



A Grounded Conceptual Model for Ownership Types in Rust

WILL CRICHTON, Brown University, USA

GAVIN GRAY, ETH Zürich, Switzerland

SHRIRAM KRISHNAMURTHI, Brown University, USA

Programmers learning Rust struggle to understand ownership types, Rust's core mechanism for ensuring memory safety without garbage collection. This paper describes our attempt to systematically design a pedagogy for ownership types. First, we studied Rust developers' misconceptions of ownership to create the Ownership Inventory, a new instrument for measuring a person's knowledge of ownership. We found that Rust learners could not connect Rust's static and dynamic semantics, such as determining why an ill-typed program would (or would not) exhibit undefined behavior. Second, we created a conceptual model of Rust's semantics that explains borrow checking in terms of flow-sensitive permissions on paths into memory. Third, we implemented a Rust compiler plugin that visualizes programs under the model. Fourth, we integrated the permissions model and visualizations into a broader pedagogy of ownership by writing a new ownership chapter for *The Rust Programming Language*, a popular Rust textbook. Fifth, we evaluated an initial deployment of our pedagogy against the original version, using reader responses to the Ownership Inventory as a point of comparison. Thus far, the new pedagogy has improved learner scores on the Ownership Inventory by an average of 9% ($N = 342$, $d = 0.56$).

CCS Concepts: • **Theory of computation** → **Type structures**; • **Human-centered computing** → *Visualization systems and tools*; • **Social and professional topics** → *Computing education*.

Additional Key Words and Phrases: Rust, ownership types, program state visualization, concept inventory

ACM Reference Format:

Will Crichton, Gavin Gray, and Shriram Krishnamurthi. 2023. A Grounded Conceptual Model for Ownership Types in Rust. *Proc. ACM Program. Lang.* 7, OOPSLA2, Article 265 (October 2023), 29 pages. <https://doi.org/10.1145/3622841>

1 INTRODUCTION

Ownership is a programming discipline for managing the aliasing and mutation of data, enforced statically through ownership types. The flagship programming language for ownership is Rust, which empowers programmers to write memory-safe code without garbage collection. Rust's ownership model synthesizes several ideas from PL research such as linear logic [Girard 1987], class-based alias management [Clarke et al. 1998], and region-based memory management [Grossman et al. 2002]. History shows that developers cannot write memory-safe C and C++ in practice [MSRC Team 2019], so the software industry is turning toward Rust. For example, Google's Android team has found zero memory vulnerabilities in 1.5 million lines of Rust code [Vander Stoep 2022].

This rosy picture of PL tech transfer belies a persistent obstacle: teaching ownership types to prospective users. Over the last four years, studies have found that Rust learners struggle to fix ownership type errors [Zeng and Crichton 2019; Zhu et al. 2022], and users self-report that

Authors' addresses: Will Crichton, Department of Computer Science, Brown University, Providence, Rhode Island, 02912, USA, wcrichto@brown.edu; Gavin Gray, Department of Computer Science, ETH Zürich, Zürich, Switzerland; Shriram Krishnamurthi, Department of Computer Science, Brown University, Providence, Rhode Island, 02912, USA.



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

© 2023 Copyright held by the owner/author(s).

2475-1421/2023/10-ART265

<https://doi.org/10.1145/3622841>

ownership is among their biggest barriers to learning Rust [Fulton et al. 2021; The Rust Survey Team 2020]. To wit: advances in the technical factors of type systems require commensurate advances in the human factors of type systems.

Our work started with the question: how can we systematically design a pedagogy for ownership types? Today, popular pedagogies for advanced type systems are driven by experts' intuitions about how people learn, as well as by intuitions about what makes type systems difficult to understand. As practicing (computer) scientists, we wanted to approach pedagogic design through scientific principles: grounding the pedagogy in observations about the struggles of Rust learners, and then evaluating the pedagogy by its effects on learning outcomes. This paper describes how we put these principles into practice:

- (1) **We ran a formative user study to identify misconceptions about ownership types held by Rust learners** (Section 2). We designed a new instrument for evaluating understanding of ownership, the Ownership Inventory, by drawing tasks from commonly-reported Rust issues on StackOverflow. We studied $N = 36$ Rust learners trying to solve Ownership Inventory problems. We found that learners can generally identify the surface-level reason for why a program is ill-typed with respect to ownership. However, they do not understand what undefined behavior (if any) would occur if an ill-typed program were executed. This misunderstanding is reflected in often inefficient and incorrect strategies that participants used to fix ownership errors.
- (2) **We developed a conceptual model of ownership types to address these misconceptions** (Sections 3.1.2 and 3.2.2). The conceptual model represents the aspects of Rust's static and dynamic semantics that are relevant to ownership, while abstracting other details. The model provides learners a foundation to understand essential concepts such as undefined behavior and the incompleteness of Rust's ownership type-checker, or "borrow checker."
- (3) **We implemented tools to visualize Rust programs under the conceptual model** (Sections 3.1.3 and 3.2.3). We leveraged existing executable models of Rust's dynamic and static semantics to generate traces for a given Rust program. We visualize these traces to help learners "see" the abstract conceptual model reified into concrete examples.
- (4) **We designed a pedagogy around our conceptual model to explain ownership types** (Section 4). We wrote a chapter on ownership that explains how and why Rust's type system prevents undefined behavior, illustrating code with diagrams generated by our tool. We integrated this chapter into a popular Rust textbook, *The Rust Programming Language* (TRPL) [Klabnik and Nichols 2022]. We setup and advertised a public website that hosts our TRPL fork.
- (5) **We A/B tested our pedagogy against the TRPL baseline** (Section 5). We measured learning outcomes with two kinds of quizzes: simpler comprehension questions about the conceptual model, and more difficult multiple-choice versions of the Ownership Inventory. Learners could correctly answer comprehension questions with 72% accuracy. Our initial deployment improved the average Inventory score from 48% to 57% ($N = 342, p < 0.001, d = 0.56$).

The thesis of our pedagogy is that to understand ownership types, Rust learners need to understand two key concepts: undefined behavior and incompleteness. What are the "stuck states" of Rust's dynamic semantics? Why does Rust's static semantics avoid stuck states? What valid programs are rejected by the static semantics? Existing Rust pedagogies accurately characterize the syntactic properties enforced by the compiler. However, they fall short of explaining the counterfactual undefined behavior that could occur without the borrow checker, especially regarding memory safety. For example, three popular Rust books explain mutable references like this:

- *Rust in Action* [McNamara 2021]: "Borrows can be read-only or read-write. Only one read-write borrow can exist at any one time." (No justification is provided.)

- *Hands-on Rust* [Wolverson 2021]: “Rust’s safety features make a mutable borrow *exclusive*. If a variable is mutably borrowed, it cannot be borrowed—mutably or immutably—by other statements while the borrow remains.” (No justification is provided.)
- *The Rust Programming Language* [Klabnik and Nichols 2022]: “If you have a mutable reference to a value, you can have no other references to that value. [...] The benefit of having this restriction is that Rust can prevent data races at compile time. [...] Users of an immutable reference don’t expect the value to suddenly change out from under them!”

Of these resources, only TRPL provides a sense of the counterfactual by abstractly gesturing towards the issue of data races. But learners likely need to see concrete “negative” examples of counterfactual behavior to better grasp the overall logic of ownership [Dyer et al. 2022]. Furthermore, none of these resources explain incompleteness, or even suggest that safe programs can be rejected by the compiler. Programmers need different strategies to fix ownership errors if a program is rejected due to incompleteness [Crichton 2020].

