Pragmatic Semantic Subtyping

LILY BROWN, ANDY FRIESEN, and ALAN JEFFREY, Roblox, USA

This paper presents the view of subtyping Luau programming language. This system has been deployed as part of the Luau programming language, used by millions of users of Roblox Studio.

CCS Concepts: • Software and its engineering \rightarrow Semantics.

ACM Reference Format:

1 INTRODUCTION

Luau is a scripting language used by Roblox creators in the IDE tool in Roblox Studio. In 2022 there were more than 4 million creators, Fig 1 [7], which is the largest user base of Semantic Subtyping. In [1] we discuss why Luau uses Semantic Subtyping:

Semantic subtyping interprets types as sets of values, and subtyping as set inclusion [3]. This is aligned with the minimize false positives goal of Luau non-strict mode, since semantic subtyping only reports a failure of subtyping when there is a value which inhabits the candidate subtype, but not the candidate supertype. For example, the program:

In the previous, syntax-driven, implementation of subtyping, this subtype check would fail, resulting in a false positive. We have now released an implementation of semantic subtyping, which does not suffer from this defect. See our technical blog for more details [5].

In Luau, we use a variant of semantic subtyping [3, 4, 6]. The important properties of semantic subtyping are:

- there is a set \mathcal{D} of semantic values,
- each type T has a semantics $[\![T]\!] \subseteq \mathcal{D}$,
- unknown and never types are interpreted as $\mathcal D$ and \emptyset ,
- union and intersection types are interpreted as set union and intersection, and
- subtyping T <: U is interpreted as $[T] \subseteq [U]$.

Authors' address: Lily Brown; Andy Friesen; Alan Jeffrey, Roblox, San Mateo, CA, USA.

This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2024 Roblox.

(CFrame, Vector3 | CFrame) -> (Vector3 | CFrame)

2475-1421/2024/9-ART?? \$15.00

https://doi.org/10.1145/nnnnnnn.nnnnnnn



Fig. 1. Creators numbers in 2022

The off-the-shelf presentation of semantic subtyping is set theoretic [3, §2.5]:

$$[\![T_1]\!] \subseteq [\![T_2]\!]$$
 if and only if $\mathcal{E}[\![T_1]\!] \subseteq \mathcal{E}[\![T_2]\!]$

where the most important case for $\mathcal{E}[T]$ is function types:

$$\mathcal{E}[\![S \to T]\!] = \mathcal{P}(\mathcal{D}^2 \setminus ([\![S]\!] \times (\mathcal{D} \setminus [\![T]\!])))$$

The set theoretical requirement has some consequences:

All functions types (never $\rightarrow T$) are identified. Consider

$$\mathcal{E}[[\text{never} \to T_1]] = \mathcal{P}(\mathcal{D}^2 \setminus ([[\text{never}]] \times (\mathcal{D} \setminus [[T_1]]))) = \mathcal{P}(\mathcal{D}^2 \setminus (\emptyset \times (\mathcal{D} \setminus [[T_1]]))) = \mathcal{P}(\mathcal{D}^2) = \mathcal{P}(\mathcal{D}^2 \setminus (\emptyset \times (\mathcal{D} \setminus [[T_2]]))) = \mathcal{P}(\mathcal{D}^2 \setminus ([[\text{never}]] \times (\mathcal{D} \setminus [[T_2]]))) = \mathcal{E}[[[\text{never}]] \to T_2]]$$

in particular, this means we cannot define a semantics-preserving function apply (T, U) such that:

$$apply(S \rightarrow T, U) = T \text{ when } U <: S$$

because there is a nasty case where *S* is uninhabited. In this presentation, the apply function used in the rule for function application:

$$\frac{D_1: (\Gamma \vdash M:T)}{D_2: (\Gamma \vdash N:U)}$$

$$\frac{pp(D_1,D_2): (\Gamma \vdash M(N): \mathsf{apply}(T,U))}{\mathsf{app}(D_1,D_2): (\Gamma \vdash M(N): \mathsf{apply}(T,U))}$$

so we have to accept that in a set-theoretic model, the type rule for function application has corner cases for uninhabited types.

```
106
107
108
109
110
111
112
113
114
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
```

```
v ::= s \mid (a \mapsto r)

a ::= () \mid (v)

r ::= diverge \mid check \mid (v)
```

