Behavioral Subtyping, Specification Inheritance, and Modular Reasoning

Gary T. Leavens
Iowa State University

David A. Naumann
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/cs_techreports

Part of the Programming Languages and Compilers Commons, Software Engineering Commons, and the Theory and Algorithms Commons

Recommended Citation
http://lib.dr.iastate.edu/cs_techreports/280
Behavioral Subtyping, Specification Inheritance, and Modular Reasoning
Gary T. Leavens and David A. Naumann

TR #06-20b
July 21, 2006, revised Aug 4, Sept. 3 2006

Keywords: Behavioral subtyping, supertype abstraction, specification inheritance, modularity, specification, verification, state transformer, dynamic dispatch, invariants, Eiffel language, JML language.


Copyright © 2006 by Gary T. Leavens and David A. Naumann.
Submitted for publication

Department of Computer Science
226 Atanasoff Hall
Iowa State University
Ames, Iowa 50011-1041, USA
Behavioral Subtyping, Specification Inheritance, and Modular Reasoning

Gary T. Leavens *
Iowa State University
Ames, IA 50011 USA
leavens@cs.iastate.edu

David A. Naumann †
Stevens Institute of Technology
Hoboken, NJ 07030 USA
naumann@cs.stevens.edu

Abstract
Behavioral subtyping is an established idea that enables modular reasoning about behavioral properties of object-oriented programs. It requires that syntactic supertypes are behavioral refinements. It validates reasoning about a dynamically-dispatched method call, say \(E.m()\), using the specification associated with the static type of the receiver expression \(E\). For languages with references and mutable objects the idea of behavioral subtyping has not been rigorously formalized as such, the standard informal notion has inadequacies, and exact definitions are not obvious. This paper formalizes behavioral subtyping and supertype abstraction for a Java-like sequential language with classes, interfaces, exceptions, mutable heap objects, references, and recursive types. Behavioral subtyping is proved sound and semantically complete for reasoning with supertype abstraction. Specification inheritance, as used in the specification language JML, is formalized and proved to entail behavioral subtyping.

1. Introduction
In object-oriented (OO) programming, subtyping and dynamic dispatch are both useful and problematic. They are useful because supertypes can abstract away details in the specifications of their subtypes, thus allowing variations in data structures and algorithms to be handled uniformly. They are problematic for modular reasoning because a dynamically-dispatched method call such as \(E.m()\) seems to require a case analysis to deal with all possible dynamic types of \(E\)’s value. The basic idea of modular reasoning, which we call supertype abstraction [22], is clear. It is a generalization of typechecking: reasoning about an invocation, say \(E.m()\), is based on the specification associated with the static type of \(E\), and constraints are imposed on implementations of \(m\) at all subtypes. While modular type safety conditions for dynamically-dispatched methods are well-known [12], a straightforward translation into conditions on overriding method specifications, while sound, is more restrictive than necessary. The translation also gives no help in reasoning about object invariants, which need to be strengthened in subtypes. Hence, for modular reasoning one needs a behavioral notion of subtyping.

Remarkably, there is no mathematically rigorous account of behavioral subtyping and its connection with modular reasoning about specifications and programs in conventional OO programming languages — although there has been much study [1,4,11,14,15,19,22,29,30,47] (see [21] for a survey). Some of the current understanding of behavioral subtyping is embodied in program logics [31,49,42,43,46] but is difficult to disentangle from various other complications. Of the current understanding is embodied languages and tools such as Eiffel [30], JML [20], ESC/Java [16], and Spec# [8,7]. But these have unsoundnesses and incompletenesses, some by engineering design and some for lack of adequate theory and methodology. A key source of unsoundness is naive treatment of object invariants, because many important design patterns invalidate the simple hierarchical notion of encapsulation on which the standard treatment [17] is based.

On one hand, behavioral subtyping has been rigorously studied in restrictive programming models (e.g., [19,47]). On the other hand, various embodiments have been implemented in static and runtime verification tools and logics that apply to rich specification and programming languages such as Java and JML [20] and Eiffel [30]. Our goal is to close the gap by providing a rigorous analysis on which can be based more specialized assessments and justifications of specific tools and logics. (With the ultimate aim of high assurance for verification tools, we have undertaken to machine-check our results.)

We believe the gap has remained because it was far from clear how to formalize a general theory that pertains directly to reasoning about code in languages of practical interest. The semantic intricacies of the languages — and of current methodologies for sound reasoning about invariants, heap encapsulation and locality of effects, etc.— are daunting. The details of the OO language are important, because some language features, such as reflection, allow programs to make observations that can distinguish between supertype and subtype objects. The achievements closest to our aim are soundness and completeness proofs for logics of Java fragments that embody supertype abstraction in some form (e.g., [19,43]). But these assess the reasoning power of a proof system, rather than assessing and explicating the connection between behavioral subtyping and supertype abstraction; and they are somewhat removed from the axiomatic semantics of some widely used verifiers (which simply postulate soundness of behavioral subtyping in some form).

The key insight that led to our results is a purely semantic formulation of supertype abstraction using two denotational semantics: in one, method calls are statically dispatched. On this basis we are able to give a formal treatment in a language with many constructs of sequential OO languages, including classes, interfaces, mutable heap objects, assignment, exceptions, inheritance, visibility, reference equality, type tests, and recursive types. Even with effective definitions in hand, it was non-trivial to find the right induction hypothesis and technique to prove the main lemma.

We make no commitment to particular specification notations or reasoning system but rather formulate modular reasoning semantically in a generic way that idealizes what is found in logics and tools. Using an operationally sound compositional semantics allows us to provide a foundation that will serve as a point of reference and as a basis for assessment and further development of specification languages and verification tools.

This paper makes the following contributions.

- We give a semantic characterization of supertype abstraction, which idealizes what is found in logics and verification tools. In contrast to related work, our definition does not rely on derived notions such as substituting one object for another [19,27,29], nor is it tied to a proof system [31,40,42,43,46].
• We formalize behavioral subtyping in terms of refinement of observable behavior in a realistic programming model. Refinement does not need to hold between all syntactically related types but only when the subtype is a (non-abstract) class.

• In contrast to the standard view [29], we define refinement intrinsically, by quantifying over satisfying implementations. Separately, we characterize refinement in terms of relations between pre- and post-conditions. Our characterization adapts previous work [13, 15, 41] that improves on the overly restrictive standard condition of postcondition implications [2, 8, 15, 29, 30]. (This also isolates the way in which characterization of completeness depends on the specification language, see Sect. 7.) An outcome of our focus on reasoning about correctness of programs rather than an abstract model is that we find abstraction functions are not an integral part of behavioral subtyping (compare, e.g., [19, 29]).

• We prove soundness and semantic completeness of behavioral subtyping for supertype abstraction. It was by seeking completeness that we were led to the surprising findings noted in the preceding items.

• We justify the standard condition on invariants [29], by precisely describing the necessary soundness conditions and how they can be enforced by recently proposed techniques for sound modular reasoning about invariants [24, 37, 32].

• We formalize specification inheritance [50] and show that it ensures behavioral subtyping. Specification inheritance is part of the semantic definition of JML, and it embodies the proof obligations whereby some logics achieve behavioral subtyping [31, 40, 42, 43, 46].

The rest of this paper is organized as follows. The next section gives a synopsis of the main ideas. We then discuss related work. Following that are technical details about the language (Sect. 3) and specifications (Sect. 4). This is followed by a formalization of supertype abstraction and behavioral subtyping (Sect. 5), with the main theorems connecting them. We then extend these results to handle invariants (Sect. 6). In Sect. 7 we characterize behavioral subtyping by a checkable condition on specifications and prove that specification inheritance ensures behavioral subtyping. Sect. 8 concludes.

2. Synopsis

In an OO language it is not obvious how to do modular reasoning, because dynamic dispatch selects different methods depending on the exact run time type of an object. What specification should one use to reason about a call, such as \( E.m() \), given that the static type of \( E \), say \( T \), is only an upper bound on its dynamic type? While we consider formal specifications and reasoning in this paper, the problem also applies to informal reasoning based on informal specifications. The first expert OO programmers used \( T \)’s specification of \( m \) to reason about such calls. This kind of reasoning, supertype abstraction [22], is modular in that it does not depend on \( E \)’s dynamic type, and hence does not have to be changed when subtypes of \( T \) are changed in compatible ways or are added to a program [27]. Supertype abstraction supports maintenance and evolutionary programming styles.

However, supertype abstraction is only valid if methods that override \( T \)’s method \( m \) satisfy \( T \)’s specification for \( m \), since that specification is the one used in reasoning about such calls. Making overrides obey the specification of overridden methods, specification inheritance [14, 50], ensures that objects of subtypes of \( T \) do not cause surprising behavior when treated as if they are objects of type \( T \); that is, it ensures behavioral subtyping [21, 31]. An alternative is to check the given specifications and treat violations as a design error [15]. Either way, a characterization is needed for soundness and to avoid unnecessary restriction.

2.1 Supertype abstraction

An influential discussion of the benefits of these ideas is Liskov’s invited talk at OOPSLA 1987 [27]. Liskov stated an easily-remembered test for behavioral subtyping (p. 25): “If for each object \( o_1 \) of type \( S \) there is an object \( o_2 \) of type \( T \) such that for all programs \( P \) defined in terms of \( T \), the behavior of \( P \) is unchanged when \( o_1 \) is substituted for \( o_2 \), then \( S \) is a supertype of \( T \).” This is often called the “Liskov Substitutability Principle” (LSP) and is a strong form of supertype abstraction. The LSP is actually too strong, because it uses the notion of “unchanged” behavior; the point of introducing supertype objects is often to change behavior in a way that is allowed by the supertype’s specification. A more flexible intuition defines observations that are not allowed by this specification as “surprising,” and says that behavioral “subtyping prevents surprising behavior” [19, Chapter 1].

As a formulation of supertype abstraction, the LSP is not easy to apply to imperative OO languages. It is not clear what it means to substitute one object for another: imperative programs are not referentially transparent, object identity matters, and the state of “an object” often depends on other objects in the heap. (Object identity is also a problem with the algebraic work on applicative languages from which the LSP is drawn [11, 19].) One of the contributions of our paper is to precisely formalize supertype abstraction in a way that captures modular reasoning about imperative OO programs.

Our definition of supertype abstraction says that properties of a command can be proved by reasoning about its method calls as if they were statically dispatched to an arbitrary implementation that satisfies the specification associated with the static type of the receiver. The definition is in terms of a denotational semantics. For a command \( S \), its meaning, \( D[S] \), is interpreted in a method environment, \( \mu \), that gives a meaning to each method at each type. Thus \( D[S] \mu \) is a function from initial states to final states. We are interested in proving that \( S \) satisfies some pre/post specification, \( \text{spec} \). Modular reasoning proves that \( D[S] \mu \) satisfies \( \text{spec} \), using only the specifications associated with the static types of receivers of method calls. We call the latter a specification table. To avoid formalizing “reasoning” as such, our definition considers semantic consequences that hold for any \( \mu \) that satisfies the specification table. Whereas the actual program semantics, \( D[-] \), uses dynamic dispatch, we define a static dispatch semantics, \( S[-] \), to formalize reasoning in terms of static types. Supertype abstraction is formalized roughly as follows: if it is true that \( S[S]\mu \) satisfies \( \text{spec} \), then the actual semantics \( D[S]\mu \) satisfies \( \text{spec} \) (provided that \( \mu \) satisfies the specification table). (For details see Def. 17.)

In a program logic, supertype abstraction is embodied by the proof rule for method invocation, which allows to derive \( \{ P \} \ E.m() \{ Q \} \) only from a specification \( (\text{pre}_m^T, \text{post}_m^T) \) associated with the static type, \( T \), of \( E \). Similarly, an automated verifier typically uses weakest precondition semantics and achieves modularity by replacing a call \( E.m() \) by the sequence “\( \text{assert} \ \text{pre}_m^T; \ \text{assume} \ \text{post}_m^T \)” (with various optimizations, e.g., [23]). Both techniques aim to produce sound conclusions about the actual semantics. We model both by the static-dispatch semantics \( S[E.m()_\mu] \). What makes both techniques sound is behavioral subtyping, imposed by proof obligations on implementations of \( m \) in subtypes of \( T \) (typically via some form of specification inheritance). The proof obligations are modeled by our specification table: behavioral subtyping, i.e., refinement of specifications, is a property of the specification table. The main result of the paper (Thms. 12, 24) says that supertype abstraction for commands is equivalent to behavioral subtyping.

The proof that behavioral subtyping implies supertype abstraction goes by induction over the syntax of the language and is the paper’s hardest technical result (Lemma 16).

2.2 Behavioral subtyping

Several authors have offered definitions of behavioral subtyping, but the most influential definition has been Liskov and Wing’s [29]. We paraphrase part of their “constraint rule” (from their Figure 4, page 1823) [3]. For type \( T \) to be a behavioral subtype of \( U \):

\[ \begin{align*}
\text{for each object } o_1 \text{ of type } S \text{ there is an object } o_2 \text{ of type } T \text{ such that for all programs } P \text{ defined in terms of } T, \text{ the behavior of } P \text{ is unchanged when } o_1 \text{ is substituted for } o_2, \text{ then } S \text{ is a supertype of } T. \end{align*} \]

\[ \begin{align*}
\text{We resist the temptation to abstract the specification table as a single method environment that uses nondeterminacy to represent a specification by the “least refined implementation”; though elegant and useful (e.g., [30, 43]), this technique requires justification with respect to the actual program semantics.} \end{align*} \]

\[ \begin{align*}
\text{Their rule allows for method renamings and an abstraction function, our semantics of OO languages does not allow the former and allows us to disregard the latter. We also ignore the part of the rule about history constraints, since we study a sequential language.} \end{align*} \]
interface Tracker {
    public model goal, curr: int;
    public invariant 0< goal ∧ goal ≤ curr;
    // ...
    method lose(kg: int)
        requires goal ≤ curr − kg;
        ensures exc = null ∧ curr = old(curr − kg);
    }

class WeightLoss extends Object implements Tracker {
    protected g, c: int;
    protected invariant 0< g ∧ g ≤ c;
    protected represents goal := g;
    protected represents curr := c;
    // ...
    method lose(kg: int)
        requires true;
        ensures (old(goal ≤ curr − kg)
            ⇒ (exc = null ∧ curr = old(curr − kg)))
            ∧ (old(goal > curr − kg) ⇒ (exc = null));
        if (goal ≤ c − kg) then setCurr(curr − kg) else throw new IAE();
    }
    method setCurr(kg: int)
        requires goal ≤ kg;
        ensures curr = null ∧ curr = kg;
    }

Figure 1. The interface Tracker and the class WeightLoss.

