(x: E): E}\ \{\mathsf{x}\}\quad \mathsf{n(x:\star):\star}\ \{\mathsf{this.m(x)}\}\ \}$

Optional:

$\textbf{class}\ \mathsf{F}\,\{\ \mathsf{m(x:\star):\star}\ \{\mathsf{x}\}\quad \mathsf{n(x:\star):\star}\ \{(\langle\star\rangle\,\mathsf{this})@\mathsf{m}_{\star\to\star}(\mathsf{x})\}\ \}$

Transient:

$\textbf{class}\ \mathsf{F}\,\{\ \mathsf{m(x:\star):\star}\ \{\langle\mathsf{E}\rangle\,\mathsf{x};\ \langle\star\rangle\,\mathsf{x}\}\quad \mathsf{n(x:\star):\star}\ \{\langle\star\rangle\,\mathsf{x};\ \langle\star\rangle\,\langle\mathsf{E}\rangle\,\mathsf{this.m}_{\star\to\star}(\langle\star\rangle\,\langle\star\rangle\,\mathsf{x})\}\ \}$

Behavioral:

$\textbf{class}\ \mathsf{F}\,\{\ \mathsf{m(x: E): E}\ \{\mathsf{x}\}\quad \mathsf{n(x:\star):\star}\ \{\blacktriangleleft\star\blacktriangleright\,\mathsf{this.m}_{\mathsf{E}\to\mathsf{E}}(\blacktriangleleft\mathsf{E}\blacktriangleright\,\mathsf{x})\}\ \}$

Concrete:

$\textbf{class}\ \mathsf{F}\,\{\ \mathsf{m(x: E): E}\ \{\mathsf{x}\}\quad \mathsf{n(x:\star):\star}\ \{\langle\star\rangle\,\mathsf{this.m}_{\mathsf{E}\to\mathsf{E}}(\langle\mathsf{E}\rangle\,\mathsf{x})\}\ \}$

Figure 12 Class translation for litmus test L3.

Two of the gradual type systems notice this invalid type. Concrete errors on **L3** because E is not a subtype of C. With behavioral, the unused type ascription is saved as a wrapper and is enforced causing a run-time error.

While this reasoning provides an intuition, it provides few detail for which we turn to our formalism. We present the translation from the top, starting with classes in Fig. 12. The optional approach does no checking whatsoever, and simply erases types. Transient also erases types, but adds argument casts on method entry. In the case of m, argument x is checked to be of type E, as the translation of type E does not include types no type error will be reported. Behavioral retains typed methods but adds behavioral casts on untyped methods. The concrete semantics retains typed methods, and adds a subtype cast when a variable of type $\star$ is passed to a method that expects and E.

Fig. 13 presents the translation of class E. For the transient semantics, when x is cast to E, all of the types on E are erased. Casting to E is tantamount to asking for the existence of the method m. In contrast, the concrete semantics retains the types of m. A cast to E is equivalent to checking if a method m that takes and returns an E exists. This comes at the cost of the ability to migrate between untyped and typed code. Suppose that both the optional and concrete versions of E existed, under a different name F. In that program, only the concrete version of E could be used with the concrete version of F. Despite implementing the same behavior, Behavioral uses the same representation for E as concrete. The behavioral cast allows to use any value that behaves like an E.

To examine the operation of the behavioral cast in more detail, Fig. 14 depicts the wrapper classes generated at the cast from C to $\star$ and from it to E. Class C_1 takes an instance of C and makes it safe against use as $\star$. In behavioral, no typed invocations can be made on

Source:

class E { m(x: D): D {x} }

Optional:

class E { m(x: ⋆): ⋆ {x} }

Transient:

class E { m(x: ⋆): ⋆ {⟨E⟩ x; ⟨⋆⟩ x} }

Behavioral:

class E { m(x: E): E {x} }

Concrete:

class E { m(x: E): E {x} }

Figure 13 Translation of E in litmus test L3.

```
class C {  a(x: C): C {x} }                    class E {  a(x: D): D {x} }

  class C1 {   that : C        a(x: ⋆): ⋆ {◀⋆▶ this.that.a_{C→C}(◀C▶ x)} }
  class C2 {   that : C1       a(x: D): D {◀D▶ this.that.a_{⋆→⋆}(◀⋆▶ x)} }
```

Figure 14 Behavioral wrappers.

a value that was cast to $\star$ (and not cast to some type somewhere); only untyped invocations are allowed. As a result, the wrapper need only generate an untyped version of C's method a, which calls the underlying C instance's a (adding suitable casts). The second wrapper class C_2 takes the C_1 wrapper and casts it back to E. This wrapper takes the untyped implementation of a and wraps it again, calling it with an argument cast to $\star$ and casting the return to D.

5.4 Discussion

These translations explicit the enforcement machinery of each of the four approaches. Programs can get stuck at dynamic invocation and casts. Inspecting where these are inserted gives a precise account of what constitutes an error in each gradual type system.