Fig. 2. Semantic values

Union does not distributed through function types. Semantic subtyping gives a natural model of overloaded functions as intersections of arrows, for example the Roblox API for matrices include an overloaded function which supports multiplication of both 2D (CFrame) and 1D (Vector3) matrices:

```
CFrame.__mul : ((CFrame, CFrame) -> CFrame) & ((CFrame, Vector3) -> Vector3)
```

Overloaded functions are a key part of the Roblox API, and we might expect that all function types can be presented as overloaded functions. We can do this if we can present unions of overloaded functions as overloaded functions. Now, union distributes through intersection, so all that is required is to distribute union through arrow:

$$[\![(S_1 \to T_1) \cup (S_2 \to T_2)]\!] = [\![(S_1 \cap S_2) \to (T_1 \cup T_2)]\!]$$

For example:

```
[(\text{number}? \rightarrow \text{number}) \cup (\text{string}? \rightarrow \text{string})] = [[\text{nil} \rightarrow (\text{number} \cup \text{string})]]
```

Unfortunately, set-theoretic models do not allow union to distributed through intersection, for example:

```
 \begin{array}{lll} \{(1, \operatorname{nil}), (\text{"hi"}, \operatorname{nil})\} & \in & \mathcal{E}[\![\!\operatorname{nil} \to (\operatorname{number} \cup \operatorname{string})]\!] \\ \{(1, \operatorname{nil}), (\text{"hi"}, \operatorname{nil})\} & \notin & \mathcal{E}[\![\![\!\operatorname{number}? \to \operatorname{number}]\!] \\ \{(1, \operatorname{nil}), (\text{"hi"}, \operatorname{nil})\} & \notin & \mathcal{E}[\![\![\!\operatorname{string}? \to \operatorname{string}]\!] \end{array}
```

This is why type normalization for function types in set-theoretic models uses a conjunctive normal form of unions of intersections of functions e.g. [6, §4.1.2].

Set-theoretc mode support negatived types. In addition, Luau does not support negation of all types, but only negation of *test types* [?], which simplifies the model, by not requiring arbitrary type negation. In particular, since the model does not support negation of function types, the normal form for function types is just overload functions, not combinations of positive and negative function types.

Concusions of this paper. In summary there is a trade-off in semantic subtyping:

- set-theoretic models, which are closer to the set-theoretic model of functions, and
- *pragmatic* models, which drop the set-theoretic requirement, and in return a) do not have corner cases on the type of function application when the argument has uninhabited type, and b) have overloaded functions (that is intersections of arrows) as the normal for function types.

Luau chooses to adopt a pragmatic semantic subtyping model.

This paper shows how Luau pragmatic model is defined, and how it supports benefits of pragmic models.

150

151

153

157

159

161 162

163

164

165

167

169

170

171

172

173 174

175

176

177

178

179

180

181

182

183

184

185 186

187

188 189

190

191

192

193

194

195 196

2 SEMANTIC SUBTYPING FOR LUAU

2.1 Semantic values for Luau

In this presentation, we presect the minimal core of Luau, which supports scalars and functions. This presentation ingores tables, mutable features, and object obects. We will ignore the details of scalar types, and assume that:

- there are scalar types, ranged over by b, such as nil, boolean, number and string,
- there are scalar values, ranged over by s, such as nil, true, false, numbers and string literals, and
- each scalar type s has a set of scalar values $\langle \langle s \rangle \rangle$, such as:

$$\langle (nil) \rangle = \{nil\} \quad \langle (boolean) \rangle = \{true, false\} \quad \langle (number) \rangle = \{0, 1, \dots\} \quad \cdots$$

The types we are considering are:

$$S, T := b \mid \text{unknown} \mid \text{never} \mid S \rightarrow T \mid S \cap T \mid S \cup T$$

which are:

- the scalar types b
- the anything type unknown,
- the uninhabited type never,
- a function type $S \to T$,
- an *intersection* type $S \cap T$, and
- a union type $S \cup T$.