• The subtype’s invariant must imply the supertype’s. That is, whenever
  T’s invariant holds for a subtype object then U’s invariant holds:

\[ \text{inv}^\text{T}(\text{self}) \Rightarrow \text{inv}^\text{U}(\text{self}) \]  \hspace{1cm} (1)

• “Subtype methods preserve the supertype method’s behavior.” That
  is, if T’s method m overrides U’s method m, then the usual static
  typing conditions hold, and for all subtype objects self: T, U’s
  precondition for m implies T’s precondition:

\[ \text{pre}^\text{U}(\text{self}) \Rightarrow \text{pre}^\text{T}(\text{self}) \]  \hspace{1cm} (2)

and T’s postcondition implies U’s:

\[ \text{post}^\text{U}(\text{self}) \Rightarrow \text{post}^\text{T}(\text{self}). \]  \hspace{1cm} (3)

The above definition is intended to be part of a “descriptive and
informal” presentation (p. 1813), which concentrates on ideas and has
only “informal justifications” (p. 1813). However, even at a conceptual
level, it has several problems.

Liskov and Wing’s postcondition rule \([\text{3}]\) is stronger (i.e., less
flexible) than necessary for the soundness of supertype abstraction \([12]\). To
see this, consider Fig. \([1]\). The interface Tracker declares two model fields,
g and curr that are only used in specifications \([25]\). They stand for the
goal of a diet and the current weight. The values of these model fields
are given by the represents clauses in class WeightLoss. The clause
“represents goal := g” declares part of the object invariant for
WeightLoss, which says that goal = self.g. This is a predicate on the
protected field and the inherited public model field. This is typical of
JML and other specification languages, where the connection between a
data representation and an abstraction are expressed by a hidden invariant.
Frame axioms (modifies clauses) could be encoded in the postcondition,
but for brevity we ignore them in our examples.

The loose method of the class WeightLoss illustrates the problem with
the postcondition rule \([\text{3}]\). That rule requires that lose’s postcondition
implies the postcondition in Tracker. However, this implication does not
hold for this example, as can be seen by considering the case where goal

\footnote{Liskov and Wing state the invariant rule for object “values,” not for objects.}
is strictly larger than the difference curr − kg, since exc ≠ null contradicts
exc = null. (The specification variable exc refers to exception results;
when exc is null in a post-state, the method returned normally.) How-
ever, in this example, supertype abstraction works fine, because the lose
method of the class WeightLoss obeys Tracker’s specification for lose
whenever Tracker’s precondition holds. Clients that reason about calls
to lose using Tracker’s specification will not be surprised, WeightLoss
be a behavioral subtype of Tracker. One way of expressing the most
flexible sound rule for U to be a behavioral subtype of T is to require that
for all subtype objects self: U, the precondition rule \([\text{2}]\) holds and

\[ \text{old}(\text{pre}^\text{U}(\text{self})) \land \text{post}^\text{U}(\text{self}) \Rightarrow \text{post}^\text{U}(\text{self}). \]  \hspace{1cm} (4)

where old(P) refers to the value of predicate P in the pre-state of a call.
The idea is that the behavior of the subtype’s method does not need to be
constrained when the supertype’s precondition does not hold in the
pre-state.

These conditions are just approximations of refinement and it is re-
finement that explains the equivalence between behavioral subtyping and
supertype abstraction (Theorem \([12]\)). Our formulation of behavioral sub-
typing (Def \([3]\) is in terms of the intrinsic refinement order on specifi-
cations. In Sect. \([7]\) we confirm that \([3]\) characterizes refinement in our
setting and can thus be used to check behavioral subtyping. Disentan-
gling the two is important since the characterization is sensitive to the
form of specifications as discussed in Sect. \([7]\).

2.3 Invariants and behavioral subtyping

While Liskov and Wing’s postcondition rule is sound but too restrictive,
their invariant rule is unsound if applied to general OO programs, unless
further restrictions on programs are applied. Object invariants are predi-
cates that can be assumed in method pre-states and must be established in
method post-states \([17]\). Liskov and Wing’s discussion implies that “the
invariant” is what the programmer declares and reasons about. This is
broken in unrestricted OO languages, because of sharing and reentrance
\([21]\). We illustrate these problems below. We also review rec-
cent methodologies for sound reasoning about invariants; in brief, the
methodologies derive from the declared invariants an effective invariant
for which Liskov and Wing’s invariant rule does work.

To explain the problems, we begin by considering how invariants
interact with pre- and postconditions. Invariants could be thought of as
conjoined to the pre- and postcondition of each method’s specification,
because invariants must hold at the beginning and end of each method’s
execution; this is the usual proof obligation. This conjunction semantics
would cause a conflict between the invariant rule \([\text{1}]\) and the precondition
rule \([\text{2}]\), as the invariant may be strengthened in a subtype (covariantly),
which will strengthen a precondition it is conjoined with, which in turn
violates the rule that preconditions may only be weakened in a subtype
(contravariantly). The standard solution to this problem \([17]\) ensures
that the implementation can assume the invariant as a precondition, by
restricting the invariant to depend only on part of the state that is within
an encapsulation boundary, say private to the class, so that it cannot be
falsified by clients between calls to methods of the class. For example,
consider the invariant in Fig. \([11]\) if field c in Fig. \([1]\) were public, instead
of protected, then clients could assign -1 to it, falsifying the invariant.
Indeed, one would like per-instance encapsulation, so that an invariant
for object o can only be invalidated by a method execution whose
receiver is o. But privacy in Java-like languages is at the level of classes,
not instances, and in this example field c is visible in subclasses.
The general rule is that if a mutable field is visible in some context, say
a client or subclass, then that code must be responsible for maintaining
invariants that depend on the field (and therefore such invariants must be
visible).

Encapsulation is partly achieved using scope and visibility, but that is
not enough, due to the problem of sharing mutable objects, between
object’s representation and client code. This is the problem of repre-
sentation exposure \([25, 22, 38]\). To see the problem, imagine that c in Fig. \([11]\)
instead of containing an integer, referred to a mutable IntCell object:

\text{protected} c: \text{IntCell};

Imagine the IntCell objects contain a (model) field val that can be
changed in a program and that the invariant is changed to refer to it:
protected invariant \(0 \leq \text{self.g} \land \text{self.g} \leq \text{self.c} \land \text{self.d} = \text{self.c} - \text{self.g};\)

\[
\text{protected invariant } 0 \leq \text{self.g} \land \text{self.g} \leq \text{self.c} \cdot \text{self.val};
\]

The problem is that the IntCell object \(c\) can be shared between this version of the WeightLoss class and clients. This allows clients to mutate \(c\), potentially violating the invariant. Liskov and Wing’s rule is thus only sound if objects cannot contain mutable subobjects, or if some other restrictions on ownership are applied [24][27][32].

Reentrance is another source of unsoundness for invariants. The class WghtDiff shown in Fig. 2 illustrates the problem. Consider a call such as \(wld.lose(5)\), where \(wld\) is an object of dynamic type \(WghtDiff\). In this case, the inherited lose method of WeightLoss is called, which reenters WghtDiff by calling its setCurr method. WghtDiff’s setCurr method makes a super call up to WeightLoss’s setCurr method, which (in the nonexceptional case) sets \(g\) and \(c\), potentially breaking WghtDiff’s invariant in the last conjunct. This violates WghtDiff’s invariant in a visible state, namely the post-state of this super call, as indicated in the code.

According to the assumption that all invariants must hold in all visible states, the last assignment in WghtDiff’s setCurr method is not necessary, even though operationally it is certainly needed to re-establish the invariant broken by the super call. This illustrates a problem with the simplest view of invariants. But the most difficult problem to solve is reentrant callbacks. The self-call to setCurr in the superclass method is dynamically dispatched to the override in WghtDiff; in this example nothing goes wrong but in general such a call can find the invariant temporarily false because there is already a method invocation (here, lose) in progress.

We expected to need visibility, ownership, etc. to formalize a sound notion of invariants — it was a pleasant surprise that in semantic terms these all amount to ways to ensure that invariants can be assumed as preconditions by method implementations, so they do not complicate our formalization (Def. 20). The effective invariant is governed by (1) and the means by which it is made sound to assume invariants as preconditions is independent from the justification of supertype abstraction (Thm. 24).

Two techniques for ensuring that invariants can be assumed as preconditions in OO languages have been investigated recently. The relevant invariant semantics is based on an ownership type system, which defines the invariants that are “relevant” to a given method [23]; in this technique a method must establish all relevant invariants that might have been broken before making a method call. The Boogie [42][43][44] technique uses a dynamic notion of ownership and strong invariants that hold in all states. Declared invariants are allowed to be temporarily broken.

To express whether the declared invariant is in effect, Boogie uses a specification-only field, invar, that ranges over the object’s supertypes. An object’s invariant takes the form

\[
\forall T \cdot \text{self.invar} \leq T \implies \text{inv}^T(\text{self})
\]

where \(\text{inv}^T\) is the invariant explicitly declared for type \(T\). During temporary violations of the declared invariant for \(T\), this field is set to a supertype of \(T\). Thus the methodology ensures that (5) holds in every reachable state. What matters is that these techniques provide invariants, derived from the declared ones, that can soundly be assumed as a precondition in verifying a method implementation.

### 2.4 Enforcing behavioral subtyping by specification inheritance

As can be seen from our examples, writing the specifications for a behavioral subtype often involves a certain amount of repetition. For example, the postcondition of the lose method in class WeightLoss (Fig. 1) has two conjuncts, and the first of these repeats both the precondition and the postcondition of the overridden method in the supertype Tracker (Fig. 1).

A similar repetition happens for invariants, as can be seen in the invariant of class WghtDiff (Fig. 2), whose first two conjuncts repeat the invariant of its supertype WeightLoss.

Such repetition can be avoided using specification inheritance [14][50]. In JML, the programmer declares two specifications, one for the supertype, and one for the subtype. The effective specification is given by specification inheritance as described below. For invariants, the extension of the subtype’s specification produces the conjunction of its declared invariant with the conjunction of the invariants from its supertypes. This would allow the invariant of WghtDiff to be stated as follows:

\[
\text{protected invariant } \text{self.d} = \text{self.c} - \text{self.g};
\]

which avoids repeating the first two conjuncts shown in Fig. 2. For methods, an overriding method declaration inherits the specifications of each declaration of that method that it overrides. Its effective specification is the join (least upper bound) of the explicit specification with the specifications of that method in all supertypes. In JML the join of a method specification with the specifications of overridden methods is indicated by the keyword also. Using this convention one could write the specification of the lose method in class WeightLoss as follows:

\[
\text{meth lose(kg: int) also requires self.goal > self.curr - kg};
\]

As we explain in Sect. 7 the meaning of the above specification under specification inheritance is essentially the same as the specification given in Fig. 2. That is, the meaning has as its precondition the conjunction of the explicit precondition and those inherited (which in this example is true), and as its postcondition a conjunction of implications, which says that if a precondition was satisfied, then the corresponding postcondition must hold. Theorem 30 says specification inheritance forces behavioral subtyping. A number of tools and logics enforce behavioral subtyping through equivalent means, though it is not always explicit.

### 2.5 Related work

Liskov and Wing’s paper [29] has already been discussed above. Their paper is famous because they clearly present the main ideas and several interesting examples. Liskov and Wing formulate something like supertype abstraction, their “Subtype Requirement” (p. 1812), but it is sketched in terms of provability and does not directly address modular reasoning about code and method contracts. They present informal arguments why behavioral subtyping ensures subtype requirements. The reasoning they permit involves only invariants and history constraints, apparently because the model of computation allows concurrency. By contrast, we use a sequential language that allows Hoare-style reasoning using invariants as well as pre- and postconditions.

Leavens and Dhara [21] survey a lot of older work on behavioral subtyping, including the pioneering work of America [2][3] and Meyer [30]. Much of it is similar to Liskov and Wing’s and has similar limitations.

Several logics have been given for sequential fragments of Java which incorporate supertype abstraction [31][12][43][50]. These logics mostly achieve behavioral subtyping by requiring that each overriding method implementation in a type satisfies the corresponding specification in each of its supertypes. This has the same effect as our definition of specification inheritance, which explicitly constructs the “effective specifications”. Some prove soundness and even completeness of a proof system with behavioral subtyping, which justifies supertype abstraction in their setting. Of these, only Müller’s [31] considers interfaces, and even this misses our insight that interfaces can be exempted from the requirements of behavioral subtyping that must apply to (non-abstract) classes. Müller concentrates on a modular treatment of frame axioms (modifies clauses) and invariants using ownership.

Parkinson’s work [42] is based on separation logic [29] and encapsulates invariants using a nonstandard form of opaque predicate [10] which can be seen as higher order quantification [3]. Parkinson says “behavioral subtyping” for the standard implications [2] and [3] and “specifi-
cation compatibility” for a proof-theoretic formulation of the intrinsic refinement ordering \[12\] Def. 3.5.1. Pierik \[43\] gives a more conventional proof system, in particular a proof outline logic with a first-order assertion language using finite sequences for heap expressions. As we do in Sect. 7, Pierik explicitly connects specification refinement with adaptation rules. Supertype abstraction and behavioral subtyping are present but intertwined with many other details. Pierik and Parkinson both prove soundness and Pierik proves completeness. But the relation between a logic and its models does not address our objective of explicating the connection between behavioral subtyping and supertype abstraction.