To address these issues, we designed our pedagogy around a new conceptual model of Rust’s dynamic and static semantics. To explain undefined behavior, we “turn off” Rust’s borrow checker to interpret and visualize programs that would otherwise be rejected by the compiler. These counterfactual visualizations help learners understand undefined behavior that is avoided by the compiler. To explain incompleteness, we reframe ownership type-checking as a form of abstract interpretation. The abstract state of the program is a mapping from paths to permissions such as “readable” or “writable.” We visualize these abstract permission states to show learners how incompleteness arises from phenomena such as field-insensitivity and limitations of the alias analysis within the borrow checker.

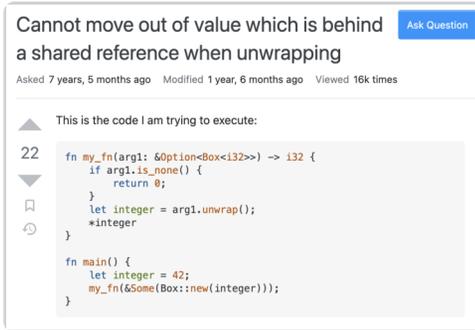
2 A CONCEPT INVENTORY FOR OWNERSHIP

To develop a pedagogy for ownership types, we first sought to understand the core misconceptions that underlie the struggles of Rust learners. Many learning resources are developed based on educators’ intuitions about what makes concepts difficult. Instead, we sought to ground our pedagogy in experimental data about the experiences of Rust learners. Our overarching methodology was the development of a *concept inventory* for ownership, henceforth called the “Ownership Inventory.”

2.1 Background

In education research, a concept inventory (CI) is a test, usually composed of multiple-choice questions, about a narrow domain where the questions and distractors are drawn from common misconceptions about the domain [Hestenes et al. 1992]. There is no singular method for devising a CI [Lindell et al. 2007], but the general idea is to articulate a range of important concepts in the target domain, and then to elicit misconceptions from learners about those concepts. The CI helps evaluate curricula for whether they effectively address such misconceptions.

CIs are an increasingly popular tool for CS education researchers. The last decade has seen a Cambrian explosion of CIs for CS topics such as CS1 [Kaczmarczyk et al. 2010], CS2 [Wittie et al. 2017], algorithms [Farghally et al. 2017], recursion [Hamouda et al. 2017], data structures [Porter et al. 2019], digital logic [Herman et al. 2010], operating systems [Webb and Taylor 2014], and cybersecurity [Poulsen et al. 2021]. The development of these CIs has proved useful in demonstrating both the existence and frequency of misconceptions. For example, the Java-focused CI of Kaczmarczyk et al. [2010] revealed that many students who completed a CS1 course ended up with a “dearth of even basic conception of an Object.” As another example, Farghally et al. [2017] developed an algorithms CI based on 17 misconceptions predicted by educators. After administering the CI to two iterations of the algorithms course at their university, they found that 7 misconceptions were



(a) StackOverflow question #32338659 that asks about moving out of a shared reference (a form of illegal borrow promotion).

```

1  /// Gets the string out of an option if
2  /// exists, returning a default otherwise
3  fn get_or_default(arg: &Option<String>)
4  -> String
5  {
6      if arg.is_none() {
7          return String::new();
8      }
9      let s = arg.unwrap();
10     s.clone()
11 }

```

(b) The program we adapted from the question for the ownership inventory. The type `Box<i32>` has been simplified to `String`, the function has been given a meaningful name, and a doc-string has been added.

Fig. 1. An example of how we created snippets for the Ownership Inventory.

held by at least 1/3 of students. This kind of data helps guide instructors in determining which misconceptions are most important to address in curricular development.

2.2 Development

We set out to design the Ownership Inventory for two reasons. First, the misconceptions we would observe in creating the Inventory would inform our later pedagogy. Second, we could use the Inventory to evaluate the efficacy of an intervention. If our pedagogy is better than before, then it should cause learners to score higher on the Inventory.

To construct the Inventory, we designed open-ended questions about ownership in situations that frequently stymie Rust learners. As we will describe in Section 2.3, we invited Rust learners to answer these questions, and then qualitatively analyzed their responses to characterize their misconceptions. Finally, we converted each open-ended question to multiple-choice by turning common misconceptions into distractors.

This method presents a chicken-and-egg problem: how do we know which situations are hard for Rust learners until we study them? So we turned to the world’s largest repository of programmer struggles: StackOverflow. We searched for the most common questions asked about Rust on StackOverflow that pertain to ownership. Specifically, we queried the top 50 “Most Frequent” questions with the `[rust]` tag and manually filtered the list to 27 questions involving ownership. We iteratively categorized each question, and identified four main categories of ownership problems:

- **Dangling pointers:** references to stack-allocated values that escape their scope.
- **Overlapping borrows:** mutating data that is aliased by another reference.
- **Illegal borrow promotion:** writing to read-only data or moving borrowed data.
- **Lifetime parameters:** taking multiple references as input and returning references as output, but failing to specify the relationship between these references’ lifetimes.

For each category, we selected a few representative StackOverflow questions and cleaned up the snippet in question. For example, Figure 1 shows a StackOverflow question in the “illegal borrow promotion” category and the corresponding clean program. This process created eight total programs. The full set of programs is provided in Appendix A.1.

- (1) The following Rust function is rejected by the Rust compiler. What error message would you expect from the compiler?
(You do not need to exactly reproduce the wording — the question is about whether you generally understand how Rust would justify rejecting this function.)
- (2) Assume that the compiler did NOT reject this function.
 - (a) What is a program that calls this function which would violate memory safety or cause a data race?
(If no such program exists, then leave this field blank and explain your reasoning below. If you are uncertain of a particular Rust syntax, you may use pseudocode notation.)
 - (b) In a few sentences, explain why you believe your program will violate memory safety or cause a data race, or why it is impossible to write such a program.
- (3) (a) How can this function be changed to pass the compiler while (1) preserving its intent and (2) minimally impacting runtime performance?
(You may use the standard library documentation and Rust compiler for this task. There is no right answer — use your judgment. You can change any aspect of the function, including the type signature.)
 - (b) In a few sentences, explain why your fix satisfies the criteria above.

Fig. 2. The template for open-ended Ownership Inventory questions about a given program that is rejected by the compiler. The small parenthetical text is provided as additional context.

We designed a single set of template questions that apply to each program. The template represents each stage of reasoning involved in fixing an ownership error. Figure 2 shows the exact wording of the questions. The phrasing of the questions is open-ended to elicit misconceptions without biasing respondents towards preconceived incorrect answers.

2.3 Methodology

After developing the open-ended Ownership Inventory, we next administered the Inventory to elicit misconceptions that Rust learners have about ownership.

2.3.1 Participants. We recruited $N = 36$ participants for the study. We found Rust learners by embedding an advertisement for the study within the online version of TRPL. Participants were required to be 18 years or older, and they were required to have completed reading TRPL before participating. Participants were compensated \$20. Participants on average had 1.7 ($\sigma = 2.2$) prior years of experience with either C or C++.

2.3.2 Materials. We created a web interface that presents participants with a program and prompts for open-ended responses to each question in Figure 2. The interface uses the Monaco code editor running a Rust language server via a WebAssembly build of Rust Analyzer. The in-browser IDE allows participants to get information about the type and functionality of unfamiliar methods.

The materials include a tutorial that guides participants through both the technical details of using the interface, as well as a sample program paired with a sample response to each of the questions. The full source code for the experiment is provided in the artifact.

An important caveat: in our initial materials, participants were instructed to answer questions Q2b and Q3b by writing a code comment in the same editor used for questions Q2a and Q3a, respectively. However, after reviewing data from the first half of the experiment (18 participants), we found that most participants either ignored or forgot this instruction. For the next 18 participants, we modified the website such that Q2b/Q3b had separate text boxes, which succeeded in eliciting responses from participants. Consequently, our results contain fewer data points for Q2b/Q3b than the other questions.

Table 1. Percentage of correct responses by program and question for the open-ended Ownership Inventory.