To give a semantic subtyping, we first declare the domain \mathcal{D} of semantic values, given by the grammar v of Figure 2. Semantic values are:

- scalar values s, and
- function values $a \mapsto r$, modeling a function that can can map an argument a to a result r.

For example:

- true and false are values in boolean,
- true and false and nil are values in the optional type boolean ∪ nil, and
- (true) \mapsto (false) is a value in the function type boolean \rightarrow boolean.

Scalar and error-suppressing values are relatively straightforward, but functions are trickier. The case where a type-correct argument is supplied and a type-correct result is returned is clean, for example:

$$((true) \mapsto (false)) \in [boolean \rightarrow boolean]$$

But there is also the case where a type-incorrect argument is supplied, in which case there is no guarantee what is returned, for example:

$$((5) \mapsto (37)) \in \llbracket boolean \rightarrow boolean \rrbracket$$

The type-correctness guarantee for results applies when a type-correct argument is provided:

$$((true) \mapsto (37)) \notin \llbracket boolean \rightarrow boolean \rrbracket$$

Those examples consider cases where one value is supplied as an argument, and one is returned, but Luau allows other cases. Luau, as is common in most functional languages, allows functions to diverge (modeled in this semantics as $a \mapsto$ diverge)., for example:

$$((true) \mapsto diverge) \in [boolean \rightarrow boolean]$$

and:

$$((5) \mapsto \mathsf{diverge}) \in \llbracket \mathsf{boolean} \to \mathsf{boolean} \rrbracket$$

Proc. ACM Program. Lang., Vol. 7, No. ICFP, Article ??. Publication date: September 2024.

197

198

205206207208

204

211 212

210

214215216217

218 219 220

221 222 223

225226227

224

228229230231

232

233

234 235 236

237

238 239

240 241

241242243

244 245

Fig. 3. Semantics of types as sets of values

Luau allows functions to check arguments, (modeled in this semantics as $a \mapsto$ check when a checked fails), for example:

 $((5) \mapsto \mathsf{check}) \in \llbracket \mathsf{boolean} \to \mathsf{boolean} \rrbracket$

but:

$$((true) \mapsto check) \notin \llbracket boolean \rightarrow boolean \rrbracket$$

Luau allows functions to be called without any arguments (modeled in this semantics as $() \mapsto r)$ for example:

 $(() \mapsto (\texttt{false})) \in [\![\texttt{boolean} \to \texttt{boolean}]\!]$

and:

and:

 $(() \mapsto \mathsf{diverge}) \in \llbracket \mathsf{boolean} \to \mathsf{boolean} \rrbracket$

 $(() \mapsto \mathsf{check}) \in \llbracket \mathsf{boolean} \to \mathsf{boolean} \rrbracket$

The restriction on zero-argument function calls is that they are allowed to return a check (since they have been passed the wrong number of arguments) but they are not just allowed to return arbitrary nonsense:

$$(() \mapsto (5)) \notin \llbracket boolean \rightarrow boolean \rrbracket$$

At this point we have introduced the semantic values used by the Luau type system, and can turn the semantics of types, from which semantic subtyping follows.

2.2 Semantics of Luau type

The semantics of Luau types are given in Fig 3. This semantics is presented mechanically in Agda in [2], and we will give the most important results here.