Pierik and de Boer \[44\] investigate various notions of completeness of logics in the presence of behavioral subtyping. One reason it is important to have disentangled behavioral subtyping from specification inheritance is that it illuminates an important design decision for specification languages. Many of the above cited works force behavioral subtyping by using specification inheritance. This decision prevents redundancy in specifications and avoids the need for subtypes to see all specifications in supertypes, some of which may be invisible (such as private invariants). The downside of this decision is that the join used in specification inheritance can lead to unsatisfiable specifications. Worse yet it can mask design mistakes.

Findler and Felleisen take the opposite decision \[15\], i.e., checking for behavioral subtyping rather than imposing it by fiat with specification inheritance. Their work is a foundational study of runtime assertion checking for pre/post specifications without invariants. Their contract checker can detect violations of behavioral subtyping as embodied in Liskov and Wing’s pre/post rules. \[2\] and \[3\]. They note that violations of the pre/post condition rule \[2\] do not occur in a specification language that constructs effective preconditions by using disjunction, as in specification inheritance. So runtime checkers for such languages cannot detect violations of the precondition rule.

3. Language

The technical development uses an idealized object-oriented language that models a large fragment of the class-based languages like Java and C#. It includes interfaces, mutually recursive classes, exceptions as first-class objects, type test and casts, inheritance and dynamic binding. It does not model: concurrency, nested classes, and reflection including dynamic class loading, to avoid complications. Constructors are also omitted, to keep the language small. For issues like evaluation order and the semantics of null casts, where reasonable languages may differ on semantic details, we follow Java. The semantics is adapted from \[6\] which in turn draws on \[13\] for formalization of syntax including the class table.

Readers familiar with conventional languages should have little difficulty with the language syntax, but some effort is needed to follow the details of the semantic definitions so we begin with design rationale.

For reasoning about specifications, a “big step” or state transformer semantics is appropriate, since specifications describe behavior in terms of initial and final states of commands. Although state transformer semantics can be defined from a small-step semantics, it is more convenient to give a direct definition that is compositional, to support proof by induction on the structure of commands. Nonetheless, our model uses a standard concrete model of state and the semantic definitions for expressions and commands are operational and unsurprising.

A key aspect of modular reasoning is that reasoning about a command is based on the specifications of methods invoked by the command — independent from any particular implementations of those methods. To model this aspect of modular reasoning without over-specializing our theory by bringing proof theory into the picture, we use a denotational semantics: the meaning of a command or expression is defined with respect to a given method environment which provides the denotations of all methods.

Four technical features of the semantics streamline the formal development but may be unexpected. First, although a distinction is made between expressions and commands, both may have effects. Reasoning systems often restrict expressions to be pure; but we include exceptions in full generality which entails heap effects in expressions. Second, threading state through the semantics of expressions adds considerable complication which is mitigated by the uniform use of a general form of state transformer, with separate variable declarations for the initial and final state spaces. In the case of a command, the initial and final state space are the same except for the addition of an exception variable (exc) in the final state space. In the case of an expression, the initial state space is as expected, but the final state space consists only of two variables, the exception variable and a variable to hold the normal result (res). In the case of a method, the initial state consists of parameters and the receiver variable self; the final state has exception and normal result variables as in the case of expressions. A state transformer of a given type is a mapping from initial states to final states or \( \bot \), the latter modeling divergence.

The third feature is the encoding of exceptions. An expression may diverge, yield a normal result, or throw an exception. The semantics uses a disjoint sum of just two kinds of outcome: either \( \bot \) or a state. But that state includes a special variable exc, of type Throwable, to encode a disjoint sum: the value of exc is either null, which signifies normal termination, or a reference to the exception object (which may contain references to other objects in the heap). Variable exc is not allowed to occur in the program text but is used in specifications.

Fourth, the syntax is in something like “A-normal form” \[49\], i.e., subexpressions in various constructs are restricted to be variables. To avoid loss of expressiveness, let-expressions are added; desugaring transformations are not difficult to define, e.g., a general equality test \( E_1 = E_2 \) can be desugared (\( \sim \)) by the rule

\[
(E_1 = E_2) =_o \text{let } x \text{ be } E_1^o \text{ in let } y \text{ be } E_2^o \text{ in } x = y
\]

which preserves order of evaluation and propagation of exceptions.

Notation summary. The metatheory is standard set theory and we often use partial functions (finite maps) treated as sets of pairs. Unqualified, the term function means total function. Much of the formalism is expressed in terms of dependently typed functions, the notation for which is deferred until later.

Application is written with juxtaposition and associates to the left as in \( f \ a \ b \) for a curried \( f \). Finite mappings are used for typing contexts and variable stores, e.g., \([self: C; x:int]\) or with brackets omitted. The extension of a finite mapping \( g \) to map \( b \) to \( z \), where \( b \notin \text{dom } g \), is written \([g; b = z]\). To override the mapping for an element \( c \in \text{dom } g \) we write \([g | c = z]\). To streamline the semantic definitions, our meta-language uses \( \sim \)-strict let-expressions: the value of “let \( x = \alpha \) in \( \beta \)” is \( \bot \) in case \( \alpha = \bot \), otherwise it is as usual. A raised dot separates a variable binding from its scope, as in \( \forall x: \text{int} \cdot a[x] > 0 \).

Syntax. The grammar is based on some given sets of names, using the following nomenclature for typical elements:

\[
C \in \text{ClassName} \quad \text{names of declared classes}
\]

\[
i \in \text{InterfaceName} \quad \text{names of declared interfaces}
\]

\[
x, y, f \quad \text{variable names (for param., fields, and locals)}
\]

\[
m \quad \text{method names}
\]

There are two distinguished variable names, self for the receiver object (this in Java) and res. The final value of res gives the return value of every method body has the form “\( S; \text{return } res; \)”. Using res fits with JML and lets us omit return statements. There is one distinguished interface name, Throwable, and three distinguished class names Object, NullDeref and ClassCastException; the latter implement Throwable.

Class and interface declarations have the following forms:

\[
\text{class } C \text{ extends } C \text{ implements } T \{ \quad \text{exc} \}
\]

\[
\text{interface } I \text{ extends } T \{ \quad \text{res} \}
\]

Here and throughout we use over-lines to indicate sequences, possibly empty. Instance fields are included in interfaces since they are needed in specifications. In a specification language they would be designated as “ghost” or “shadow” fields and our results apply to programs that are properly annotated for ghost fields, but the distinction is not needed for our results.
The remaining syntactic categories are as follows. (Bold keywords and punctuation marks including "{" and "}" are terminal symbols.)

\[
\begin{align*}
T &::= C \mid I \mid \text{bool} \mid \text{int} \mid \text{unit} &\quad& \text{data type} \\
\text{vis} &::= \text{private} \mid \text{protected} \mid \text{public} &\quad& \text{visibility modifier} \\
\text{msig} &::= m(\overline{x} : T) : T &\quad& \text{method signature} \\
\text{mdec} ::= \text{meth} \msig \{ S \} &\quad& \text{method declaration} \\
S &::= x := E \mid x.f := x \mid \text{var} \ x : T \in S \mid S ; S \mid \text{if} E \text{ then } S \text{ else } S &\quad& \text{assign to var., to field} \\
&\quad& \text{local variable block} \\
&\quad& \text{sequence, conditional} \\
&\quad& \text{throw } x \quad \text{throw exception} \\
&\quad& \text{try-catch} \\
&\quad& \text{try-finally } S \quad \text{try-finally} \\
E &::= x \mid \text{null} \mid \text{true} \mid 0 \ldots \mid x.f \mid x = x \mid x \in T \mid (T) x \mid \text{new } C() \mid x.m(\overline{x}) \mid \text{let } x \text{ be } E \in E &\quad& \text{variable, constants} \\
&\quad& \text{field access, equality test} \\
&\quad& \text{type test, cast} \\
&\quad& \text{object construction} \\
&\quad& \text{method call} \\
&\quad& \text{sequenced local binding}
\end{align*}
\]

For brevity the term ref type is used to mean any non-primitive data type, i.e., a class or interface name, and we define

\[\text{RefType} = \text{ClassName} \cup \text{InterfaceName}\]

A complete program is a collection of class and interface declarations. Formally, we consider a class table CT which is a mapping sending each class name C to its declaration CT(C) and each interface name I to its declaration CT(1). We eschew the term “statement” but use identifier S for commands since C is mnemonic for class names.

The typing rules are syntax directed. The rules for commands and expressions use judgments in which the variable context is explicit, giving names and types of local variables and parameters that are in scope (namely, the method parameters, any locals in scope, and the special variables self, res). The rules also depend on the complete class table, as is needed to deal with recursive class declarations —this is not made explicit in the judgments.

Various auxiliary notations are used that depend on the class table, again without explicit indication. Suppose CT(C) is

\[\text{class } C \text{ extends } D \text{ implements } T \{ \text{vis } f : T ; \text{ mdec } \}\]

Then we define super C = D and superinterfaces C = T. Let mdec be in the list mdec of method declarations, so mdec has the form

\[\text{meth } m(\overline{x} : T) : T \{ S \} \]

We record the type and parameter names by defining mtype(C, m) = \(\overline{x} : T \rightarrow T_1\). If m is inherited in C from D (i.e., is defined in D but not declared in C) then mtype(C, m) is defined to be mtype(D, m). Thus mtype(C, m) is defined iff m is declared or inherited in C. Similarly for interfaces: if I extends T then superinterfaces I = T and mtype(I, m) is defined the same as for classes.

For declared fields of the class displayed above we define dfields C = \(\text{vis } f : T \) and similarly dfields I are the fields declared by interface I. To include inherited fields for such a class C we define

\[\text{fields } C = \text{fields } D \cup \text{dfields } C \cup \{ \cup \overline{1} \in \text{superinterfaces } C \cdot \text{fields I} \}\]

In a well formed class table each of these unions will be disjoint. For interfaces define fields I = dfields I \cup fields(superinterfaces I). We define Meths T = \{ m | mtype(T, m) \} is defined \} noting that mtype(T, m) is defined just when T is a ref type and m is declared or inherited in T.

The subtype relation \(\leq\) is defined inductively by

- \(C \leq D\) if super C = D
- \(T \leq I\) if I \(\in\) superinterfaces T (in which case T is a ref type)
- \(I \leq \text{Object}\) if I \(\in\) dom(C) is an interface
- \(\leq\) is reflexive and transitive

Note that for primitive types, \(T \leq U\) holds just if T = U.

A class table CT is well formed provided it satisfies standard constraints such as acyclicity of \(\leq\). Every type in a field declaration, parameter, etc., is one of the primitives or in RefType. Every method declaration \(m(\overline{x} : T) : T \{ S \}\) in CT(C) is typable in the sense that \(\Gamma \vdash S\) where \(\Gamma = \{ \text{self} : C, \text{res} : T, \overline{x} : T \}\). Method overrides must not change the signature. Rules that define \(\Gamma \vdash S\) are in Appendix A. We give here just two.

\[
\begin{align*}
\Gamma \vdash E : T &\quad \Gamma, x : T \vdash E1 : U \\
\Gamma \vdash \text{let } x \text{ be } E \in E1 : U \\
\Gamma \vdash x : T &\quad T \in \text{RefType} \\
\text{mtype}(T, m) &\quad \overline{x} : T \rightarrow U \\
\Gamma \vdash \gamma : \overline{V} &\quad \overline{V} \leq \overline{T}
\end{align*}
\]

### 3.1 Semantic domains

Details of the semantics for expressions and commands are not very important, although they are needed in some proofs. But familiarity with the semantic domains —especially state transformers and method environments—is essential since specifications are treated semantically.

We assume a given set Ref of references —abstract addresses. A ref context is a finite partial function \(\rho\) that maps references to class names. The idea is that if \(o \in \text{dom } \rho\) then o is allocated and moreover o points to an object of type \(\rho\). We define the set RefCtx = Ref → ClassName, where → denotes partial functions. Note that for \(\rho\) and \(\rho'\) in RefCtx, we can write \(\rho \subseteq \rho'\) to express that the domain of \(\rho'\) includes at least the objects in \(\rho\) and for objects allocated in \(\rho\) the types are the same in \(\rho'\).

Semantic domains are the sets of possible meanings for various kinds of phrases such as data types, expressions, and method bodies. The definitions capture important invariants about the semantics, e.g., that the value of a field or variable of a given static type is a value of that type. Moreover, if T is a ref type then its values include references to objects in subclasses. Finally, there are no dangling references. Most of the semantic domains are defined in terms of a given ref context. For example, \(\text{Val}(T, \rho)\) is the set of values of type T in a state where \(\rho\) is the ref context. In case T is a primitive type, the definition of \(\text{Val}(T, \rho)\) is independent from \(\rho\). But if T is a class type, then \(\text{Val}(T, \rho)\) is some set containing null and some references —all the allocated objects of some type U such that \(U \leq T\).

Here is a guide to the domains. Those marked with * are just special cases of the others as explained below.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Description</th>
<th>Metavariables</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{Val}(T, \rho))</td>
<td>value of type T</td>
<td>(\alpha, \nu)</td>
</tr>
<tr>
<td>(\text{Store}(\Gamma, \rho))</td>
<td>stores for (\Gamma)</td>
<td>(r)</td>
</tr>
<tr>
<td>(\ast \text{Obrecond}(C, \rho))</td>
<td>fields of C-objects</td>
<td></td>
</tr>
<tr>
<td>(\text{Heap}(\rho))</td>
<td>heaps</td>
<td>(h)</td>
</tr>
<tr>
<td>(\text{State}(\Gamma))</td>
<td>states for (\Gamma)</td>
<td>(s, t)</td>
</tr>
<tr>
<td>(\text{STrans}(\Gamma_1, \Gamma_2))</td>
<td>state transformers</td>
<td>(\sigma, \tau)</td>
</tr>
<tr>
<td>(\ast \text{SemExpr}(T, \rho))</td>
<td>semantic expression</td>
<td></td>
</tr>
<tr>
<td>(\ast \text{SemCommand}(\Gamma))</td>
<td>semantic command</td>
<td></td>
</tr>
<tr>
<td>(\ast \text{SemMeth}(C, \rho))</td>
<td>semantics of method (m) in C</td>
<td></td>
</tr>
<tr>
<td>(\text{MethEnv})</td>
<td>method environment</td>
<td>(\mu)</td>
</tr>
<tr>
<td>(\text{XMethEnv})</td>
<td>extended meth. env. (incl. interfaces)</td>
<td>(\mu)</td>
</tr>
</tbody>
</table>

For data types T the definition is by cases on T:

- \(\text{Val}(\text{bool}, \rho) = \{ \text{true}, \text{false} \}\)
- \(\text{Val}(\text{int}, \rho) = \{ \ldots, -2, -1, 0, 1, 2, \ldots \}\)
- \(\text{Val}(\text{unit}, \rho) = \{ \text{it} \}\)
- \(\text{Val}(C, \rho) = \{ \text{null} \} \cup \{ o | o \in \text{dom } \rho \land \rho \leq C \}\)
- \(\text{Val}(I, \rho) = \{ o | \exists C \cdot C \leq I \land o \in \text{Val}(C, \rho) \}\)

The next definitions involve dependent function spaces or dependent pairs, for which we use the following.