Program	Category	Q1	N	Q2a	N	Q2b	N	Q3a	N	Q3b	N
make_separator	Dangling pointer	77%	15	50%	15	35%	10	80%	15	65%	10
get_or_default	Illegal borrow promotion	80%	15	73%	15	0%	4	80%	15	0%	4
remove_zeros	Overlapping borrows	85%	13	54%	13	25%	4	42%	13	75%	4
reverse	Overlapping borrows	75%	16	22%	16	10%	10	72%	16	45%	10
find_nth	Illegal borrow promotion	88%	17	15%	17	5%	11	12%	17	18%	11
apply_curve	Overlapping borrows	60%	15	30%	15	28%	9	37%	15	33%	9
concat_all	Lifetime parameters	10%	15	3%	15	11%	9	43%	15	11%	9
add_displayable	Lifetime parameters	38%	16	6%	16	5%	10	6%	16	10%	10
Average accuracy:		64%		31%		15%		46%		31%	

2.3.3 Procedure. Participants provided informed consent via the web interface, filled out information about their programming background, and then followed the tutorial. After learning about the style of task, participants were given three randomly selected programs in a random order. Participants had up to 15 minutes to answer all questions for a given program. Participants could complete the experiment at any time, and were not supervised by research staff.

At the end of the experiment, participants were given the option to provide open-ended feedback on their experience in the experiment. During the experiment, we continually monitored feedback for confusions with the materials. We did not ultimately make any changes based directly on participant feedback, which was most commonly of the form “I found some of the questions difficult.”

2.3.4 Analysis. To evaluate the overall accuracy of participants, the first two authors independently coded each response as correct or incorrect. After the first round of coding, the authors resolved major disagreements, then independently re-coded the data. After the second round, the inter-rater reliability was 91% in terms of raw agreement and $\kappa = 0.81$ as measured by Cohen’s κ , which is generally considered “excellent” [Fleiss et al. 2013] or “almost perfect” [Landis and Koch 1977] agreement. We considered this sufficient agreement to proceed with the analysis. For the quantitative results, we report scores as the average of the two raters’ scores on each item.

To characterize the specific misconceptions that led to incorrect answers, we performed a thematic analysis of participant responses. The first author coded each incorrect response for the category of error displayed in the response, such as “changing a type from `&Option<String>` to `Option<String>`.” Error categories were further categorized based on similarities across problems, such as “using `clone` to satisfy the borrow checker.”

2.4 Results

Table 1 shows the percentage of total correct responses per-question. Participants could usually predict why the borrow checker would reject a program (Q1). However, participants could only fix the program in 46% of cases (Q3a), and could only create a counterexample in 31% of cases (Q2a). Their accuracy further drops when asked to justify their answer (Q2b and Q3b). Participants could sometimes create counterexamples and fixes without understanding why their answer is correct.

2.4.1 Misconceptions about undefined behavior. Participants’ reasonable performance on Q1 suggests that Rust learners generally understand the surface-level reason for why a program is rejected. However, participants’ comparatively poor performance on Q2a and Q2b suggests that Rust learners do not understand the deeper reasons that justify the ownership rules. Participants’ incorrect attempts to construct counterexamples reveal a range of misconceptions about undefined behavior.

Participants frequently struggled to construct a correct counterexample to an unsafe function. For example, consider the `make_separator` program, shown on the right, which returns a dangling pointer to the variable `default`. 5/15 responses to Q2a provided a counterexample that called `make_separator("")` but did not use the output of the function call. These participants considered calling this function sufficient to violate memory safety, not realizing the need to actually use the dangling pointer.

```

1 fn make_separator(user_str: &str) -> &str
2 {
3     if user_str == "" {
4         let default = "".repeat(10);
5         &default
6     } else {
7         user_str
8     }
9 }

```

```

1 fn add_displayable<T: Display>(
2     v: &mut Vec<Box<dyn Display>>,
3     t: T
4 ) {
5     v.push(Box::new(t));
6 }

```

As a more complex example, consider the `add_displayable` program (left) which lacks correct lifetime parameters. If the type T contains a reference, then the lifetime of that reference is erased when converting T into a trait object `Box<dyn Display>`. Without specifying how T relates to `dyn Display`, that erasure allows `v` to outlive references in its elements.

No participants managed to write a correct counterexample for this task. The answer closest to correct is shown on the right (4/15 participants gave a comparable answer). This answer is not a counterexample because `add_displayable` moves the input string into the vector, so no data is deallocated upon exiting the nested scope. This snippet becomes a correct counterexample if a reference is added to `v`, e.g. the function call is changed to `add_displayable(&mut v, &some_string)`. Then `v` contains a dangling pointer to `some_string`, and a read of that pointer would be undefined behavior.

```

1 let mut v = Vec::new();
2 {
3     let some_string = String::from("Some string");
4     add_displayable(&mut v, some_string);
5 }
6 dbg!(v.last().unwrap());

```

```

1 fn reverse(v: &mut Vec<i32>) {
2     let n = v.len();
3     for i in 0 .. n / 2 {
4         std::mem::swap(
5             &mut v[i],
6             &mut v[n - i - 1]
7         );
8     }
9 }

```

Participants also struggled to identify when a function is actually safe and no counterexample exists. Consider the `reverse` program (left), a case of overlapping borrows. Rust considers `&mut v[i]` and `&mut v[n - i - 1]` to possibly alias. But $i \neq n - i - 1$ for $i \in [0, n/2)$ so this program is actually safe. Only 3/15 participants identified this fact, and only 1 of those 3 gave a correct justification. Participant performance was comparably poor on Q2a and Q2b for the other two safe programs, `find_nth` (6% / 0%) and `apply_curve` (29% / 25%).

Participants also struggled to identify when a function is actually safe and no counterexample exists. Consider the `reverse` program (left), a case of overlapping borrows. Rust considers `&mut v[i]` and `&mut v[n - i - 1]` to possibly alias. But $i \neq n - i - 1$ for $i \in [0, n/2)$ so this program is actually safe. Only 3/15 participants identified this fact, and only 1 of those 3 gave a correct justification. Participant performance was comparably poor on Q2a and Q2b for the other two safe programs, `find_nth` (6% / 0%) and `apply_curve` (29% / 25%).

2.4.2 Misconceptions about fixing ownership errors. Participants could usually change a broken function to pass the borrow checker, but these fixes were not always correct and idiomatic.

One common strategy we observed is the use of the `.clone()` method. In Rust, cloning avoids aliasing by creating a deep copy of data. However, participants often incorrectly used `clone`. As shown on the right, when fixing the `reverse` program, 2/16 participants avoided overlapping borrows by cloning the input vector `v`, and then swapping between the two vectors. This “fix” only reverses the first half of the input vector.

```

1 fn reverse(v: &mut Vec<i32>) {
2     let n = v.len();
3     let mut v2 = v.clone();
4     for i in 0 .. n / 2 {
5         std::mem::swap(
6             &mut v[i], &mut v2[n - i - 1]);
7     }
8 }

```

```

1 fn add_displayable<T: Display + 'static>(
2     v: &mut Vec<Box<dyn Display>>,
3     t: T
4 );
5
6 fn add_displayable<'a, T: Display + 'a>(
7     v: &mut Vec<Box<dyn Display + 'a>>,
8     t: T
9 );

```

When fixes required editing the type signature of a function, participants often created type signatures that were too restrictive or non-idiomatic. For example, returning to the `add_displayable` program, 12/16 participants fixed the function by adding the `+ 'static` bound to the generic type `T`, as shown on the left on lines 1-4. (Notably, this solution is suggested in the compiler error for the original function.) However, this type signature is unnecessarily restrictive — it prevents `add_displayable` from being

used on any type containing a non-static reference. An idiomatic solution is shown on lines 6-9, where a lifetime parameter `'a` is added to the bounds of both `T` and the trait object `dyn Display` to indicate the aliasing relationship between the two types. Only one participant provided the correct and idiomatic solution.