For example, two of the important rules are for functions, in the case where functions are called with argument values, and return result values. The rules are:

- Type-incorrect argument: if $v \notin \llbracket S \rrbracket$ then $((v) \mapsto r) \in \llbracket S \to T \rrbracket$
- Type-correct result: if $w \in \llbracket T \rrbracket$ then $(a \mapsto (w)) \in \llbracket S \to T \rrbracket$

This is the same as the semantics of Coppo types [?] as used in the fully abstract semantics of Lazy Lambda Calculus [?] using Domain Theory In Logical Form [?].

$$((v) \mapsto (w)) \in \llbracket S \to T \rrbracket \text{ if and only if } (v \in \llbracket S \rrbracket) \Rightarrow (w \in \llbracket T \rrbracket)$$

In order to give a constructive presentation of the semantics, rather than usual negative presentation of $v \notin \llbracket S \rrbracket$, we give a positive presentation $v \in \llbracket S \rrbracket^{\mathbb{C}}$, as given in Fig 4.

```
246
247
248
249
250
251
252
253
254
255
256
257
```

 Fig. 4. Complemented semantics of types as sets of values

It is routine to check that $[S \cap T]^{\mathbb{C}}$ is a constructive presentation of $\mathcal{D} \setminus [S \cap T]$.

Lemma 2.1. $(v \in \llbracket T \rrbracket^{\complement})$ if and only if $(v \notin \llbracket T \rrbracket)$.

Moreover there is a decision procedure for $v \in \llbracket T \rrbracket$ or $v \in \llbracket T \rrbracket^{\complement}$.

Lemma 2.2. $(v \in [\![T]\!]) \lor (v \in [\![T]\!]^{\widehat{\mathbb{C}}}).$

2.3 Properties of semantic subtyping

From the semantics of types as sets of semantic values, semantic subtyping S <: T is a proof than if $v \in \llbracket S \rrbracket$, then $v \in \llbracket T \rrbracket$. Constructively this a dependent function:

$$(S \mathrel{<:} T)$$
 if and only if $\forall v \: . \: (v \in \llbracket S \rrbracket) \to (v \in \llbracket T \rrbracket)$

for example number <: number? since:

$$\forall v . (v \in \llbracket \text{number} \rrbracket) \rightarrow (v \in \llbracket \text{number}? \rrbracket)$$

Subtyping can be viewed as a dependent function from $[\![T]\!]^{\mathbb{C}}$ to $[\![S]\!]^{\mathbb{C}}$

Lemma 2.3.
$$(S \lt: T)$$
 if and only if $\forall v . (v \in \llbracket T \rrbracket^{\complement}) \rightarrow (v \in \llbracket S \rrbracket^{\complement})$

PROOF. For "if", for any v, if $v \in [S]$, then by Lemma 2.2 either $v \in [T]$ or $v \in [T]$. In the first case, $v \in [T]$ as needed. In the second case $v \in [T]$ and so by "if" hypothesis $v \in [S]$, but by Lemma 2.1 we have a contradiction from $v \in [S]$ and $v \in [S]$. So we have established that S <: T.

For "only if", for any v, if $v \in \llbracket T \rrbracket^{\mathbb{C}}$, then by Lemma 2.2 either $v \in \llbracket S \rrbracket$ or $v \in \llbracket S \rrbracket^{\mathbb{C}}$. In the first case $v \in \llbracket S \rrbracket$ and so by "only if" hypothesis, $v \in \llbracket T \rrbracket$, but by Lemma 2.1 we have a contradiction from $v \in \llbracket T \rrbracket$ and $v \in \llbracket T \rrbracket^{\mathbb{C}}$. In the second case, $v \in \llbracket S \rrbracket^{\mathbb{C}}$ as needed. So we have established that $\forall v \cdot (v \in \llbracket T \rrbracket^{\mathbb{C}}) \to (v \in \llbracket S \rrbracket^{\mathbb{C}})$.