**Notation for dependent types in metatheory**

Suppose X is a set and for every \(x \in X\) a set \(Z_x\) is given. Then the dependent product \(\{ y : X \times Z_x \mid x \in X \times z \} \) such that \(x \in X\) and \(z \in Z_x\) is the dependent function space \(\{ y : X \rightarrow Z_y \mid y \in \{ y : X \times Z_y \mid x \in X \times z \} \} \) such that \(f \in Z_y\) is the set of functions \(f \) from \(X\) to \(\cup_{y \in X} Z_y\) such that \(f \in Z_y\).
for all $x \in X$. Note that $y$ is a bound variable in these notations.\(^4\)

The first use is in the definition

$$\text{Store}(\Gamma, \rho) = (x : \text{dom } \Gamma) \to \text{Val}(\Gamma, x, \rho)$$

What this means is that $\text{Store}(\Gamma, \rho)$ is a set of functions; and for any $r$ in $\text{Store}(\Gamma, \rho)$ (the name $r$ is mnemonic for store), the domain of $r$ is $\text{dom } \Gamma$ and $r$ is an element of $\text{Val}(\Gamma, x, \rho)$ for each $x \in \text{dom } \Gamma$.

Next we build up to program states.

$$\text{Obrecord}(C, \rho) = \text{Store}(\text{fields } C, \rho)$$

$$\text{Heap}(\rho) = (o : \text{dom } \rho) \to \text{Obrecord}(\rho, o, \rho)$$

$$\text{State}(\Gamma) = (p : \text{RefCtx}) \times \text{Heap}(\rho) \times \text{Store}(\Gamma, \rho)$$

Recall that $\text{obcontext}$ is the variable context obtained by removing visibility markers from $\text{fields}$ $C$. So a heap $h$ is map sending each allocated reference $o$ to a record, $h(o)$, of the object’s current field values.

The most important domain is state transformers:

$$\text{STrans}(\Gamma, \Gamma') = (s : \text{State}(\Gamma)) \times \{ s' | s' \in \text{State}(\Gamma') \land \text{extState}(s, s') \}$$

Relation $\text{extState}$ is used to say that one state’s ref context extends the other’s: $\text{extState}(\rho, \rho', h, r') \iff \rho \subseteq \rho'$. Elements of $\text{STrans}(\Gamma, \Gamma')$ are functions that map a state in $\text{State}(\Gamma)$ to either $\perp$ or a state in $\text{State}(\Gamma')$.

The domain of state transformers subsumes meanings for methods, expressions and commands.

$$\text{SemExpr}(\Gamma, T) = \text{STrans}(\Gamma, \text{res : } T, \text{exc : } \text{Thrwbl})$$

$$\text{SemCommand}(\Gamma) = \text{STrans}(\Gamma, \text{Gamma : } T, \text{exc : } \text{Thrwbl})$$

$$\text{SemMeth}(T, m) = \text{STrans}(\text{self : } T, \text{expr : } T', \text{res : } U, \text{exc : } \text{Thrwbl})$$

where $\text{mtype}(m, T) = \text{tau} : T \to U$

A (normal) method environment is defined to be a table of meanings for all methods in all classes:

$$\text{MethEnv} = (C : \text{ClassName}) \times (m : \text{Meths } C) \to \text{SemMeth}(C, m)$$

The idea is that a method environment $\mu$ is defined for pairs $(C, m)$ where $C$ is a class with method $m$ and moreover $\mu(C, m)$ is a state transformer suitable to be the meaning of a method of type $\text{mtype}(C, m)$. In case $m$ is inherited in $C$ from $B$, $\mu(C, m)$ will be the restriction of $\mu(B, m)$ to receiver objects of type $C$.

The semantics of a command depends on its method environment and a complete program $CT$ denotes a method environment, described later.

In the formulation of modular reasoning based on static types it turns out to be convenient to use an extended kind of method environment which associates method meanings to interfaces as well as to classes (even though the receiver of an invocation is always an object of some class, cf. the definition of $\text{Val}(\Gamma, \rho)$).

An extended method environment is defined to be a table that also includes meanings for methods of interfaces:

$$\text{XMethEnv} = (T : \text{RefType}) \times (m : \text{Meths } T) \to \text{SemMeth}(T, m)$$

The metavariable $\mu$ ranges over normal method environments and $\tilde{\mu}$ is used to range over extended method environments — the dot is intended to be mnemonic for the dotted in interface.

3.2 Semantics of expressions, commands, and class table

The semantic definitions for expressions and commands are straightforward but mostly relegated to Appendix 8. We explain just what is needed for the proofs. For an expression $\Gamma \vdash E : T$, the semantics $\text{Eval}(\Gamma, E, T) = \mu(\rho, h, r)$ gets applied to a method environment $\mu$ to yield a state transformer $\text{Eval}(\Gamma, E, T) = \mu(\rho, h, r)$ in turn applied to a state, e.g., $\text{Eval}(\Gamma, E, T) = \mu(\rho, h, r)$. Similarly for commands. The definition is

\(^4\)More standard notations are $\Sigma y : X, Z_y$ for $(y : X) \times Z_y$ and $\Pi y : X, Z_y$ for $(y : X) \to Z_y$. Ours are similar to the PVS prover’s.

\(^5\)Since $\rho$ and $\rho'$ are partial functions which we treat as sets of pairs, $\rho \subseteq \rho'$ says that $\rho'$ has at least the domain of $\rho$ and they agree on their common domain.

Syntax-directed, e.g., here is the semantics of an assignment:

$$\text{Eval}(\Gamma, x := E) = \mu(\rho, h, r)$$

let $(p_1, h_1, r_1) = \text{Eval}(\Gamma, E, T)$ in

if $r_1 \text{ exc } = \text{null}$ then $(p_1, h_1, [r : x : r_1 = \text{null}, \text{res : } \text{null}])$

else $(p_1, h_1, [r, \text{exc : } r_1 = \text{null}])$

if $E$ yields $\bot$ then so does the assignment. If $E$ throws no exception then its value is assigned to $x$ and the final state is extended with $\text{exc set null}$. Otherwise, the final state is extended with $\text{exc mapped to the exception}$.

This definition and the ones in the appendix use the notation $[x : \Gamma]$ for the semantics function, but this abbreviates two definitions. They are defined in the same way, except in the case of method call. The dynamic dispatch semantics, for which we decorate the semantics brackets as $\text{Eval}([x : \Gamma])$, is the operationally accurate one. It dispatches to a method meaning in the method environment based on the dynamic type of the receiver. Let $T = \text{Gamma } x : \text{tau} : T \to U$ and $\text{mtype}(m, T)$ as in the typing rule. Define

$$\text{Eval}(\Gamma, x : T, m(y) : U, \text{exc : } V)$$

$$\text{Eval}(\Gamma, x : T, m(y) : U, \text{exc : } V) =$$

if $r \text{ exc } = \text{null}$ then $\text{exc}(p, h, U, \text{NullDeref})$

else let $r_1 = [\text{self : } x : r : x, \text{exc : } V]$ in $\mu(p, h, r_1)$

Because the dynamic type of the receiver is a class (specifically, $p(x)$), this semantics is based on a normal method environment. The helping function $\text{exc}$ builds a state with $\text{exc}$ set to a new $\text{NullDeref}$ object (see Appendix 8).

The static dispatch or nominal 22 semantics of method call applies a method meaning determined by the static type $T$ of the receiver. Since interfaces are included among the static types, the static dispatch semantics is defined in terms of an extended method environment $\mu$.

$$\text{Eval}(\Gamma, x : T, m(y) : U, \text{exc : } V)$$

$$\text{Eval}(\Gamma, x : T, m(y) : U, \text{exc : } V) =$$

if $r \text{ exc } = \text{null}$ then $\text{exc}(p, h, U, \text{NullDeref})$

else let $r_1 = [\text{self : } x : r : x, \text{exc : } V]$ in $\mu(p, h, r_1)$

One of the proofs deals with the case of let-expressions in detail. Here is the semantics, where $[\bot]$ is either $[\text{Eval}([\bot : \Gamma])]$ or $[\text{Eval}([\bot : \Gamma])]$ throughout.

$$\text{Eval}(\Gamma, \text{let } x \text{ be } E : E \to U) =$$

let $(p_0, h_0, r_0) = \text{Eval}(\Gamma, E : U, \text{exc : } V)$ in

if $r_0 \text{ exc } = \text{null}$ then $\text{null}$

else let $r_1 = [\text{self : } x : r : x, \text{exc : } V]$ in $\mu(p_0, h_0, r_1)$

The semantics is defined with respect to an arbitrary allocator. An allocator is just a choice function for unused locations, i.e., a function $\text{fresh}$ that maps a pair $(p, h)$, with $h \in \text{Heap}(p)$, to a location such that $\text{fresh}(p, h) \not\in \text{dom } \rho$.

4. Specifications and refinement

This section formalizes method specifications and satisfaction by state transformers. On this basis we define specification tables and satisfaction for them as well as the induced refinement relation between specifications. Refinement is used in the following section to define behavioral subtyping.

From now on it is assumed that $CT$ is some well formed class table.

Specification languages have long used special syntax in postconditions to refer to initial state, such as “$\text{old}(x)$” in JML, so that postconditions are two-state predicates (but see Sect. 7 in regards to auxiliary variables).

Specifications for methods and commands are obviously needed, but expression specifications are also needed, to prove one of the main results. So it is convenient to use the notation of a state transformer type $\Gamma \vdash \text{Gamma } T$ for specifications of state transformers in $\text{STrans}(\Gamma, \Gamma')$.\(^7\)

\(^7\)As a simple example, $\text{Ref}$ can be taken to be the naturals and $\text{fresh}(p, h)$ can be the least $n$ not in $\text{dom } \rho$. A realistic allocator depends on program state which is why we include $h$ here.
Definition 1 (state transformer specification) A specification of type \( \Gamma \leadsto \Gamma' \) is a pair \((\text{pre}, \text{post})\) such that

- \text{pre} is a subset of \(\text{State}(\Gamma)\)
- \text{post} is a subset of \(\text{State}(\Gamma) \times \text{State}(\Gamma')\)

For ref type \(T\), a method specification of type \((T, m)\) is a specification of type \([\text{self} \colon T, T \vdash T \leadsto U] \vdash \text{res} : U, \text{exc} : \text{Thrwbl}\) where \(\text{mtype}(m, T) = T : T \rightarrow U\). A \(\Gamma\)-specification is one of type \(\Gamma \leadsto \Gamma'\) \(\Gamma'\) is the "effective specification", which is typically obtained from declared specifications by means of context-dependent interpretation of modifiers such as \([\text{self} \colon U, T \vdash U] \vdash \text{res} : U, \text{exc} : \text{Thrwbl}\). It models what might be called the "effective specification", which is typically obtained from declared specifications by means of context-dependent interpretation of modifiers clauses \([\text{self} \colon U, T \vdash U] \vdash \text{res} : U, \text{exc} : \text{Thrwbl}\), etc. Roughly speaking, a method environment satisfies a specification table if each method satisfies its specification; formally there are two notions corresponding to the two kinds of method environment.

Definition 2 (satisfaction by state transformer) Let \((\text{pre}, \text{post})\) be a specification of type \(\Gamma \leadsto \Gamma'\) and \(\text{be} \in \text{STrans}(\Gamma, \Gamma')\). Then \(\sigma\) satisfies \((\text{pre}, \text{post})\), written \(\sigma \models (\text{pre}, \text{post})\), iff

\[
\forall t \in \text{pre} \Rightarrow (t, \sigma \cdot t) \in \text{post}
\]

A specification table provides specifications for all methods (we add an invariant for each class in Sect.\(\square\)). It models what might be called the "effective specification", which is typically obtained from declared specifications by means of context-dependent interpretation of modifiers clauses \([\text{self} \colon U, T \vdash U] \vdash \text{res} : U, \text{exc} : \text{Thrwbl}\), etc. Roughly speaking, a method environment satisfies a specification table if each method satisfies its specification; formally there are two notions corresponding to the two kinds of method environment.

Definition 3 (specification table) A specification table, \(ST\), contains a method specification \(ST(T, m)\) of type \(\text{mtype}(m, T)\) for each ref type \(T\) and each \(m \in \text{Meths} T\).

For perspicuity we sometimes write \(((\text{pre}^P, \text{post}^P))\) for \(ST(T, m)\).

Definition 4 (satisfaction by method environment) Let \(ST\) be a specification table. An extended method environment \(\mu\) satisfies \(ST\), written \(\mu \models ST\), iff \(\mu(T, m) \models ST(T, m)\) for all ref types \(T\) and \(m \in \text{Meths} T\).

A normal method environment \(\mu\) satisfies \(ST\), written \(\mu \models ST\), iff \(\mu(C, m) \models ST(C, m)\) for all classes \(C\) and \(m \in \text{Meths} C\).

Note this does not require \(\mu(C, m)\) to satisfy specifications of the interfaces implemented by \(C\), nor of its superclass. The idea is that \(ST(C, m)\) is the actual behavioral condition specified for \(C, m\). In Sect.\(\square\) we define specification inheritance as a means to derive \(ST(C, m)\) from declared specifications of \(C\)'s interfaces and superclasses.