2.5 Discussion

These results show that participants were quite capable at understanding the surface rules of ownership types. Excluding the two questions about lifetime parameters, participants could correctly predict the compiler’s reason for rejection in 78% of cases. However, the subsequent questions reveal that this understanding is shallow. On average, participants could not construct counterexamples to demonstrate undefined behavior, nor could they effectively fix an ownership error.

Given these results, the key question for pedagogy is: why do Rust learning resources like TRPL lead to these learning outcomes? Learning is a complex process, so it is difficult to point to a specific passage and say, “this is the problem.” But in light of the misconceptions observed during this study, we hypothesized that a major learning challenge is that TRPL does not provide the foundations to reason counterfactually about undefined behavior. Nor does TRPL explain how the borrow checker actually works, especially with respect to soundness vs. completeness.

3 A CONCEPTUAL MODEL FOR OWNERSHIP

To understand undefined behavior and incompleteness, a person needs to understand Rust’s dynamic and static semantics. At least, they need to understand a *way of thinking* about these semantics that is *viable* for tasks like debugging ownership errors. The responses to the open-ended Ownership Inventory showed that our participants had a fragile mental model of Rust’s semantics. Therefore, we designed a new way of thinking (a “conceptual” model) of Rust that is precise enough to explain the relevant aspects of ownership, but approximate enough to avoid unnecessary detail.

To characterize the tension between precise and approximate conceptual models, consider a person who wants to understand integer addition in Rust. That is, they want a conceptual model of the semantics of the statement `let z = x + y` where `x, y : i32`. The true dynamic semantics of Rust’s integer addition include aspects like two’s complement overflow and auto-vectorization due to LLVM’s optimizations. But for the average Rust user, these details are usually irrelevant for correctly using addition in routine programming tasks. A mental model that approximates the semantics as “ $x, y \in \mathbb{Z}$ and $x + y$ is standard integer addition” is a generally viable model.

In this section, we describe our conceptual model of Rust’s semantics designed to give learners a viable understanding of ownership. We provide a model for Rust’s dynamic semantics (Section 3.1) and for Rust’s static semantics (Section 3.2). For each model, we articulate three aspects. First, the *informal model*, an intuitive and visual representation of the model, as it would be presented to a Rust learner. Second, the *formal model*, a precise and logical representation of the model for

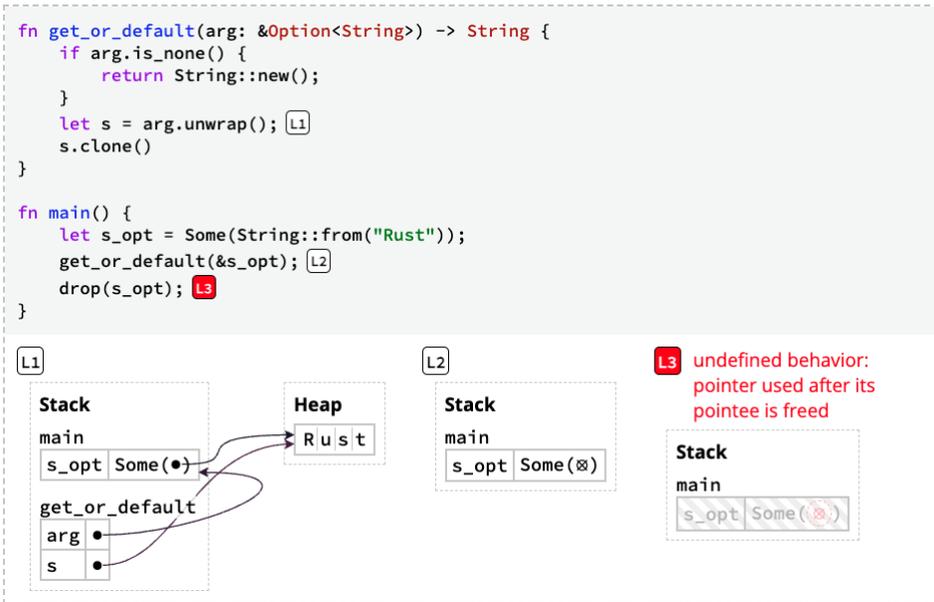


Fig. 3. The dynamic model visualized for the `get_or_default` Ownership Inventory program. The state of memory is shown at three locations L1-L3. At L1, the diagram includes a heap pointer to the string data “Rust” and a stack pointer from the callee to the caller. At L2, the string has been deallocated on behalf of `s` after the call to `get_or_default` ends. At L3, undefined behavior occurs upon a double-free of `s_opt`. Note that the L-*n* labels and the undefined behavior text are part of the visualization, not edited into the figure.

communication and reasoning within this paper. And third, the *implementation*, a description of the tool that executes Rust programs under the model and generates the visualizations.

For a Rust learner, our models must be described in terms of Rust’s surface syntax. Rust learners do not see or think in terms of core calculi or intermediate representations. However, this fact is in tension with our own need to design models that are both (1) simple to formally reason about and (2) feasible to implement. To resolve this tension, the formal model and implementation are described using the “Mid-level Intermediate Representation,” or MIR [2023], a control-flow graph IR within the Rust compiler. The informal model is described using the surface syntax, and the implementation uses source-mapping information to lift the analysis from MIR to the surface.

3.1 Dynamic Model

An essential property of the dynamic model is that it must be able to express the undefined behavior in Rust programs that is normally caught by the borrow checker. Conveniently, a similar need already exists in the Rust ecosystem to find undefined behavior caused by `unsafe` blocks. Miri [2023a] is a MIR interpreter that instruments a program’s runtime to detect undefined behavior like out-of-bounds memory accesses and use-after-free, comparable to Valgrind. Miri provides a *de facto* dynamic semantics to the MIR that can express undefined behavior while avoiding unnecessary details like compiler optimizations.

Therefore, our dynamic model is basically “what Miri does.” The MIR does not have a formal semantics, although multiple projects are currently underway to design one [Niko Matsakis 2023; Ralf Jung 2023]. Rather than provide a complete formal semantics for MIR, we will instead provide a didactic subset of Miri’s semantics that suffices to explain our pedagogy.

PROGRAMS	Variable x	Number n	Function Name f
	Projection $q ::= \varepsilon \mid q.n$		Rvalue $rv ::= c \mid p \mid \ell \mid \langle \bar{p}_i \rangle \mid \mathbf{box} \ p$
	Path $p ::= x \mid p.q \mid *p$		Instruction $I ::= p := rv \mid \mathbf{if} \ p \ \mathbf{then} \ n_1 \ \mathbf{else} \ n_2 \mid f(\bar{p}_i) \mid$
	Constant $c ::= n \mid \mathbf{true} \mid \mathbf{false}$		return $p \mid \mathbf{drop} \ p$
	Ownership Qualifier $\omega ::= \mathbf{shrd} \mid \mathbf{uniq}$		CFG $G ::= \langle \bar{I}_i \rangle$
	Loan $\ell ::= \&\omega \ p$		Function $F ::= \mathbf{fn} \ f \langle \bar{Q}_1 \succ \bar{Q}_2 \rangle (\bar{x}_i : \bar{\tau}_i) \rightarrow \tau_o \{G\}$
TYPES		RUNTIME	Heap Location \mathcal{L}
	Concrete Lifetime r		Segment $s ::= \mathbf{frame} \ (n, x) \mid \mathbf{heap} \ \mathcal{L}$
	Abstract Lifetime ϱ		Address $a ::= s.q$
	Lifetime $l ::= r \mid \varrho$		Value $v ::= c \mid a \mid (v_1, \dots, v_n)$
	Type $\tau ::= \mathbf{u32} \mid \mathbf{bool} \mid (\tau_1, \dots, \tau_n) \mid$		Env $\psi ::= \langle \bar{x}_i \mapsto \bar{v}_i \rangle$
	$\&l \ \omega \ \tau \mid \mathbf{box} \ \tau$		Frame $\sigma ::= (f, n, \psi)$
			Stack $S ::= \langle \bar{\sigma}_i \rangle$
			Heap $H ::= \langle \bar{\mathcal{L}}_i \mapsto \bar{v}_i \rangle$

Fig. 4. The syntax and runtime structure of MIR Rust programs.