More interestingly is the constructive presentation of *anti-subtyping* $S \nleq T$. Normally this is presented negatively, but it can be read constructively since $S \nleq T$ is witnessed by a value v where $v \in \llbracket S \rrbracket$ but $v \in \llbracket T \rrbracket^{\complement}$.

$$(S \nleq : T)$$
 if and only if $\exists v . (v ∈ \llbracket S \rrbracket) \land (v ∈ \llbracket T \rrbracket^{\complement})$

 for example number? $\not <$: number since we can pick our witness v to be nil:

$$nil \in [number?]$$
 $nil \in [number]^{C}$

Now, by Lemma 2.1, it is direct that $S \nleq T$ is a contradiction of $S \lt T$:

LEMMA 2.4.
$$(S \angle: T) \rightarrow \neg (S <: T)$$

Unfortunately, this does not give a decision procedure for subtyping, for the usual reason that it tricky to build an algorithm for checking semantic subtyping, which requires type normalization [?]. We will return to this in §2.6.

It is direct to show that <: is transitive.

LEMMA 2.5.
$$(S <: T) \land (T <: U) \rightarrow (S <: U)$$

More interestingly there is a dual property for $\not\prec$:. Classically this is the same as transitivity, just stated in terms of $\not\prec$: rather than \prec :. But constructively this is a choice function, that states that if $S \not\prec$: U then for any T we have a witness for either $S \not\prec$: T or $T \not\prec$: U. For example number? $\not\prec$: number (witnessed by nil), so for a mid-point string we have nil \in [string] $^{\mathbb{C}}$ which means we chose the constructive anti-subtype number? $\not\prec$: string witnessed by nil \in [number?] and nil \in [string] $^{\mathbb{C}}$.

```
Lemma 2.6. (S \not<: U) \rightarrow (S \not<: T) \lor (T \not<: U)
```

PROOF. $(S \nleq U)$ must have a witness v where $v \in \llbracket S \rrbracket$ and $v \in \llbracket U \rrbracket^{\mathbb{C}}$. Now by Lemma 2.2 (the decision procedure for type semantics) we either have $v \in \llbracket T \rrbracket$ or $v \in \llbracket T \rrbracket^{\mathbb{C}}$. In the first case, $v \in \llbracket T \rrbracket$ and $v \in \llbracket U \rrbracket^{\mathbb{C}}$, which witnesses $T \nleq U$. In the second case, $v \in \llbracket S \rrbracket$ and $v \in \llbracket T \rrbracket^{\mathbb{C}}$, which witnesses $S \nleq T$. In either case, we have a decision procedure for $(S \nleq T) \lor (T \nleq U)$.

2.4 Co- and contra-variance of subtyping for fuctions

We now turn to co- and contra-variant subtyping of functions. These come in two flavors: when function types are introduced, and when function types are eliminated. When a function type is introduced, we we check that the arguments respect contravariant subtyping, and that the results respect covariant subtyping.

```
Lemma 2.7. If S' <: S and T <: T' then (S \rightarrow T) <: (S' \rightarrow T')
```

PROOF. If $u \in \llbracket S \to T \rrbracket$ then from Fig 3, either

- $u = (a \mapsto (w))$ and $w \in [T]$, so T <: T' implies $w \in [T']$, and so $(a \mapsto (w)) \in [S' \to T']$,
- $u = ((v) \mapsto r)$ and $v \in \llbracket S \rrbracket^{\widehat{\mathbb{C}}}$, so S' <: S and Lemma 2.3 implies $v \in \llbracket S' \rrbracket^{\widehat{\mathbb{C}}}$, and so $((v) \mapsto r) \in \llbracket S' \to T' \rrbracket$,
- $u = (a \mapsto \text{diverge})$, so $(a \mapsto \text{diverge}) \in [S' \to T']$, or
- $u = () \mapsto \text{check}$, so $(() \mapsto \text{check}) \in [\![S' \to T']\!]$.

In any case, $u \in \llbracket S' \to T' \rrbracket$.