Our account of the behavioral approach matches its implementation in Typed Racket. However, one could imagine a slightly less restrictive implementation, one which does not have a check for method names at wrapper creation. That check is pragmatic but perhaps too strict – it will rule out programs that may be fine just because a method is missing. One could have a wrapper that simply reports an error if a missing method is called.

Performance is a perennial worry for implementers of gradual type systems. It is difficult to guess how a highly optimizing language implementation will perform, as these implementations are likely to optimize away the majority of the casts and dynamic dispatches. Consider the progress in the performance of Typed Racket reported in the literature [22, 2]. What we can tell by looking at the translations is that with optional there is no obvious benefit or cost to having type annotations. Transient has checks on reads, which are common, and typed function calls. Furthermore, those checks are needed even if the entire program is typed. Both concrete and behavioral can benefit from type information in typed code. The difference is that the cost of boundary crossing are low in the concrete approach, as it uses a subtype check whereas behavioral requires allocation of a wrapper. Wrappers may also complicate the task of devirtualization and unboxing.

6 Conclusion

This paper introduced KafKa, a framework for comparing the design of gradual type systems for object-oriented languages. Our approach is to provide translations with different gradual semantics from a common surface language into KafKa. These translations highlight the different run-time enforcement strategies deployed by the languages under study. The differences between gradual type systems are highlighted explicitly by the observable differences of their behavior in our litmus tests, demonstrating how there is no consensus on the meaning of error. These litmus tests motivated the need to have a common framework to explore the design space.

KafKa demonstrates that to express the different gradual approaches, one needs a calculus with two casts (structural and behavioral), two invocation forms (dynamic and static), the ability to extend the class table at run-time, and wrappers that expose their underlying unwrapped methods. We provide a mechanized proof of soundness for KafKa that includes run-time class generation. We also demonstrate that KafKa can be straightforwardly implemented on top of a stock virtual machine.

A open question for gradual type system designers is performance of the resulting implementation. Performance remains a major obstacle to adoption of approaches that attempt to provide soundness guarantees. Under the optional approach, types are removed by the translation; as a result, performance will be identical to that of untyped code. The transient approach checks types at uses, so the act of adding types to a program introduces more casts and may slow the program down (even in fully typed code). In contrast, the behavioral approach avoids casts in typed code. The price it pays for this soundness, however, is that heavyweight wrappers inserted at typed-untyped boundaries. Lastly, the concrete semantics is also sound and has low overheads, but comes at a cost in expressiveness and ability to migrate from untyped to typed.

Going forward, there are several issues we wish to investigate. We do not envision that supporting nominal subtyping within KafKa will pose problems, it would only take adding a nominal cast and changing the definition of classes. Then nominal and structural could coexist. A more challenging question is how to handle the intricate semantics of Monotonic Reticulated Python. For these we would need a somewhat more powerful cast operation. Rather than building each new cast into the calculus itself, it would be interesting to axiomatize the correctness requirements for a cast and let users define their own cast semantics. The goal would be to have a collection of user defined pluggable casts within a single framework.

References

1. Esteban Allende, Oscar Callaú, Johan Fabry, Éric Tanter, and Marcus Denker. Gradual typing for Smalltalk. *Science of Computer Programming*, 96, 2014. doi:10.1016/j.scico.2013.06.006.
2. Spenser Bauman, Carl Friedrich Bolz-Tereick, Jeremy Siek, and Sam Tobin-Hochstadt. Sound gradual typing: Only mostly dead. *Proc. ACM Program. Lang.*, 1(OOPSLA), 2017. doi:10.1145/3133878.
3. Gavin Bierman, Martin Abadi, and Mads Torgersen. Understanding TypeScript. In *European Conference on Object-Oriented Programming (ECOOP)*, 2014. doi:10.1007/978-3-662-44202-9_11.