Ordering specifications. The behavioral subtyping property is expressed in terms of a refinement ordering on specifications, saying that if \(T\) is a subtype of \(U\) then \(ST(T, m)\) is a stronger specification than \(ST(U, m)\) in the sense that any method satisfying \(ST(T, m)\) also satisfies \(ST(U, m)\). This intrinsic ordering on specifications is determined by the nature of command denotations and the definition of satisfaction. Some care needs to be taken with the details, because if \(T\) is a class, a method in class \(T\) is defined to act on receiver objects of type \(T\) whereas a specification of type \((U, m)\) imposes a requirement on state transformers for target type \(U\). Owing to the semantics of dynamic dispatch, however, it is sound for a method in class \(T\) to assume a strengthened precondition saying that the receiver object has type \(T\). (This is explicit in the proof obligation for method bodies in proof systems, e.g. \([\text{self} \colon U, T \vdash U] \vdash \text{res} : U, \text{exc} : \text{Thrwbl}\).)

For a method \(m\) of class \(U\) with \(\text{mtype}(U, m) = T : T \rightarrow U\), the relevant state transformers are in \(\text{SemMeth}(U, m)\), i.e., of type \([\text{self} : U, T \vdash U] \vdash \text{res} : U, \text{exc} : \text{Thrwbl}\). For \(T\), a method meaning will have type \([\text{self} : T, T \vdash T] \vdash \text{res} : U, \text{exc} : \text{Thrwbl}\) —only the type of \(\text{self}\) varies.\(\square\) This leads us to define a nonstandard notion of satisfaction. To that end it is convenient to define an operation that converts a predicate to a two-state predicate. For \(p\) a set of states, define \(\text{old}(p)\) by

\[
(s, t) \in \text{old}(p) \iff s \in p
\]

It is well known how to vary parameter and result types; our results can be easily adapted to such variation but we use invariant method overriding as in Java to avoid unelaborating notational complications.

5. Supertype abstraction and behavioral subtyping

This section states and proves the supertype abstraction theorem. Behavioral subtyping is defined in terms of the intrinsic refinement order on method specifications and then an alternative formulation is given. The rest of the section is devoted to stating and proving the main theorem.

The definition of behavioral subtyping is slightly weaker than one might expect, as needed for an equivalence with supertype abstraction.

Definition 8 (behavioral subtyping) A specification table \(ST\) has behavioral subtyping if and only if for all ref types \(U\) and classes \(C\) with \(C \leq U\) and all \(m \in \text{Meths} U\) we have \(ST(C, m) \subseteq ST(U, m)\).

If \(\geq\) is any preorder relation on some set, an instance \(a \geq b\) is equivalent to \(\forall c : b \geq c \Rightarrow a \geq c\). Supertype abstraction for methods is roughly a restatement of behavioral subtyping in this manner, though taking into account the change of type.

Definition 9 (supertype abstraction for method specifications) Specification table \(ST\) has supertype abstraction for method specifications iff the following holds for all ref types \(T\), all \(m \in \text{Meths} T\), and all \(\sigma\).

\[
(\exists \text{pre}^P, \text{post}^P) \models (\text{pre}, \text{post}), \text{if} \sigma \models (\text{pre}, \text{post}) \Rightarrow \sigma \models ST(T, m)
\]

The notation \(\models\) doesn't fully capture the contextual assumptions but it should suffice to avoid ambiguity in our use of it.

9If desired, antisymmetry can be achieved without loss of generality by normalizing specifications.
method specifications spec of type mtype(T, m): If \(ST(T, m) \supseteq \text{spec}\) then \(ST(C, m) \supseteq \text{spec}\) for every class \(C\) with \(C \leq T\).

**Lemma 10** A satisfiable specification table \(ST\) has behavioral subtyping iff it has supertype abstraction for method specifications.

**Proof** For the implication left to right, suppose \(ST\) has behavioral subtyping. Consider any pair \((T, m)\) and method specification spec for \((T, m)\) such that \(ST(T, m) \supseteq \text{spec}\) and any \(C\) with \(C \leq T\). By behavioral subtyping we have \(ST(C, m) \supseteq \text{spec}\) \(ST(T, m)\). Since \(ST(T, m)\) is satisfiable, we can apply weak transitivity (Lemma 7) to get \(ST(C, m) \supseteq \text{spec}\).

For the implication right to left, suppose \(ST\) has supertype abstraction for method specifications. Consider \(C \leq U\) where \(U\) is any ref type. Instantiate supertype abstraction with \(T := U\) and \(\text{spec} := ST(U, m)\). Since \(ST(U, m) \supseteq \text{spec}\) \(ST(T, m)\), supertype abstraction yields \(ST(C, m) \supseteq \text{spec}\) \(ST(U, m)\). □

**Supertype abstraction for commands.** The idea is that a modular reasoning system is a sound means to prove that some command \(\Gamma \vdash S\) satisfies some specification spec. For the proof to be modular means that reasoning about method calls in \(S\) is based only on the specifications of those methods. Thus our semantic formulation says that \(S\) satisfies spec when \(S\) is interpreted by the static dispatch semantics. Of course the static dispatch semantics of a command has many properties that are inconsistent with its standard semantics, so reasoning on the basis of static dispatch semantics with a particular method environment would be unsound. To capture that reasoning about method calls is based only on their specifications, our formulation quantifies over all environments that satisfy \(ST\). Reasoning is expressed in terms of program semantics, so this is a property of a class table together with a specification table.

**Definition 11** Let \(ST\) be a specification table for class table \(CT\). Supertype abstraction is valid for \(ST, CT\) iff for all \(\Gamma \vdash S\) and all \(\Gamma\)-specifications spec, (5) implies (6), where

\[
\forall \mu \in \text{MethEnv} \cdot \mu \vdash ST \Rightarrow S[\Gamma \vdash \text{spec}] \mu \vdash \text{spec} \quad (6)
\]

\[
\forall \mu \in \text{MethEnv} \cdot \mu \vdash ST \Rightarrow D[\Gamma \vdash \text{spec}] \mu \vdash \text{spec} \quad (7)
\]

The idea is that modular reasoning establishes (6) but it is then a consequence that \(S\) satisfies spec in the sense of the standard semantics, i.e., \(D[\Gamma \vdash \text{spec}] \mu \vdash \text{spec}\) where \(\mu\) is the semantics of the class table —provided that \(\mu\) does satisfy \(ST\) (i.e., there is a proof obligation that each method implementation satisfies its specification). In fact, owing to modularity of reasoning about satisfaction as described by (6), the stronger conclusion (7) can be drawn.

In light of Lemma 10 the main result says that behavioral subtyping validates such reasoning, and indeed it is necessary.

**Theorem 12** (supertype abstraction) For any satisfiable \(ST\) the following are equivalent.

(a) \(ST\) has supertype abstraction for method specifications.

(b) supertype abstraction is valid for \(ST, CT\).

We sketch the argument here and return to it after laying some groundwork. The idea for (b) \(\Rightarrow\) (a) is to instantiate \(S\) with suitable method call and unfold the semantics. The idea for (a) \(\Rightarrow\) (b) is to prove (b) by structural induction on \(S\) assuming that \(ST\) has supertype abstraction. A key lemma to prove (b) is an analogous result for expressions, also proved by induction. In these proofs there are three kinds of cases:

- \(S\) is a method call — then the semantics is used to reduce implication (6) \(\Rightarrow\) (7) to the implication given by the supertype abstraction property.

- \(S\) is some other primitive expression or command — then the semantics is used to prove (7) directly from (6).

- \(S\) is a compound command or expression. In that case we decompose it to some state transformers used in its semantic definition and appeal to the induction hypothesis for these state transformers and certain specifications obtained using weakest preconditions.

It is important that the quantifiers are arranged as they are in Def. 11 so that the induction hypothesis is of the form “for all spec and \(S\), (6) \(\Rightarrow\) (7), because for a given \(S\) of the third kind we need multiple instantiations of spec and \(S\).

Putting the Theorem together with lemma 10 we obtain the following.

**Corollary 13** (semantic soundness and completeness) Suppose \(ST\) has behavioral subtyping. Suppose \(S[\Gamma \vdash \text{spec}] \mu\) can be proved to satisfy some \(\Gamma\)-specification spec, assuming only that \(\mu\) satisfies \(ST\). Then the actual semantics \(D[\Gamma \vdash \text{spec}] \mu\) satisfies spec, provided that the semantics, \(\bar{\mu}\), of the class table satisfies \(ST\). Moreover, if such reasoning is sound then \(ST\) has behavioral subtyping.

**Proving the Theorem.** The most difficult part of the proof is to prove (7) by structural induction on \(S\) (Lemmas 16 and 17). We need to argue in a way that can be illustrated by considering the case that \(S\) is a sequence \(S_0; S_1\). Dropping the semantic brackets etc., the idea is that to show \(S_0; S_1 \models \text{spec}\) we find specifications \(\text{spec}_0\) and \(\text{spec}_1\) such that \(S_0 \models \text{spec}_0\) and \(S_1 \models \text{spec}_1\). If \(S_0; S_1 \models \text{spec}\) in the static dispatch semantics then owing to our careful choice of \(\text{spec}_0\) and \(\text{spec}_1\) we get \(S_0 \models \text{spec}_0\) and \(S_1 \models \text{spec}_1\) in the static dispatch semantics.

Now the induction hypothesis can be invoked to yield that these hold in the dynamic dispatch semantics, whence the semantic equation for sequence can be used to get \(S_0; S_1 \models \text{spec}\). Suitable \(\text{spec}_0\) and \(\text{spec}_1\) can be obtained by using the weakest precondition of \(S_1\) with respect to the given post. So the technical details involve defining the weakest precondition for a state transformer and proving decomposition results for sequential and conditional composition of state transformers.

The semantics for expressions and commands are based on explicitly defined state transformers (described by the mathematical notations for updates, dropping variables, etc.) and those given by the method environment. These are composed sequentially and conditionally. (The definitions use the \(\downarrow\)-strict let construct which combines sequence with test ing for \(\downarrow\).

Suppose \(\sigma\) is a state transformer of type \(\Gamma \rightarrow \Gamma'\) and post is a subset of \(\text{State}(\Gamma) \times \text{State}(\Gamma')\). Define the weakest precondition of \(\sigma\) with respect to post, written \(\wp \sigma\) post, to be a subset of \(\text{State}(\Gamma)\) as follows:

\[
t \in \wp \sigma\) post \iff (t, s) \in post\]

Clearly \(\sigma\) is \((\wp \sigma\) post\), post). Now we can give the decomposition lemma for sequenced transformers.

**Lemma 14** (sequential decomposition) Suppose \(\sigma\) is a state transformer of type \(\Gamma_0 \sim \Gamma_2\). Suppose moreover that \(\sigma = \sigma_0; \sigma_1\) where each \(\sigma_i\) has type \(\Gamma_i \sim \Gamma_{i+1}\) and semicolon is function composition. Suppose \(\text{spec} = (\pre, \text{post})\) and define \(\text{spec}_0 = (\pre, (\wp \sigma_1 \text{post})\) and \(\text{spec}_1 = ((\wp \sigma_1 \text{post}), \text{post})\).

Then

\[
\sigma \models \text{spec} \iff \sigma_0 \models \text{spec}_0\) and \(\sigma_1 \models \text{spec}_1\]

and moreover for any \(\sigma_0', \sigma_1',\) if \(\sigma_0' \models \text{spec}_0\) and \(\sigma_1' \models \text{spec}_1\) then \(\sigma_0' \sigma_1' \models \text{spec}\).

These are well known facts about weakest preconditions; the lemma merely spells them out in a particular way because their use later is a little intricate. Similarly for the following.

**Lemma 15** (conditional decomposition) Suppose \(\sigma\) is a state transformer of type \(\Gamma \sim \Gamma'\) such that for all states \(s\) we have

\[
\sigma s = (\{s \in P \mid \sigma_0 s \text{ else } \sigma_1 s\}
\]

where \(P \subseteq \text{State}(\Gamma)\) and each \(\sigma_i\) has type \(\Gamma_i \sim \Gamma_{i+1}\). Suppose \(\text{spec}\) has type \(\Gamma \rightarrow \Gamma'\) and let \(\text{spec} = (\pre, \text{post})\) to define \(\text{spec}_0 = (\pre \circ \wp \sigma_1 \text{pre, post})\) and \(\text{spec}_1 = ((\wp \sigma_1 \text{post, post})\).

Then

\[
\sigma \models \text{spec} \iff \sigma_0 \models \text{spec}_0\) and \(\sigma_1 \models \text{spec}_1\]

and moreover (if \(s \in P\) then \(\sigma_0 s \text{ else } \sigma_1 s\) \(\models \text{spec}\) for any \(\sigma_0' \sigma_1'\) that satisfy \(\text{spec}_0\), \(\text{spec}_1\).

**Lemma 16** (supertype abstraction for expressions) Suppose \(ST\) has supertype abstraction and is satisfiable. Then for all \(\Gamma, E, T, \text{spec}\) such that \(\Gamma \models E : T\) and \(\text{spec}\) is of type \(\Gamma \sim \text{res : T, exc : Thwbl}\) we have that (8) implies (9) where

\[
\forall \mu \in \text{MethEnv} \cdot \mu \vdash ST \Rightarrow S[\Gamma \vdash E : T]\mu \vdash \text{spec} \quad (8)
\]
∀μ ∈ MethEnv · μ |≡ | ST ⇒ D[Γ] ⊢ E : T |μ |≡ spec 

Proof: By structural induction on E. For the case Γ ⊢ x : T, the result follows from a much stronger property:

S[Γ ⊢ x : T]μ = D[Γ ⊢ x : T]μ for all μ and all μ

This holds because the two semantics are identical, there are no sub-expressions, and the semantics is independent of the method environment. The argument is the same for null and other constants, as well as for x, f, x = y, E = T, (T) x, and new C().

The remaining cases are x.m(π), which involves the method environment and supertype abstraction, and let x be E in E1, for which the induction hypothesis is used.