When a function type is eliminated, we we check that the arguments reflect contravariant subtyping, and that the results reflect covariant subtyping.

```
Lemma 2.8. If (S \rightarrow T) <: (S' \rightarrow T') then S' <: S and T <: T'
```

PROOF. If $v \in \llbracket S \rrbracket^{\complement}$ then from Fig 3 $((v) \mapsto \text{check}) \in \llbracket S \to T \rrbracket$, so since $(S \to T) <: (S' \to T')$, we have $((v) \mapsto \text{check}) \in \llbracket S' \to T' \rrbracket$ and so $v \in \llbracket S' \rrbracket^{\complement}$. Thus, using Lemma 2.3, we have established S' <: S.

If $w \in \llbracket T \rrbracket$ then from Fig 3 (() \mapsto (w)) $\in \llbracket S \to T \rrbracket$, so since $(S \to T) <: (S' \to T')$, we have (() \mapsto (w)) $\in \llbracket S' \to T' \rrbracket$ and so $w \in \llbracket T' \rrbracket$. Thus we have established T <: T'.

Note that these Lemmas rely on pragmatic semantic subtyping. Lemma 2.7 is true for set-theoretic semantic subtyping, but Lemma 2.8 is only true for set-theoretic models when S and T are inhabited types. In pragmatic sematic subtyping, we do not have special corner cases, in particular Lemma 2.8 is true for all types, and does not require special cases about inhabitance.

2.5 Distritution of intersection and union over fuctions

Finally, we turn to cases where intersection and union disturiute through functions. Since Luau uses intersections of functions as types for overloaded functions, it is unsupringing that interesection does not in general distribute through functions. For example:

$$(boolean \rightarrow boolean) \cap (number \rightarrow number)$$

does not distribute to:

$$(boolean \cup number) \rightarrow (boolean \cap number)$$

For example:

```
 (false \mapsto true) \in [\![ (boolean \to boolean) \cap (number \to number) ]\!] \\ (false \mapsto true) \in [\![ (boolean \cup number) \to (boolean \cap number) ]\!]^{\mathbb{C}}
```

is a witness for:

Now in general we can not distribute intersetion through functions, on both the left and right, but we can distrutue on just left, and just right. This is similar to the situation with premonids categories [?] which are functorial on both sides, but are not binary functorials.

LEMMA 2.9.

```
(1) [[(S_1 \to T) \cap (S_2 \to T)]] = [[(S_1 \cup S_2) \to T]]
(2) [[(S \to T_1) \cap (S \to T_2)]] = [[S \to (T_1 \cap T_2)]]
```

In contract, union does distructibute over fuctions.

Lemma 2.10.
$$[(S_1 \to T_1) \cup (S_2 \to T_2)] = [(S_1 \cap S_2) \to (T_1 \cup T_2)]$$

2.6 Type normalization

3 FURTHER WORK

REFERENCES

- L. Brown, A. Friesen, and A. S. A. Jeffrey. 2023. Goals of the Luau Type System: Two Years Only. In Proc. Human Aspects of Types and Reasoning Assistants. https://asaj.org/papers/hatra23.pdf
- [2] L. Brown and A. S. A. Jeffrey. 2023. Luau Prototype Typechecker. https://github.com/luau-lang/agda-typeck
- [3] G. Castagna and A. Frisch. 2005. A Gentle Introduction to Semantic Subtyping. In *Proc. Principles and Practice of Declarative Programming*.
- [4] A. Frisch, G. Castagna, and V. Benzaken. 2008. Semantic subtyping: Dealing set-theoretically with function, union, intersection, and negation types. J. ACM 55, 19 (2008).
- [5] A. S. A. Jeffrey. 2022. Semantic Subtyping in Luau. Roblox Technical Blog. https://blog.roblox.com/2022/11/semantic-subtyping-luau/
- [6] A. M. Kent. 2021. Down and Dirty with Semantic Set-theoretic Types (a tutorial). https://pnwamk.github.io/sst-tutorial/
- [7] Roblox. 2022. Roblox Coporation 2022 Annual Report. https://ir.roblox.com/financials/annual-reports/default.aspx