4 Gavin M. Bierman, Erik Meijer, and Mads Torgersen. Adding dynamic types to C#. In *European Conference on Object-Oriented Programming (ECOOP)*, 2010. doi:10.1007/978-3-642-14107-2_5.
5 Bard Bloom, John Field, Nathaniel Nystrom, Johan Östlund, Gregor Richards, Rok Strnisa, Jan Vitek, and Tobias Wrigstad. Thorn: Robust, concurrent, extensible scripting on the JVM. In *Conference on Object Oriented Programming Systems Languages and Applications (OOPSLA)*, 2009. doi:10.1145/1639950.1640016.
6 Gilad Bracha. Pluggable type systems. In *OOPSLA 2004 Workshop on Revival of Dynamic Languages*, 2004. doi:10.1145/1167473.1167479.
7 Gilad Bracha and David Griswold. Strongtalk: Typechecking Smalltalk in a production environment. In *Conference on Object Oriented Programming Systems Languages and Applications (OOPSLA)*, 1993. doi:10.1145/165854.165893.
8 Luca Cardelli. Amber. In *LITP Spring School on Theoretical Computer Science*, pages 21–47. Springer, 1985.
9 Gilles Dubochet and Martin Odersky. Compiling structural types on the JVM: A comparison of reflective and generative techniques from Scala's perspective. In *Workshop on the Implementation, Compilation, Optimization of Object-Oriented Languages and Programming Systems (ICOOLPS)*, 2009. doi:10.1145/1565824.1565829.
10 Robert Findler and Matthias Felleisen. Contracts for higher-order functions. In *Proceedings of the ACM SIGPLAN International Conference on Functional Programming*, 2002. doi:10.1145/581478.581484.
11 Robert Bruce Findler, Matthew Flatt, and Matthias Felleisen. Semantic casts: Contracts and structural subtyping in a nominal world. In *European Conference on Object-Oriented Programming (ECOOP)*, 2004. doi:https://doi.org/10.1007/978-3-540-24851-4_17.
12 Ben Greenman and Matthias Felleisen. A spectrum of soundness and performance. *Proc. ACM PL (ICFP), to appear*, 2018.
13 Atsushi Igarashi, Benjamin C. Pierce, and Philip Wadler. Featherweight Java: a minimal core calculus for Java and GJ. *ACM Trans. Program. Lang. Syst.*, 23(3), 2001. doi:10.1145/503502.503505.
14 Timothy Jones and David J. Pearce. A mechanical soundness proof for subtyping over recursive types. In *Workshop on Formal Techniques for Java-like Programs (FTfJP)*, 2016. doi:10.1145/2955811.2955812.
15 Fabian Muehlboeck and Ross Tate. Sound gradual typing is nominally alive and well. *Proc. ACM Program. Lang.*, 1(OOPSLA), 2017. doi:10.1145/3133880.
16 Benjamin C. Pierce and David N. Turner. Local type inference. In *Symposium on Principles of Programming Languages (POPL)*, 1998. doi:10.1145/345099.345100.
17 Gregor Richards, Ellen Arteca, and Alexi Turcotte. The VM already knew that: Leveraging compile-time knowledge to optimize gradual typing. *Proc. ACM Program. Lang.*, 1(OOPSLA), 2017. doi:10.1145/3133879.
18 Gregor Richards, Francesco Zappa Nardelli, and Jan Vitek. Concrete types for TypeScript. In *European Conference on Object-Oriented Programming (ECOOP)*, 2015. doi:10.4230/LIPIcs.ECOOP.2015.76.
19 Jeremy Siek. Gradual typing for functional languages. In *Scheme and Functional Programming Workshop*, 2006. URL: http://ecee.colorado.edu/~siek/pubs/pubs/2006/siek06_gradual.pdf.
20 Jeremy Siek and Walid Taha. Gradual typing for objects. In *European Conference on Object-Oriented Programming (ECOOP)*, 2007. doi:10.1007/978-3-540-73589-2_2.
21 Jeremy G. Siek, Michael M. Vitousek, Matteo Cimini, and John Tang Boyland. Refined criteria for gradual typing. In *Summit on Advances in Programming Languages (SNAPL)*, 2015. doi:10.4230/LIPIcs.SNAPL.2015.274.

22 Asumu Takikawa, Daniel Feltey, Ben Greenman, Max New, Jan Vitek, and Matthias Felleisen. Is sound gradual typing dead? In *Symposium on Principles of Programming Languages (POPL)*, 2016. doi:10.1145/2837614.2837630.

23 Asumu Takikawa, T. Stephen Strickland, Christos Dimoulas, Sam Tobin-Hochstadt, and Matthias Felleisen. Gradual typing for first-class classes. In *Conference on Object Oriented Programming Systems Languages and Applications (OOPSLA)*, 2012. doi:10.1145/2398857.2384674.

24 The Dart Team. Dart programming language specification, 2016. URL: http://dartlang.org.

25 The Facebook Hack Team. Hack, 2016. URL: http://hacklang.org.

26 Sam Tobin-Hochstadt and Matthias Felleisen. Interlanguage migration: from scripts to programs. In *Symposium on Dynamic languages (DLS)*, 2006. doi:10.1145/1176617.1176755.

27 Sam Tobin-Hochstadt and Matthias Felleisen. The design and implementation of typed Scheme. In *Symposium on Principles of Programming Languages (POPL)*, 2008. doi:10.1145/1328438.1328486.

28 Michael Vitousek, Andrew Kent, Jeremy Siek, and Jim Baker. Design and evaluation of gradual typing for Python. In *Symposium on Dynamic languages (DLS)*, 2014. doi:10.1145/2661088.2661101.

29 Michael M. Vitousek, Cameron Swords, and Jeremy G. Siek. Big types in little runtime: Open-world soundness and collaborative blame for gradual type systems. In *Symposium on Principles of Programming Languages (POPL)*, 2017. doi:10.1145/3009837.3009849.

30 Tobias Wrigstad, Francesco Zappa Nardelli, Sylvain Lebresne, Johan Östlund, and Jan Vitek. Integrating typed and untyped code in a scripting language. In *Symposium on Principles of Programming Languages (POPL)*, 2010. doi:10.1145/1706299.1706343.