3.1.1 Informal Model. A Rust program acts upon memory organized into a stack of function frames and a heap of long-lived data. Each stack frame maps syntactically-scoped variables to values, which are primitives (ints, bools, etc.), composites (structs, enums), or pointers. A path describes a particular value in memory, and pointers are essentially “paths as values” (as opposed to numeric addresses that can be arithmetically manipulated).

Figure 3 shows an example of how we visualize the dynamic model. The state of memory is visualized at multiple points throughout the program. The diagram is fairly similar to prior work in program state visualization [Sorva et al. 2013], so we will not belabor its design. In brief: one key aspect is that states are unrolled over time. Unrolling lets a person more easily compare changes between states, and it allows a person to see all the information in the diagram without interaction. This contrasts with tools like Python Tutor [Guo 2013], which shows one state at a time and requires users to actively scrub a slider between states. Another key aspect is that certain core data-types like boxes, strings, and vectors can be abstracted. For example, the abstracted version of `s_opt` just shows a pointer to heap data, while the expanded version would show a struct containing fields like the string’s length (a button in the diagram permits toggling between these views).

3.1.2 Formal Model. Figure 4 provides the syntax for a subset of the MIR. A control-flow graph G consists of a sequence of indexed instructions I . Instructions are either assignments, conditionals, function calls, returns, or deallocations. The basic primitives are standard (numbers, booleans, tuples, functions), but the interesting operations are those that involve memory.

Memory is arranged into two segments: a stack S of frames σ , and a heap H that maps locations \mathcal{L} to values v . Stack allocations are created as frame-local variables through instructions such as $x := (0, 1)$, and heap allocations are created with boxes such as $y := \mathbf{box} \ 2$. Data in memory are accessed via paths p , such as $x.0$ and $*y$. Finally, references to paths can be created with loans

$\ell ::= \&\omega p$, where ω qualifies the loan as either shared (shrd) or unique (uniq). (We will say more about the type system when discussing the static model.)

Miri's operational semantics for this model is a judgment $(S, H) \rightarrow (S', H')$. The key idea is that Miri's semantics can express undefined behavior such as dangling pointers. Consider this program:

```
x := box 0; drop x; y := *x
```

Miri will give the following execution trace for this program:

$$\begin{aligned} (S = (x := \mathbf{box} 0, \quad \emptyset), H = \emptyset) &\rightarrow \\ (S = (\mathbf{drop} x, \quad \{x \mapsto \mathcal{L}\}), H = \{\mathcal{L} \mapsto 0\}) &\rightarrow \\ (S = (y := *x, \quad \{x \mapsto \mathcal{L}\}), H = \emptyset) &\rightarrow \\ \text{undefined behavior} \end{aligned}$$

Miri's semantics are not expressive enough to represent aspects of the compiler like register allocation and auto-vectorization. But if those optimizations are implemented correctly, a Rust program that does not have undefined behavior under Miri's semantics should also not have undefined behavior under the Rust compiler's actual semantics for its various assembly targets.

3.1.3 Implementation. Given a Rust program, we run Miri to collect an execution trace that describes the state of the stack and heap after each instruction. We use the debug information within Miri (e.g., the types of variables on the stack) to reconstruct the structure of the data in memory. Based on source-mapping information in the Rust compiler, we group contiguous subsequences of the trace into steps that represent source-level expressions and statements. Then this data is passed to a script in a web browser which visualizes the data through a combination of HTML, CSS, and Javascript.

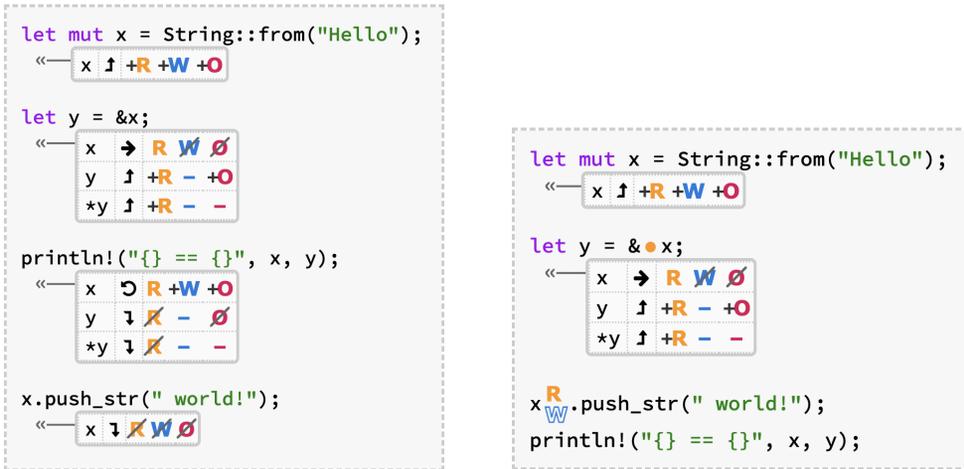
One subtlety is that Miri does not normally execute programs which would be rejected by the borrow checker, such as the one in Figure 3. However, in the current implementations of Rust and Miri, the borrow checker does not substantially influence the translation of a program into an Miri-interpretable CFG. Therefore, we can configure the compiler to simply ignore borrow checking errors, and then ask Miri to attempt to execute programs regardless. Miri is already well-suited to catching undefined behavior (just normally in `unsafe` code), so it suffices for catching undefined behavior in unchecked code in the safe subset of Rust as well.

3.2 Static Model

The dynamic model provides the foundation for understanding how programs can go wrong. The static model should then help learners understand how Rust's borrow checker catches programs that could go wrong (soundness), as well as when safe programs may be rejected (incompleteness).

As a flow-sensitive analysis, borrow checking is more complex than the usual type system encountered by today's programmers. So in designing a conceptual model of the borrow checker, our main goal was to condense the complex intermediate state of the analysis into a comprehensible, visualizable object. The result is the *permissions model* of borrow checking.

3.2.1 Informal Model. At compile-time, Rust's borrow checker uses a system of *permissions* to check whether an operation might cause undefined behavior. The borrow checker tracks whether a path is readable (**R**), writable (**W**), or ownable (**O**). Focusing on Figure 5a, a variable has **RO** permissions by default, and it has the **W** permission if declared with `let mut`. The string `x` therefore has **RWO** permissions. With these permissions, the string can be read like `x.len()`, written like `x.push_str(...)`, and owned like `drop(x)`. The plus sign indicates that the permissions were gained, and the cause of the change is indicated by the icon (up-arrow \uparrow for variable initialization). The



(a) Each table shows the changes in permission state after a given statement. (b) Operations on paths indicate which permissions are expected, and whether those permissions exist.

Fig. 5. Visualizations of the permissions model over two programs that borrow a string.

borrow `let y = &x` removes `WO` permissions from `x` (right-arrow \rightarrow for path borrowed); this action provides `RO` permissions to `y` and the `R` permission to `*y`. Once `y` is no longer used, then its permissions are eliminated (down-arrow \downarrow for the end of a variable's live range) and `x` regains `WO` permissions (cycle-left \curvearrowleft for regaining borrowed permissions).