For case Γ ⊢ let x be E in E1 : U, assume

∀μ ∈ XMedEnv · μ |≡ | ST ⇒ D[Γ] ⊢ let x be E in E1 : U |μ |≡ spec

for some spec of type Γ ⊢ [res : U, exc : Throwbl]. Let μ be any normal method environment such that μ |≡ | ST. We must show

D[Γ ⊢ let x be E in E1 : U]μ |≡ spec

By assumption (10) and satisfiability of ST is there some μ such that

S[Γ ⊢ let x be E in E1 : U]μ |≡ spec

Let us write out the definition of S[Γ ⊢ let x be E in E1 : U]μ in a way that makes explicit the manipulation of ⊥. (The definitions in Table 3 use the monadic let to suppress ⊥.)

S[Γ ⊢ let x be E in E1 : U]μ(r, h, r) = let X = S[Γ ⊢ E : T]μ(r, h, r) in

if X = ⊥ then ⊥ else let (r0, h0, r0) = X in

if r0 exc ≠ null then (r0, h0, [res : default U, exc : r0 exc])
else r1 = [r, x : r0 res] in S[Γ ⊢ E1 : U]μ(r0, h0, r1)

The dynamic-dispatch semantics is identical except for replacing the three occurrences of S[−] with D[−] and using a normal method environment.

The point of writing out the semantics is to make clear that it is just the alternative/sequential composition of certain state transformers:

• S[Γ ⊢ E : T]μ of type Γ ⊢ [res : T, exc : Throwbl]


• others, such as one we will call f, that sends (ρ0, h0, r0) to (ρ0, h0, [res : default U, exc : r0 exc : Throwbl]).

By the decomposition lemmas there are specifications specE, specF1, specF2, ... such that (11) holds iff each of the component transformers satisfies its specification.

Since assumption (10) holds for all μ, it follows that S[Γ ⊢ E : T]μ |≡ specE for all μ and S[Γ ⊢ E : U]μ |≡ specF1 for all μ. (Some readers will want to check that the arrangement of quantifiers in the Lemma does just this step and that plausible simplifications do not work.) As a consequence, we may appeal to the induction hypothesis for Γ, E, T, specE and for [Γ, x : T] E1, U, specF1. This yields that D[Γ ⊢ E : T]μ |≡ spec and D[Γ, x : T] ⊢ E1 : U]μ |≡ specF1 for our arbitrarily chosen μ. The other component transformers like f are same in both the static and dynamic dispatch semantics. Having established that the component transformers of D[Γ ⊢ let x be E in E1 : U]μ all satisfy the component specifications, we obtain D[Γ ⊢ let x be E in E1 : U]μ |≡ spec which was to be proved.

Finally, consider the case of Γ ⊢ x.m(π) : U. Recall the static dispatch semantics for methods:

S[Γ ⊢ x.m(π) : U]μ(r, h, r) =

if r x = null then except (ρ, h, U, NullDeref)
else let r1 = [self : r x, x : r π] in μ(T, m)(ρ, h, r1)

Suppose mtype(m, T) = π : T → U as in the typing rule for method call. Suppose spec has type Γ ⊢ [res : U, exc : Throwbl] and choose some μ such that S[Γ ⊢ x.m(π) : U]μ |≡ spec. (Such μ exists owing to satisfiability of ST and the assumption (10).) By decomposition we obtain spec of type π : T → U such that μ(T, m) |≡ spec. Moreover, noting that if μ |≡ | ST then so does μ |≡ | ST for any r with σ |≡ | ST(T, m), it follows from assumption (10) that ST(T, m) ⊑ spec'.

Now suppose μ is any normal method environment that satisfies ST and recall the dynamic dispatch semantics which differs in using the dynamic type ρ(r x) of the receiver, rather than its static type T, to look up the method in the environment.

D[Γ ⊢ x.m(π) : U]μ(r, h, r) =

if r x = null then except (ρ, h, U, NullDeref)
else let r1 = [self : r x, x : r π] in μ(r(x), r)(ρ, h, r1)

By supertype abstraction for methods (Def. 3), ST(T, m) ⊑ spec implies that C ⊑ ST(C, m) ⊑ spec' for all C. Since μ |≡ | ST we have for each C ⊑ T that μ(C, m) |≡ | ST(C, m) and thus μ(C, m) |≡ | C spec'. To complete the proof of D[Γ ⊢ x.m(π) : U]μ it is not enough to use decomposition backwards; we also unfold the definition of |≡ | and since μ(C, m) is used just in case r(x) ⊑ C.

The satisfiability hypothesis is necessary. Suppose that ST(C, m) is unsatisfiable for some C. Then implies (9) because both have false antecedents. This does not let us drop the satisfiability hypothesis because it can happen that the only unsatisfiable part is some interface specification ST(I, m), falsifying the antecedent only of (9).

The following result amounts to the (a) ⇒ (b) part of Theorem 12.

Lemma 17 (supertype abstraction for commands) Suppose ST has supertype abstraction and is satisfiable. Then for all Γ, S, spec such that Γ ⊢ S and spec is of type Γ ⊢ [self : T, π : T]μ |≡ spec implies ∀μ ∈ MethEnv · μ |≡ | ST ⇒ S[Γ ⊢ S]μ |≡ spec

Proof: By structural induction on S.

In the cases that S is x.f := y and throw x, the semantics using S[−] and D[−] are identical so the proof is immediate.

In the cases of conditional, sequence, try-catch and try-finally, the argument is by induction following the pattern for let-expression in the proof of Lemma 10. That is also the pattern for the remaining command form, x := E, except that instead of the induction hypothesis there is an appeal to Lemma 10 for E.

It remains to prove the (b) ⇒ (a) part of Theorem 12 for this, (b) can be specialized to the case of method calls. For lack of space the proof is in an appendix.

Lemma 18 If ST is satisfiable then it has supertype abstraction for method specifications provided that (12) implies (13) for all T and all m ∈ Meths T with mtype(m, T) = π : T → U, where

∀μ ∈ XMedEnv · μ |≡ | ST ⇒ S[Γ] ⊢ E : T, y : U]μ |≡ spec

∀μ ∈ MethEnv · μ |≡ | ST ⇒ D[Γ] ⊢ E : T, y : U]μ |≡ spec

6. Object invariants

This section formalizes a semantic account of invariants in which we can express what is achieved by invariant methodologies (Sec. 2.3) and prove soundness of supertype abstraction with invariants.

An object invariant is a predicate that holds of a particular object together with some others on which it depends. So an invariant suitable for objects of type T can be taken to be a subset of State[(self : T)]. In this section we extend the notion of specification table as follows.

Definition 19 A specification table, ST, consists of method specifications ST(T, m) exactly as in Definition 5 together with a predicate ST(T, m) ⊑ State[(self : T)], also written invT, for each ref type T.

Invariants serve as pre- and post-conditions for methods. For a method specification of the form ST(T, m) = (preT, postT) where

11 2006/9/3
mtpe(m, T) = \_T \rightarrow U, we will be concerned with specifications of the form

\[ \langle \text{pre}^T \cap \text{inv}^T, \text{post}^T \cap \text{inv}^T \rangle \] (14)

Taken literally this is nonsense —the intersections are empty— because \text{inv}^T is a subset of State[\{self : T\}] whereas \text{pre}^T is a subset of State[\{self : T, \_T \rightarrow T\}] and post^T is a subset of State[\{self : T, \_T \rightarrow T\} \times State[\{res : U, exc : Threw\}]. But there is a natural way to make sense of the intersections. For the precondition, we can define \text{inv}^T \subseteq State[\{self : T, \_T \rightarrow T\}] by \((p, h, r) \in \text{inv}^T \iff (p, h, r - \_T) \in \text{inv}^T. (Here r - \_T means to remove it from the domain of r.) Now the intersection of \text{inv}^T with \text{pre}^T makes sense. For the postcondition we interpret post^T \cap \text{inv}^T to mean the set of \((s, (p, h, r)) \text{ such that } (s, (p, h, r)) \in \text{post}^T and \((p, h, r - \_T) \in \text{inv}^T.\)

What is achieved by invariant methodologies is to arrange that in fact \text{inv}^T holds before any method call, so that the reasoning is sound. In this section our formalization adapts Def. 22 as follows.

**Definition 20 (satisfaction by method environment, \[\|\_\|\])** Let ST be a specification table. An extended method environment \(\mu\) satisfies ST, written \(\mu \models \|ST\|\), iff for all ref types \(T\) and \(m \in \text{Meths } T\) we have

\[ \mu(T, m) \models \langle \text{pre}^T \cap \text{inv}^T, \text{post}^T \cap \text{inv}^T \rangle \]

where \(\langle \text{pre}^T, \text{post}^T \rangle = ST(T, m)\) and \(\text{inv}^T \equiv ST(T)\).

A normal method environment \(\mu\) satisfies ST, written \(\mu \models \|ST\|\) iff for all classes \(C\) and \(m \in \text{Meths } C\) we have

\[ \mu(C, m) \models \langle \text{pre}^C \cap \text{inv}^C, \text{post}^C \cap \text{inv}^C \rangle \]

Our formalization in the preceding section treated a specification \(ST(T, m)\) both as the proof obligation for an implementation and as an assumption to be made for an invocation. Now we are considering the situation where the proof obligation is \(\|\_\|\).\)

What is achieved by a sound methodology has a simple semantic formulation: reasoning about a command \(S\) is carried out using the static-dispatch semantics as before but now the proof obligation for method implementations is in terms of \(\mu \models \|ST\|\) from Def. 20. That is, we adapt Def. 21 as follows.

**Definition 21** Let ST be a specification table for class table CT. Supertype abstraction is valid for ST, CT iff for all \(\Gamma \vdash S\) and all \(\Gamma\)-specifications \(\text{spec}\). implies \(\|\_\|\), where

\[ \forall \mu \in \text{MethEnv} \cdot \mu \models \|ST\| \Rightarrow S[\Gamma \vdash S] \mu \models \text{spec} \] (15)

\[ \forall \mu \in \text{MethEnv} \cdot \mu \models \|ST\| \Rightarrow D[\Gamma \vdash S] \mu \models \text{spec} \] (16)

The point is that (15) will be false unless \text{inv}^T holds as a precondition of every invocation on an object of dynamic type \(T\). Of course it is not an explicit precondition, since \text{inv}^T is a conjunction of several invariants which may be implicit in the methodology and/or not visible at the call site. The soundness result for a methodology must nonetheless ensure that the effective invariant, which we refer to as \text{inv}^T, does hold before every call.

The authors were surprised to find how little impact the addition of invariants has on the theory of supertype abstraction and behavioral subtyping. Definition of supertype abstraction does not need to change. Recall that ST has supertype abstraction for method specifications just if for all ref types \(T\), all \(m \in \text{Meths } T\), and all method specifications \(\text{spec}\) of type \(mtpe(T, m)\); if \(ST(T, m) \supseteq \text{spec}\) then \(ST(C, m) \supseteq \text{spec}\) for every class \(C\) with \(C \subseteq T\).

**Definition 22 (behavioral subtyping, with invariants)** A specification table ST has behavioral subtyping if and only if for all ref types \(T\) and classes \(C\) with \(C \subseteq T\) and all \(m \in \text{Meths } U\) we have \(ST(C, m) \supseteq ST(U, m)\) and moreover \(ST(C) \subseteq ST(U)\).

The revised version of Lemma 10 is no longer an equivalence.

**Lemma 23 (behavioral subtyping, with invariants)** Any satisfiable specification table with behavioral subtyping has supertype abstraction for method specifications.

This is an immediate consequence of Lemma 10; the relevant definitions have not changed except for the added invariant condition in Def. 22.

Unlike in Lemma 10 the converse does not hold. Supertype abstraction as we have defined it is only concerned with reasoning about method calls. Behavioral subtyping also imposes a condition on invariants. It would be straightforward to add “invariant reasoning” to the notion of supertype abstraction and thereby get an equivalence like Lemma 10, but it would shed no light. Thus the analog of Corollary 13, which we reframe from stating, is only an implication.

**Theorem 24 (supertype abstraction with invariants)** For any satisfiable ST the following are equivalent.

(a) ST has supertype abstraction for method specifications.

(b) Supertype abstraction is valid for ST, CT.

Remarkably, the proof looks just the same as the proof of Theorem 12. The reason is that the proof involves conditions dependent on \(\mu \models ST\) and \(\mu \models ST\) and various arguments about refinement and specifications. But it does not involve the definition of \(\mu \models ST\) or \(\mu \models ST\) per se. The argument goes through using \(\|\_\|\) instead of \(\models\).

What has changed is the interpretation of the result. Hypothesis 15 describes modular reasoning that \(S\) satisfies spec under the assumption that methods satisfy their specifications, in the revised sense of satisfaction that adjoints the invariant. In practice, this means the reasoning system has restricted the invariants and restricted \(S\) in such a way that an object’s invariant indeed holds as a precondition for every invocation.

The conclusion 10 says that indeed \(S\) is correct in the context of the real program semantics, provided that the implementations do satisfy their specifications with invariant adjoined.

Although supertype abstraction is based on \(\|\_\|\), behavioral subtyping (Def 22) is still defined using the relation \(\square\) applied to the pre/post specifications in the specification table.

7. Checking and ensuring behavioral subtyping

In light of the importance of behavioral subtyping, it is obviously useful to check whether given specifications satisfy the refinement condition. A characterization of \(\square\) is needed since the definition of quantities over all state transformers. This is the first topic of this section. The second topic is an alternative to checking: to impose behavioral subtyping by fiat.

**Characterizing refinement.** The most common formulation of behavioral subtyping uses the implications (2) and (3) that correspond to the rule of consequence in Hoare logic which derives a weaker specification from a stronger one. Even in Hoare logic for simple procedures these are incomplete. Hoare proposed an “adaptation rule” which is not complete and some subsequent proposals were found to have subtle unsoundness in connection with auxiliary variables in specifications (corrected in 4).

By now sound and complete rules are known and the connection with specification refinement has been made clear [13, 33, 41, 43].