Turning to Figure 5b, operations expect permissions from paths. The expectations are visually placed between the operator requiring permissions and the path operand. The borrowing operation `&x` expects that `x` has the `R` permission, represented as a yellow circle. The circle is filled in because `x` has the `R` permission. The operation `x.push_str(...)` expects `RW` permissions, represented by the stack of the two letters. (The user of the diagram generator can place annotations on the source code to make an expectation be represented as either a letter or a circle, depending on where the user wants to focus their readers.) However, by swapping the order of the `push_str` and `println` lines, `x` no longer has the `W` permission, visualized as a hollow letter. Therefore the borrow checker will reject this program.

3.2.2 Formal Model. Rust's ownership types are often viewed as enforcing "aliasing XOR mutability", which recent work has started to articulate through the metaphor of permissions [Yanovski et al. 2021]: what can or can't a program do on a particular path at a particular program location? We advance this idea by designing a conceptual model of the borrow checker that reifies permissions into formal objects that can be analyzed, visualized, and taught to Rust learners.

To precisely characterize the permissions model, we first need to provide a model of how Rust's borrow checker actually works. Polonius [Matsakis 2018] is a model of the borrow checker that is maintained by the Rust compiler developers and implemented in Datalog. The former aspect means that Polonius is likely to be a plausible model of Rust's implementation. The latter aspect enables us to easily implement our own alternative model that shares a base of facts about properties like liveness. Sharing facts simplifies both our implementation and our proof of model equivalence.

Polonius model of borrow checking. Figure 6 shows a subset of the inference rules for the Polonius model. Both the Polonius and permission models rely on a shared set of judgments about aspects like liveness and mutation. We do not define these judgments back to their axioms within this

$\frac{\text{POL-BORROW-CONFLICT}}{\ell \text{ live at } I \quad \ell \text{ invalidated at } I} \text{access-error } G$	$\frac{\text{POL-MOVE-CONFLICT}}{p \text{ moved before } I \quad p \text{ read at } I} \text{access-error } G$	
$\frac{\text{POL-READ-INVALID}}{p' \text{ read at } I \quad p \# p'} \&\text{uniq } p \text{ invalidated at } I$	$\frac{\text{POL-WRITE-INVALID}}{p' \text{ written at } I \quad p \# p'} \&\omega p \text{ invalidated at } I$	$\frac{\text{POL-MOVE-INVALID}}{p' \text{ moved at } I \quad p \# p'} \&\omega p \text{ invalidated at } I$

Fig. 6. The core subset of inferences rules for the Polonius model of the borrow checker.

paper — we refer interested readers to the Polonius source code [2023b]. Instead, we just provide enough context to understand the differences and equivalences of the two model.

Given a control-flow graph G , Polonius will reject G (written as “**access-error** G ”, consistent with Polonius’ naming conventions)¹ under one of two conditions (note that $I \in G$ for all rules):

POL-BORROW-CONFLICT: “ ℓ live at I ” means that a loan $\ell ::= \&\omega p$ was created somewhere and is live at an instruction I , i.e., ℓ is used at I or at some instruction reachable from I . “ ℓ invalidated at I ” means I performs an operation that conflicts with ℓ . The **invalidated at** judgment is defined in three ways. **POL-READ-INVALID** states that a unique loan on a path p is invalidated by a read of a conflicting path p' (written $p \# p'$). **POL-WRITE-INVALID** states that any loan on a path p is invalidated by a conflicting write. **POL-MOVE-INVALID** states that any loan on a path p is also invalidated by a conflicting move.

For example, consider the program on the right. Because y is used in $z := *y$, $x := (0, 0)$; then the loan $\&\text{shrd } x$ is live at the instruction $x.0 := 1$. However, $x.0 := 1$ $y := \&\text{shrd } x$; invalidates that loan because $x.0 \# x$, and borrowed data cannot be mutated. $x.0 := 1$; Therefore this program has a loan conflict and is rejected. $z := *y$;

POL-MOVE-CONFLICT: “ p moved before I ” means that the path p has been moved before reaching I . Any use of a movable data type (like a box) that is not through a reference will cause a move. “ p read at I ” means that p is read at I .

For example, consider the program on the right. Because x is moved by $x := \text{box } 0$; $y := x$; and x is read later at $*x$, then this program has a move conflict and is rejected by the borrow checker. $z := *x$

Permissions model of borrow checking. Next, we describe our permissions model and its relationship to the borrow checker. The basic idea is that the five different judgments used within the two Polonius **access-error** rules can be abstracted into two higher-level judgments: “ p needs c at I ” and “ p missing c at I ”, where $c ::= \mathbf{R} \mid \mathbf{W} \mid \mathbf{O}$ is a permission to read, write, or own a path, respectively. A program has a permission violation (written as “**permission-error** G ”) under the **PERM-FAIL** rule, where a permission c on a path p is needed but missing at an instruction I .

Figure 7 shows the rules for the permissions model. The **needs at** rules are straightforward: a path p needs the **R** permission if read, the **W** permission if written, and the **O** permission if moved. The **missing at** rules describe the conditions under which a path lacks a particular permission. The **PERM-MISSING-R** rule states that a place p cannot be read while a loan to a conflicting place p' is live. This is analogous to **POL-READ-INVALID**, i.e. that a read of place p' invalidates any loan of a conflicting place p . The analogy can be formalized as a correctness theorem: we want to show that

¹Type systems are normally formalized as a “positive” judgment, e.g., a program type-checks by constructing a proof of $\Gamma \vdash e : \tau$. But Polonius is formulated as a “negative” judgment: a program does not type-check if a proof of **access-error** G is constructed. For consistency with Polonius, we follow the negative judgment convention in our own model.

$\frac{\text{PERM-FAIL}}{p \text{ needs } c \text{ at } I \quad p \text{ missing } c \text{ at } I}$ <p style="text-align: center;">permission-error G</p>	$\frac{\text{PERM-NEEDS-R}}{p \text{ read at } I}$ <p style="text-align: center;">$p \text{ needs } \mathbf{R} \text{ at } I$</p>	$\frac{\text{PERM-NEEDS-W}}{p \text{ written at } I}$ <p style="text-align: center;">$p \text{ needs } \mathbf{W} \text{ at } I$</p>	$\frac{\text{PERM-NEEDS-O}}{p \text{ moved at } I}$ <p style="text-align: center;">$p \text{ needs } \mathbf{O} \text{ at } I$</p>
$\frac{\text{PERM-MISSING-R}}{\&\text{uniq } p' \text{ live at } I \quad p \# p'}$ <p style="text-align: center;">$p \text{ missing } \mathbf{R} \text{ at } I$</p>	$\frac{\text{PERM-MISSING-WO}}{\&\omega p' \text{ live at } I \quad p \# p'}$ <p style="text-align: center;">$p \text{ missing } c \in \{\mathbf{W}, \mathbf{O}\} \text{ at } I$</p>	$\frac{\text{PERM-MISSING-*}}{p \text{ moved before } I}$ <p style="text-align: center;">$p \text{ missing } c \in \{\mathbf{R}, \mathbf{W}, \mathbf{O}\} \text{ at } I$</p>	

Fig. 7. The inference rules for the permissions model of borrow checking.

permission-error soundly approximates **access-error**, i.e., that $\text{access-error } G \vdash \text{permission-error } G$. A simple rearrangement of terms proves this entailment for the case of **POL-READ-INVALID**:

$$\frac{\&\text{uniq } p' \text{ live at } I \quad \frac{p \text{ read at } I \quad p \# p'}{\&\text{uniq } p' \text{ invalidated at } I}}{\text{access-error } G} \vdash \frac{\frac{p \text{ read at } I}{p \text{ needs } \mathbf{R} \text{ at } I} \quad \frac{\&\text{uniq } p' \text{ live at } I \quad p \# p'}{p \text{ missing } \mathbf{R} \text{ at } I}}{\text{permission-error } G}$$

We can see a similar correspondence for the remaining rules. **PERM-MISSING-WO** states that a path p is missing the write and own permissions while there exists a live loan to a conflicting path p' . Correspondingly, **POL-WRITE-INVALID** states that a loan on p' is invalidated by a write to p . **PERM-MISSING-*** states that a path p is missing all permissions after being moved. Correspondingly, **POL-MOVE-CONFLICT** states that a use of a p cannot occur if p is moved. These correspondences are formalized in each case of the proof in Appendix A.2.

An open problem: lifetime parameter errors. Access errors are the most common kind of borrow checker error encountered by Rust users, accounting for 6/8 programs in the Ownership Inventory. The other kind of error is a *lifetime parameter error*.

Consider the identity function on the right. Because x flows into the return value, the lifetime ϱ_1 must “outlive” the lifetime ϱ_2 . Polonius uses this information to modularly analyze the live ranges of references across function calls, e.g. $b := \text{id}(a)$ should cause a to be live as long as b is live. However, Rust does not infer outlives-constraints for function types, so the user must explicitly specify $\varrho_1 \text{ :> } \varrho_2$ or else this function would be rejected.

```

fn id( $\varrho_1, \varrho_2$ )( $x$ : & $\varrho_1$  uniq u32)
   $\rightarrow$  & $\varrho_2$  uniq u32
{
   $y := x$ ;
  return  $y$ 
}

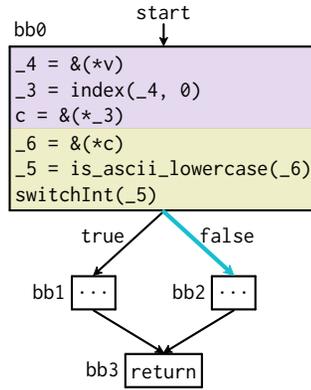
```

Permissions are not a perfect analogy to explain lifetime parameter errors. For example, an outlives-constraint cannot always be blamed on a path, while our model is structured around path-specific permissions. We have experimented with a fourth kind of “flow” permission **F**:

$\frac{\text{POL-LIFETIME-CONFLICT}}{\varrho_1 \text{ outlives } \varrho_2 \text{ at } I \quad \varrho_1 \not\triangleright \varrho_2}$ <p style="text-align: center;">lifetime-param-error G</p>	$\frac{\text{PERM-NEEDS-FLOW}}{\varrho_1 \text{ outlives } \varrho_2 \text{ in } p \text{ at } I}$ <p style="text-align: center;">$p \text{ needs } \mathbf{F} \text{ at } I$</p>	$\frac{\text{PERM-MISSING-FLOW}}{\varrho_1 \text{ outlives } \varrho_2 \text{ in } p \text{ at } I \quad \varrho_1 \not\triangleright \varrho_2}$ <p style="text-align: center;">$p \text{ missing } \mathbf{F} \text{ at } I$</p>
---	--	---

The first rule states that Polonius finds a lifetime parameter error when an instruction I requires ϱ_1 to outlive ϱ_2 , but the function is not annotated with that outlives-constraint. The corresponding permission rules narrow the scope to outlives-constraints that can be blamed on a place p (such as y in the `id` example). The rules state that p needs **F** if such a constraint exists, and that p is missing **F** if the function lacks the necessary outlives annotation. Due to the narrowing of scope, our model does not soundly approximate lifetime parameter errors. In future work, we will investigate either extensions to the permission model or alternative conceptual models that can better represent lifetime parameter errors.

```
fn ascii_capitalize(v: &mut Vec<char>) {
  let c = &v[0];
  if c.R.is_ascii_lowercase() {
    let c_up = c.to_ascii_uppercase();
    v[0] = c_up;
  } else {
    println!("... {v:?}");
  }
}
```



(a) The function `ascii_capitalize` capitalizes the first character in a vector of ASCII characters. It demonstrates flow-sensitive changes in permissions.

(b) A simplified MIR control-flow graph for `ascii_capitalize` highlighting relevant parts of the CFG and how they map to the source-level.

Fig. 8. Example of how the MIR CFG relates to source-code constructs.

3.2.3 Implementation. The structure of our static model visualization parallels the `PERM-FAIL` rule: one component for the `needsat` relation and one component for the missing at relation, corresponding to the letters and tables shown in Figure 5. After generating these relations at the MIR level following Figure 7, the main implementation challenge is to lift the relations to the source-level. We will briefly discuss why this source-mapping is non-trivial and how we approach it, using the `ascii_capitalize` function in Figure 8 as a running example.

Needs-at analysis. One challenge for the `needsat` analysis arises with desugared conversions. For example, consider the method call `c.is_ascii_lowercase()` in Figure 8a. Intuitively, the method’s type signature `&char -> bool` means that the `R` permission should be needed from the method’s receiver, the path `c`. However, as shown in `bb0` of Figure 8b, the MIR-level method is not called directly on `c`, but rather an automatically generated temporary `_6` that is a reborrow of `c`. Therefore, to lift the needs-at analysis for method calls, we have to search backwards from the MIR call instruction to find the first use of the source-visible receiver path.

Missing-at analysis. The missing at relation defines a permission state, or a location-specific set of missing permissions for each path. Rather than visualize the entire permission state at each instruction, we instead visualize the differences in permission state (or “steps”) caused by each instruction, which help readers better see how operations affect permissions. It is straightforward to compute steps between adjacent MIR instructions, but the implementation challenge is to determine which clusters of MIR instructions correspond to meaningful source-level steps.

For example, at the beginning of `ascii_capitalize` in Figure 8-left, `c` is defined as a shared borrow of `v[0]`. As a result, `v` loses `O`, `*v` loses `W`, `c` gains `RO`, and `*c` gains `R`. Using compiler source map information, we compute the contiguous subsequence of MIR instructions that correspond to the source-level statement; in Figure 8 these are highlighted in purple. Then we compute the step as the difference in permission state between the first and last instructions.

4.1 What is Ownership?	4.2 References and Borrowing
4.1.1 Safety is the Absence of Undefined Behavior	4.2.1 References Are Non-Owning Pointers
4.1.2 Ownership is a Discipline for Memory Safety	4.2.2 Dereferencing a Pointer Accesses Its Data
4.1.3 Variables Live in the Stack	4.2.3 Rust Avoids Simultaneous Aliasing and Mutation
4.1.4 Boxes Live in the Heap	4.2.4 References Change Permissions on Paths
4.1.5 Rust Does Not Permit Manual Memory Management	4.2.5 The Borrow Checker Finds Permission Violations
4.1.6 A Box's Owner Manages Deallocation	4.2.6 Mutable References Provide Unique and Non-Owning Access to Data
4.1.7 At Runtime, A Move is Just a Copy	4.2.7 Permissions Are Returned At The End of a Reference's Lifetime
4.1.8 Collections Use Boxes	4.2.8 Data Must Outlive All Of Its References
4.1.9 Variables Cannot Be Used After Being Moved	
4.1.10 Cloning Avoids Moves	
4.3 Fixing Ownership Errors	
4.3.1 Fixing an Unsafe Program: Reference to the Stack	
4.3.2 Fixing an Unsafe Program: Not Enough Permissions	
4.3.3 Fixing an Unsafe Program: Aliasing and Mutating a Data Structure	
4.3.4 Fixing an Unsafe Program: Copying vs. Moving Out of a Collection	
4.3.5 Fixing a Safe Program: Mutating Different Tuple Fields	
4.3.6 Fixing a Safe Program: Mutating Different Array Elements	

Fig. 9. The table of contents for the three sections of our chapter on ownership, designed as a drop-in replacement for TRPL Chapter 4: “Understanding Ownership.”