To characterize when \(\text{pre}', \text{post}' \subseteq \text{pre}, \text{post}\) holds, it is not difficult to show necessity of \(\text{pre} \subseteq \text{pre}'\); it is the postconditions that are tricky. In our setting, one characterization is \(\text{old}(\text{pre}) \cap \text{post}' \subseteq \text{post}, \text{cf. (4)}\). There is an equivalent condition [14] that is similar to the join of specifications investigated later.

\[ \text{pre} \subseteq \text{pre}' \land \sim(\text{old}(\text{pre}') \cup \text{post}') \subseteq \sim(\text{old}(\text{pre}) \cup \text{post}) \]

where we write \(\sim p\) for the complement of \(p\) with respect to the set of all states of its type. Taking the type of self into account we have the following (the proof can be adapted from arguments in 13).

**Lemma 25 (characterization of refinement)** Let \(\text{pre}, \text{post}\) be a specification of type \(\Gamma \rightarrow \Gamma'\) and \(\text{pre}', \text{post}'\) be of type \(\Gamma[\text{self} : T] \rightarrow \Gamma'\) where \(T \subseteq \text{Gamma}\). Then \(\text{pre}', \text{post}' \supseteq \text{pre}, \text{post}\) if and only if

\[ \text{pre} \cap \text{p} \subseteq \text{pre}' \land \text{old}(\text{pre}) \cap \text{post}' \subseteq \text{post} \]

where \(p = \text{State}([\Gamma[\text{self} : T]])\).
Completeness of such a condition depends on the program semantics, since the definition of $\sqsubseteq^T$ quantifies over all program meanings. Moreover the presence of general auxiliary variables scoped over specifications can require a more complicated condition [33,41,43]. For the results in this paper, we found that using general auxiliary variables led to quite complicated semantic definitions so for perspicuity we chose the simpler form of specification with two-state postconditions. By disentangling the characterization of refinement from the use of refinement to define and reason about behavioral subtyping, we make it straightforward to adapt our results to other forms of specification.

**Specification inheritance.** Specification inheritance was pioneered by Wills [50,51] and is found in JML where it is made explicit by the `also` keyword. It is defined in terms of the join of specifications. Regardless of whether the class table is finite, there are only finitely many $U$ that are supertypes of a given $T$ so binary join suffices. The key property of specification inheritance is that it forces behavioral subtyping [14], as we show below.

Definition 26 (join of specifications) Let $(pre, post)$ and $(pre', post')$ be specifications of type $\Gamma \rightsquigarrow \Gamma'$. Define $(pre, post) \sqcup (pre', post')$ to be $(p, q)$ of the same type, where

\[
p = pre \cup pre',
q = (\neg old(pre) \cup post) \cap (\neg old(pre') \cup post').
\]

Lemma 27 (least upper bound) Consider specifications of some type $\Gamma \rightsquigarrow \Gamma'$. The specification $(pre, post) \sqcup (pre', post')$ is a least upper bound of $(pre, post)$ and $(pre', post')$ among specifications of type $\Gamma \rightsquigarrow \Gamma'$.

The proof is in Appendix C. Join only involves specifications at a single type, but for specification inheritance we need to consider its effect in terms of different types. To that end we need the following result which has a tedious but elementary proof in which the definitions are unfolded all the way to satisfaction.

Lemma 28 (monotonicity) Suppose $spec_0$ and $spec_1$ are specifications of some type $\Gamma \rightsquigarrow \Gamma'$ and suppose $T \subseteq \Gamma$ self. Suppose $spec_0$ and $spec_1$ are specifications of type $[\Gamma']$ self; $T$ and moreover $\forall i \geq 0$ and $i = 1$. Then $(spec_0 \cup spec_1) \sqsubseteq^T (spec_0 \cup spec_1)$.

Definition 29 (inheriting specifications) Let $ST$ be a specification table

\[
ST = (pre, post, T) \sqcup (pre', post', T')
\]

Recall from Def. 3 that $(pre, post)\ T$ denotes the restriction of the specification to states where self is $\leq T$.

Theorem 30 If $ST$ is the extension of $ST$ by specification inheritance then $ST$ has behavioral subtyping.

Proof If $T \leq V$ then $ST(T) \subseteq ST(V)$ because if $V \leq U$ then $T \leq U$ so $ST(T)$ is an intersection of more sets.

For methods, suppose $T \leq V$. Let $X$ be the set of $U$ such that $U \geq V$ and observe that $X$ is a subset, possibly a proper subset, of the set of $U$ with $U \geq V$. Now $ST(V, m)$ is the join of all $ST(U, m) \setminus V$ with $U \in X$. Let $JT$ be the join of $ST(U, m) \setminus V$ with $U \in X$. By the least upper bound property (Lemma 27), $ST(T, m) \sqsubseteq JT$ since $ST(T, m)$ joins over a set containing $X$. By definitions we have for every $T \geq V$ that $ST(U, m) \setminus V \sqsubseteq T$ and so by Lemma 28 we have that $JT \sqsubseteq^T ST(V, m)$. Provided that $JT$ is satisfiable, this can be put together with $ST(T, m) \sqsubseteq JT$ to get $ST(T, m) \sqsubseteq^T ST(V, m)$ (by Lemma 7). If $JT$ is unsatisfiable, then since $ST(T, m)$ is unsatisfiable and $ST(T, m) \sqsubseteq^T ST(V, m)$ is immediate by definition of $\sqsubseteq^T$.

8. Conclusions

We have formalized behavioral subtyping in terms of the intrinsic refinement ordering on specifications, where $spec \sqsubseteq spec'$ means that any program satisfying $spec$ also satisfies $spec'$. For $T$ to be a behavioral subtype of $U$ involves conditions like $spec_{U,m} \sqsubseteq U$ and $spec_{U,m}$ where $spec_{U,m}$ is the specification for method $m$ in class $U$. We have also formalized supertype abstraction: static reasoning about a program fragment using only specifications of methods it calls. The main result, Corollary 13 says that behavioral subtyping not only validates supertype abstraction but is equivalent to it, which we view as semantic completeness (see Thms. 12 and 24).

The semantic model has been encoded in the PVS theorem prover and type soundness proved, building on previous work [35]. Machine checking of the results of this paper is planned in the near future.

It is impractical to check $\sqsubseteq$ directly since its definition quantifies over all implementations. The contravariant/covariant relations $\sqsubseteq$ and $\sqsubseteq$, are a sound approximation but not complete. A corollary of Lemma 25, that behavioral subtyping is equivalent to a simple logical condition relating pre- and post-conditions in the relevant specifications. However, to apply the condition to check, say $spec_{T,m} \sqsubseteq spec_{U,m}$, requires reasoning about the entire specification whereas in practice the specification should be defined in terms of state in supertypes that is often not visible in the subtype, and vice versa. This is especially true for postconditions that express modifies specifications. This is one motivation for constructing the “effective specifications” by specification inheritance from arbitrary invariants and method specifications explicitly declared by the programmer. Theorem 20 confirms that the result has behavioral subtyping.

On the other hand, joining specifications in this way can lead to unimplementable specifications. It can be argued that, at least in situations where both super- and sub-type specifications are visible, failure of refinement should be detected as a design error [15] rather than masked by specification inheritance. Further investigation would be worthwhile to reconcile the issues pertaining to information hiding and reuse with the need to detect design flaws, leading to better tools and specification notations.

Another direction for future work is to extend our treatment of behavioral subtyping to other kinds of specifications and programming languages, especially concurrency and temporal logic. Such an extension would involve answering questions such as: what is the right notion of specification inheritance for temporal logic? Another direction would be to extend our treatment to explicitly deal with frame axioms (modifies clauses and other features of rich specifications) which is already quantified (auxiliaries).

An active area of research is to find easily used disciplines for making local definitions of relevant pieces of specifications — especially invariants — which can give an interpretation that validates modular reasoning. Our results serve to specify the semantic properties to be achieved by any sound discipline. Tools may be unsound or incomplete by design. Our results will help ensure that it is not for lack of theory or understanding.

Acknowledgments

Thanks to Paulo Borga and Augusto Sampaio for hosting us in Recife during our initial work on this paper and for stimulating discussions. Thanks to Shenghao Qin, Joseph Kiniry, Andreas Podelski, and Erik Poll for comments on earlier drafts.

References


A. Syntax

A class table CT is well formed provided it satisfies the following.

1. The subtype ordering is acyclic and T ≤ Object for all T.
2. Any ref type that appears as a field type, superclass, local variable type, cast etc. is declared in CT.
3. Field names are not shadowed, that is, if vis f : T is in fields T and super T = U then f is not in dfields U. (Note that the effect of the definition of fields is that fields are inherited.)
4. Method types—including parameter names—are invariant, that is, if mtype(U, m) is defined and T ≤ U then mtype(T, m) = mtype(U, m).
5. For any C, every method declaration m(Γ | T) : T {S} in CT(C) is typable in the sense that Γ ⊢ S where Γ = [self : C, res : T, θ : T]. Rules that define Γ ⊢ S appear just below.
6. For any C, any I ∈ superinterfaces C, and any method signature m(Γ | T) : T declared or inherited in I, there is a declared or inherited method in C with the same signature.
7. Variable exc does not occur anywhere. (It must be available for use in the semantics, and postconditions can refer to it.)

Typing rules for expressions

\[\begin{align*}
\Gamma \vdash 0 : \text{int} & & \Gamma \vdash x : \Gamma x & & \Gamma \vdash \text{new } C() \\
\Gamma \vdash \text{null} : T & & \Gamma \vdash x : T_1 & & \Gamma \vdash y : T_2 \\
\Gamma \vdash x : U & & (\text{vis } f : T) \in \text{dfields } V \\
U \leq V & & (\text{vis } = \text{private} \Rightarrow \Gamma \text{ self } = V) \\
& & (\text{vis } = \text{protected} \Rightarrow \Gamma \text{ self } \leq V) \\
\Gamma \vdash x : T & & U \leq T \quad T \in \text{RefType} \\
\Gamma \vdash (U) x : U & & & & & & \Gamma \vdash x : T & & U \leq T \quad T \in \text{RefType} \\
\Gamma \vdash x : T & & T \in \text{RefType} \\
\Gamma \vdash x : U \quad T \in \text{RefType} \\
\Gamma \vdash x.m(\overline{y}) : U & & \text{mtyp}(T, m) = \overline{z} : T \rightarrow U \\
\Gamma \vdash E : T & & [\Gamma, x : T] \vdash E_1 : U \\
\Gamma \vdash \text{let } x \text{ be } E \text{ in } E_1 : U
\end{align*}\]

Typing rules for commands

\[\begin{align*}
\Gamma \vdash E : T & & T \leq \Gamma x & & \Gamma \vdash x : T \\
\Gamma \vdash x : U & & (\text{vis } f : T) \in \text{dfields } V \\
U \leq V & & (\text{vis } = \text{private} \Rightarrow \Gamma \text{ self } = V) \\
& & (\text{vis } = \text{protected} \Rightarrow \Gamma \text{ self } \leq V) \\
\Gamma \vdash x : \text{bool} & & \Gamma \vdash S_1 & & \Gamma \vdash S_2 \\
\Gamma \vdash \text{if } x \text{ then } S_1 \text{ else } S_2 \\
\Gamma \vdash \text{var } x : T \text{ in } S \\
\Gamma \vdash S_1 & & \Gamma \vdash S_2 \\
\Gamma \vdash x : T & & T \leq \text{Thrwbl} \\
\Gamma \vdash \text{throw } x \\
\Gamma \vdash S_1 & & [\Gamma, x : T] \vdash S_2 \\
\Gamma \vdash \text{try } S_1 \text{ catch } (x : T) S_2 \\
\Gamma \vdash \text{try } S_1 \text{ finally } S_2
\end{align*}\]

B. Semantics of commands and expressions

The semantics of commands and expressions are defined in figures [B] and [D]. The semantics of expressions is defined by recursion on the structure of E. In case E has subexpressions, the definition uses nomenclature from the corresponding typing rule.

To streamline the semantics of commands, we define a helping function to create exceptional result states. Given ref context ρ, heap h ∈ H(e), class name C ≤ Thrwbl, and any type T we define except(ρ, h, T, C) to be an element of State[ρ, res : T, exc : Thrwbl] as follows.

except(ρ, h, T, C) = let o = fresh(ρ, h) in
let ρo = [ρ, o : C] in
let ho = [h, o : defaultObrecord C] in
(p0, ho, [res : default T, exc : o])

This is similar to the semantics of new C(), but the new object is assigned to exc rather than to res.

A similar helping function is used in the semantics of commands. Given (ρ, h, r) in State(Γ) and class name C ≤ Thrwbl we define except(ρ, h, r, C) to be an element of State[ρ, res : Thrwbl] as follows.

except(ρ, h, r, C) = let o = fresh(ρ, h) in
let ρo = [ρ, o : C] in
let ho = [h, o : defaultObrecord C] in
(p0, ho, [r, exc : o])

For any derivable typing Γ ⊢ E : T, any method environment μ, and any state (ρ, h, r, C) ∈ State(Γ), the definition of [Γ ⊢ E : T]μ(ρ, h, r) yields either ⊥ or an element of State[ρ, res : Thrwbl]. That is, [Γ ⊢ E : T] is an element of SemExpr(Γ, T).

For cast and type test, the typing rule ensures that type T (and hence type U) is a class or interface; this justifies application of ρ in the semantics.

For method call, the receiver object is rx so ρ(rx) is the dynamic type of the object; thus to look up in method environment μ the meaning of the dynamically dispatched method we write μ(ρ(rx))m. Since the argument expressions y are variables we can write r y for their values.

The semantics of commands is also defined by recursion on structure and using nomenclature from the typing rules. For any derivable typing Γ ⊢ S, the definition of [Γ ⊢ S] yields an element of SemCommand(Γ). For semantics of local variables we use another bit of notation: To remove an element from the domain of a function we use the minus sign, e.g., if r is a store then r − exc is the same store but with exc removed from its domain.


Semantics of class table. The semantics of a complete program is a method environment defined as a limit. The first step is to give a meaning for a method declaration mdec of the form meth m(τ; T) : T S in some class C. We define [mdec] to be a function in

\[ \text{MethEnv} \rightarrow \text{STrans}([\text{self} : C, \tau : T], [\text{res} : T, \text{exc} : \text{Thrwb}]) \]

as follows, using \( \Gamma = [\text{self} : C, \text{res} : T, \tau : T] \) so that \( \Gamma \vdash S \) (owing to condition 5 in the definition of well formed class table). For any method environment \( \mu \) and state \( (\rho, h, r) \) in \( \text{State}([\text{self} : C, \tau : T]) \), define

\[ [\text{mdec}]\mu(\rho, h, r) = \begin{cases} \text{let } r_0 = [r, \text{res} : \text{default } T] \text{ in } \mu(\rho, h, r_0) \text{ in (} (\rho_1, h_1, r_1) \in (\mu(\rho, h, r_0)) \text{ with default) } \end{cases} \]

where default is a fixed default value. The next step is to define an ascending chain \( \mu \in \mathbb{N} \rightarrow \text{MethEnv} \) of method environments as follows.

\( \mu_0(C, m) = \lambda S \cdot \bot \), for any \( m \) declared or inherited in \( C \).
\( \mu_j + 1(C, m) = [\text{mdec}]\mu_j \), if \( m \) is declared as mdec in \( C \).
\( \mu_j + 1(C, m) = \text{restr}((\mu_j + 1(B, m)), C) \), if \( m \) inherited in \( C \) from \( B \).

Here restr restricts the function the \( \mu_j + 1(B, m) \), which is defined on stores with self : \( B \), to stores with self : \( C \). This works because \( C \leq B \) implies \( [C] \subseteq [B] \) which in turn induces an inclusion for stores.

We take pains to make such conversions explicit throughout the paper. It is necessary for machine-checking the results. More importantly a key aspect of behavioral subtyping is the need for a method declared in some class \( C \) to satisfy a specification in which self has some different type \( T \geq C \).

Method environments are ordered by \( \mu \leq \mu' \) if \( \mu(C, m) \leq \mu'(C, m) \) for all \( C, m \). This refers to the usual ordering on state transformers: For \( \sigma \) and \( \tau \) in \( \text{STrans}(\Gamma, \Gamma') \), define \( \sigma \leq \tau \) if for all \( s \) in \( \text{State}(\Gamma) \) we have either \( \sigma = \tau \) or \( \sigma = \bot \). The everywhere-\( \bot \) function is the least element in the set of state transformers of a given type, and this induces a least method environment. These sets are closed under limits of ascending chains. For lack of space we refrain from proving that for any \( \Gamma \vdash S \), the semantics \( [\Gamma \vdash S] \) is a monotonic function from method environments to state transformers. Similarly, the semantics of a method declaration is monotonic in the method environment.

It follows that \( i \leq j \Rightarrow \mu_i \leq \mu_j \) for the approximation chain. The semantics \( \mu \) is defined to be the least upper bound of the approximation chain.\[ \square \]

Remark on closed world. The semantics is given for a class table which is a closed collection of class declarations. Although it is compositional at the level of commands, the semantics is not compositional at the level of classes: the semantics of a class table is not defined by composing a separate semantics for each class. Such a semantics, for a language with mutually recursive methods and mutually recursive class declarations, would be far more complex and less operationally transparent. But one may wonder whether our semantics is suitable for formulating results concerning modular reasoning. In fact we lose little or nothing, since our results quantify over arbitrary class tables. Indeed, there may be infinitely many classes and infinitely many methods in a class. So in some sense there is a universal class table containing all programs.

C. Additional proofs

**Proof of Lemma 18** Proof For any \( T, m \) and any spec of type mtype(m, T) we need to show that \( \text{ST}(T, m) \models^T \text{spec} \) implies \( \text{ST}(C, m) \models^C \text{spec} \) for all \( C \leq T \). This follows by weak transitivity from \( \text{ST}(C, m) \models^C \text{ST}(T, m) \) which we will prove.

The command \( z := \text{self}.m(\overline{y}) \) is chosen because we can unfold the semantics of \( z := \ldots \) as in the proof of Lemma 16 so that this command satisfies spec just if \( \text{self}.m(\overline{y}) \) satisfies an associated expression specification of type mtype(m, T). We refrain from giving the transformation on specifications and simply observe that for any \( \sigma \) that satisfies \( \text{ST}(T, m) \) there is \( \mu \) with \( \mu(T, m) = \sigma \) and \( \mu \models \text{ST} \). Moreover it is only \( \mu(T, m) \) that has any bearing on whether

\[ S[[\text{self} : T, \overline{y} : \overline{T}, z : U] \models z := \text{self}.m(\overline{y})] \text{ satisfies spec. So if we instantiate the antecedent by spec := ST(T, m) it amounts to } \forall \sigma \cdot \sigma \models \text{ST}(T, m) \Rightarrow \sigma \models \text{ST}(T, m) \text{ which is true. Thus we obtain the consequent with spec := ST(T, m), which boils down for any } C \leq T, \forall \sigma \in \text{SemMeth}(C, m) \cdot \sigma \models \text{ST}(C, m) \Rightarrow \sigma \models \text{ST}(T, m). \]

\[ \square \]

**Proof of Lemma 22** Proof To show that it is an upper bound, instantiate Lemma 25 with \( \text{pre}' := \text{pre} \cup \text{pre}' \) and \( \text{post}' := (\text{old}(\text{pre}) \cup \text{old}(\text{pre})) \cup \text{post}' \). Thus the upper bound property \((\text{old}(\text{pre}) \cup \text{post}) \cap (\text{old}(\text{pre}') \cup \text{post}') \subseteq (\text{pre}, \text{post}) \) follows from \( \text{pre} \leq \text{pre} \cup \text{pre}' \) which is immediate and \( \text{old}(\text{pre}) \cap (\text{old}(\text{pre}) \cup \text{post}) \cap (\text{old}(\text{pre}') \cup \text{post}') \subseteq \text{post} \) which also reduces to true.

To show that \( \cup \) gives the least upper bound, suppose \( (\text{pre}, \text{post}) \not\subseteq (\text{pre}_0, \text{post}_0) \) and \( (\text{pre}, \text{post}) \not\subseteq (\text{pre}_1, \text{post}_1) \). Now \( (\text{pre}, \text{post}) \not\subseteq (\text{pre}_0, \text{post}_0) \) and \( (\text{pre}_1, \text{post}_1) \) iff \( \sigma \models (\text{pre}, \text{post}) \) implies \( \sigma \models (\text{pre}_0, \text{post}_0) \) and \( (\text{old}(\text{pre}) \cup \text{post}) \cap (\text{old}(\text{pre}_0) \cup \text{post}_0) \) for all \( \sigma \). To prove the consequent for arbitrary \( \sigma \), consider any state \( t \) and suppose \( t \in \text{pre}_0 \cup \text{pre}_1 \). We must show that \( (t, \sigma t) \) is in \( (\text{old}(\text{pre}) \cup \text{post}) \) (and a symmetric condition). In the case \( t \in \text{pre}_0 \), we have \( (t, \sigma t) \in (\text{old}(\text{pre}_0) \cup \text{post}_0) \) immediately. For the case \( t \in \text{pre}_1 \), note that \( (\text{pre}, \text{post}) \not\subseteq (\text{pre}_0, \text{post}_0) \) have \( \text{pre} \not\subseteq \text{pre}_0 \). So we can use the antecedent \( \sigma \models (\text{pre}, \text{post}) \) to yield \( (t, \sigma t) \in \text{post} \), which concludes the proof that \( (t, \sigma t) \) is in \((\text{old}(\text{pre}) \cup \text{post}) \). \( \square \)

13 A similar semantics is used in [3], a characterization of least upper bounds and machine-checked proofs of the monotonicity properties etc. appears in [35].
\[ \Gamma \vdash x : T \mu(\rho, h, r) = (\rho, h, [\text{res: } r x, \text{exc: } \text{null}]) \]
\[ \Gamma \vdash \text{true : bool} \mu(\rho, h, r) = (\rho, h, [\text{res: true, exc: } \text{null}]) \]
\[ \Gamma \vdash 0 : \text{int} \mu(\rho, h, r) = (\rho, h, [\text{res: } 0, \text{exc: } \text{null}]) \]
\[ \Gamma \vdash \text{null : } T \mu(\rho, h, r) = (\rho, h, [\text{res: null, exc: } \text{null}]) \]
\[ \Gamma \vdash x = y : \text{bool} \mu(\rho, h, r) = \text{let } v = (\text{if } (r x = r y) \text{ then } \text{true else false}) \text{ in } (\rho, h, [\text{res: } v, \text{exc: } \text{null}]) \]
\[ \Gamma \vdash \text{new } C() : C \mu(\rho, h, r) = \]
\[ \quad \text{let } o = \text{fresh}(\rho, h) \text{ in let } o = [\rho, o : C] \text{ in let } o = [h, o : \text{defaultObrecord } C] \text{ in } (\rho o, h o, [\text{res: o, exc: } \text{null}]) \]
\[ \Gamma \vdash x : T \mu(\rho, h, r) = \text{if } r x \neq \text{null} \text{ then } (\rho, h, [\text{res: } h(r x) f, \text{exc: } \text{null}]) \text{ else except}(\rho, h, T, \text{NullDeref}) \]
\[ \Gamma \vdash (U) x : U \mu(\rho, h, r) = \text{if } r x = \text{null} \lor \rho(r x) \leq U \text{ then } (\rho, h, [\text{res: } r x, \text{exc: } \text{null}]) \text{ else except}(\rho, h, U, \text{ClassCast}) \]
\[ \Gamma \vdash x \in U : \text{bool} \mu(\rho, h, r) = \text{let } v = \text{if } r x \neq \text{null} \land \rho(r x) \leq U \text{ then } \text{true else false} \text{ in } (\rho, h, [\text{res: } v, \text{exc: } \text{null}]) \]
\[ \Gamma \vdash \text{let } x \text{ be } E \text{ in } E1 : U \mu(\rho, h, r) = \]
\[ \quad \text{let } (\rho o, h o, r o) = (\Gamma \vdash E : T) \mu(\rho, h, r) \text{ in } \]
\[ \quad \text{if } r o \text{ exc} \neq \text{null} \text{ then } (\rho o, h o, [\text{res: default } U, \text{exc: } r o \text{ exc}]) \text{ else let } r 1 = [r, x : r o \text{ res}] \text{ in } (\Gamma \vdash E1 : U) \mu(\rho o, h o, r 1) \]
\[ \Gamma \vdash x.m(\gamma) : U \mu(\rho, h, r) = \]
\[ \quad \text{if } r x = \text{null} \text{ then except}(\rho, h, U, \text{NullDeref}) \]
\[ \quad \text{else let } \tau : T \rightarrow U = mtype(m, T) \text{ in let } r 1 = [\text{self: } r x, \tau : r \gamma] \text{ in } \mu(\rho(r x)) m(\rho, h, r 1) \]

---

**Figure 3.** Semantics of expressions.

\[ \Gamma \vdash x := E \mu(\rho, h, r) = \]
\[ \quad \text{let } (\rho 1, h 1, r 1) = (\Gamma \vdash E : T) \mu(\rho, h, r) \text{ in if } r 1 \text{ exc} = \text{null} \text{ then } (\rho 1, h 1, [r : r 1 \text{ res}, \text{exc: } \text{null}]) \text{ else } (\rho 1, h 1, [r, \text{exc: } r 1 \text{ exc}]) \]
\[ \Gamma \vdash x.f := y \mu(\rho, h, r) = \text{if } r x \neq \text{null} \text{ then } (\rho, h, [r x.f, r y], r) \text{ else except}(\rho, h, r, \text{NullDeref}) \]
\[ \Gamma \vdash x \text{ if then } S 1 \text{ else } S 2 \mu(\rho, h, r) = \text{if } r x = \text{true} \text{ then } (\Gamma \vdash S 1) \mu(\rho, h, r) \text{ else } (\Gamma \vdash S 2) \mu(\rho, h, r) \]
\[ \Gamma \vdash \text{var } x : T \text{ in } S \mu(\rho, h, r) = (\rho 1, h 1, r 1) = (\Gamma \vdash S) \mu(\rho, h, [r x : \text{default } T]) \text{ in } (\rho 1, h 1, r 1 - x) \]
\[ \Gamma \vdash S 1; S 2 \mu(\rho, h, r) = \text{let } (\rho 1, h 1, r 1) = (\Gamma \vdash S 1) \mu(\rho, h, r) \text{ in if } r 1 \text{ exc} = \text{null} \text{ then } (\Gamma \vdash S 2) \mu(\rho 1, h 1, r 1 - \text{exc}) \text{ else } (\rho 1, h 1, r 1) \]
\[ \Gamma \vdash \text{throw } x \mu(\rho, h, r) = \text{if } r x \neq \text{null} \text{ then } (\rho, h, [r, \text{exc: } r x]) \text{ else except}(\rho, h, r, \text{NullDeref}) \]
\[ \Gamma \vdash \text{try } S 1 \text{ catch}(x : T) S 2 \mu(\rho, h, r) = \]
\[ \quad \text{let } (\rho 1, h 1, r 1) = (\Gamma \vdash S 1) \mu(\rho, h, r) \text{ in } \]
\[ \quad \text{if } r 1 \text{ exc} = \text{null} \lor \rho(r 1, \text{exc}) \leq T \text{ then } (\rho 1, h 1, r 1) \]
\[ \quad \text{else let } r 3 = [r 1 : r 1 \text{ res}, \text{exc}] \text{ in } \]
\[ \quad \text{let } (\rho 2, h 2, r 2) = (\Gamma ; x : T \vdash S 2) \mu(\rho 1, h 1, r 3) \text{ in } (\rho 2, h 2, r 2 - x) \]
\[ \Gamma \vdash \text{try } S 1 \text{ finally } S 2 \mu(\rho, h, r) = \]
\[ \quad \text{let } (\rho 1, h 1, r 1) = (\Gamma \vdash S 1) \mu(\rho, h, r) \text{ in } \]
\[ \quad \text{let } (\rho 2, h 2, r 2) = (\Gamma ; x : T \vdash S 2) \mu(\rho 1, h 1, r 1 - \text{exc}) \text{ in } \]
\[ \quad \text{if } r 2 \text{ exc} = \text{null} \text{ then } (\rho 2, h 2, [r 2, \text{exc: } r 1 \text{ exc}]) \text{ else } (\rho 2, h 2, r 2) \]

---

**Figure 4.** Semantics of commands.