Steps are not always intra-basic-block subsequences — they can also span across basic blocks, as shown in Figure 8-right when the `else` branch is taken. In this case the sequence doesn't even include any instructions, just the branch between blocks highlighted in cyan. These changes occur due to the flow-sensitive liveness of `c`, which is indicated in our diagram by the down-arrow next to the permission changes. Using these techniques, we lift the formal model to a visual description displayed on the source language.

4 A PEDAGOGY FOR OWNERSHIP

The models in Section 3 provide the conceptual foundation for understanding the aspects of ownership identified in Section 2 such as undefined behavior and incompleteness. Next, we designed a pedagogy that could help Rust learners internalize these models. Rather than designing an entire Rust curriculum from scratch, we instead forked the popular open-source Rust textbook *The Rust Programming Language* (TRPL) [Klabnik and Nichols 2022]. TRPL covers most of the language's core features, and it is the official Rust learning resource endorsed by the Rust project.

We designed a new pedagogy of ownership as a replacement for the existing chapter on ownership in TRPL. The structure of the pedagogy is apparent in the sequence of headings used to organize each section, shown in Figure 9. We start by explaining the core ideas of undefined behavior and memory safety through boxes and moves (the dynamic model). We then introduce references, the borrow checker, and permissions (the static model). Finally, we synthesize these ideas by providing multiple examples of how a Rust programmer can interpret and fix ownership errors, emphasizing

the distinction between soundness and completeness. The full text of the chapter is available online at this link: <https://rust-book.cs.brown.edu/>

4.1 An Illustrative Excerpt

We will provide a sense of the pedagogic principles we used in writing the chapter by walking through the design of the book’s §4.3.4: “Fixing an Unsafe Program: Copying vs. Moving Out of a Collection”. Each excerpt is boxed in gray, and the pedagogic justification is beside the box.

The goal of this section is to help learners understand the distinction between movable types (like `String` or `Vec`) and copyable types (like `i32` or `bool`). The section starts like this:

A common confusion for Rust learners happens when copying data out of a collection, like a vector. For example, here’s a safe program that copies a number out of a vector:

```
let v: Vec<i32> = vec![0, 1, 2];
let n_ref: &i32 = &v[0];
let n: i32 = *n_ref;
```

Each section is anchored around a concrete running example, like copying an element from a vector. Initially, the example is a valid, compiling program. The permission annotations show both the expected permissions on the relevant operation (**R** on `*n_ref`) and how those permissions were gained (via the borrow `&v[0]`).

Then a small change is made to the program such that it no longer compiles (and in this case, is also now unsafe). The change is as small as possible so the reader can maximally transfer their understanding of the previous snippet onto the current one. The contrast in the permission diagrams emphasizes how the change in type has affected the internal state of the borrow checker.

The dereference operation `*n_ref` expects just the **R** permission, which the path `*n_ref` has. But what happens if we change the type of elements in the vector from `i32` to `String`? Then it turns out we no longer have the necessary permissions:

```
let v: Vec<String> =
  vec![String::from("Hello world")];
let s_ref: &String = &v[0];
let s: String = *s_ref;
```

The first program will compile, but the second program will not compile. Rust gives the following error message:

```
error[E0507]: cannot move out of `*s_ref` which is behind
             a shared reference
--> test.rs:4:9
|
4 | let s = *s_ref;
|         ^^^^^^
|         |
|         move occurs because `*s_ref` has type `String`,
|         which does not implement the `Copy` trait
```

The issue is that the vector `v` owns the string “Hello world”. When we dereference `s_ref`, that tries to take ownership of the string from the vector. But references are non-owning pointers – we can’t take ownership through a reference. Therefore Rust complains that we “cannot move out of [...] a shared reference”.

The reader is then given the actual output of the Rust compiler. These error messages will be the actual text encountered by Rust learners in their day-to-day practice, so it is important to explicitly relate the text of the error to the permissions model. (In future work we hope to incorporate the permissions visualizer into the IDE.)

After establishing that a program is rejected by the compiler, we then engage in counterfactual reasoning: what would happen if the program were allowed to execute? In this instance, the code is already executable (i.e., not an abstract function), so we do not need to construct a separate counterexample.

The steps in the diagram are purposefully selected: first, we show a reasonable initial state of memory with live pointers. Then, we show where a pointer becomes dangling. Finally, we show where the dangling pointer is used.

In the remainder of the section (which we elide for brevity), the case study is generalized into a pithy principle: “if a value does not own heap data, then it can be copied without a move”. Then the text explores a space of solutions to avoid the move, such as: only using a reference, deep copying the data, or consuming ownership by removing the element from the collection.

We believe that this style of exposition provides readers with the requisite foundations to reason about errors like “cannot move out of a shared reference” from first principles. Notably, this is the same kind of error that stymied the StackOverflow questioner in Figure 1. In their SO post, that person wrote: “I see there is already a lot of documentation about borrow checker issues, but after reading it, I still can’t figure out the problem.” We set out to determine whether our new pedagogy would suffice to help learners like this one in such cases.

5 EVALUATION

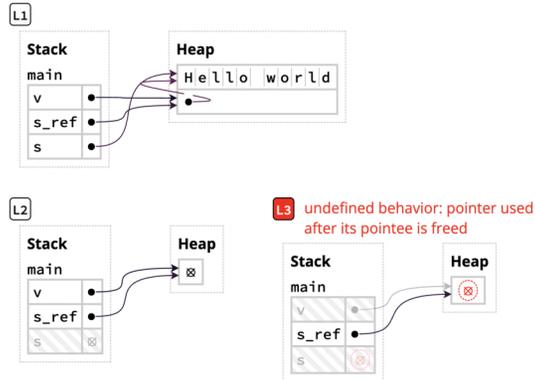
We sought to evaluate our pedagogy on whether it helps learners understand ownership in Rust. This raised two immediate questions. First, how do we find learners to try out our pedagogy? The vast majority of CS education research takes place in a classroom, but we explored an alternative route: free online textbooks. These resources provide access to a larger and more diverse population of learners than CS undergraduates at a single institution. To that end, we set up a publicly-accessible website that hosts our TRPL fork, and it has been visited by tens of thousands of Rust learners to date. This site provides a research platform for analyzing and intervening in the Rust learning process. The intervention described in this paper is the first step in an ongoing experiment to leverage the platform for systematically improving Rust education at scale.

The second key question is: how do we know if learners understand ownership after following our pedagogy? “Understand” is difficult to define — ideally, a longitudinal study might measure understanding as learners’ ability to productively write safe and performant Rust code in their context of use. But for lack of such data, we instead opted for a common substitute: quiz questions.

But why is this unsafe? We can illustrate the problem by simulating the rejected program:

```
let v: Vec<String> =
  vec![String::from("Hello world")];
let s_ref: &String = &v[0];
let s: String = *s_ref; L1

// These drops are normally implicit, but we've added them for clarity.
drop(s); L2
drop(v); L3
```



What happens here is a **double-free**. After executing `let s = *s_ref`, both `v` and `s` think they own “Hello world”. After `s` is dropped, “Hello world” is deallocated. Then `v` is dropped, and undefined behavior happens when the string is freed a second time.

For example, [Ongaro and Ousterhout \[2014\]](#) faced a similar situation: evaluating a novel conceptual model (Raft) for a complex problem (distributed consensus) against a baseline technique (Paxos). They measured understanding by presenting graduate students with a 1-hour lecture on either system followed by a quiz, finding that Raft quiz scores were 8% higher than Paxos quiz scores.

We used a similar approach, but adapted to the setting of an online textbook. Rather than present a monolithic lecture followed by a monolithic quiz, we diffused the quiz questions throughout the ownership chapter and the rest of the book. Furthermore, we distinguished between two kinds of quiz questions designed to answer two research questions:

RQ1. Does our pedagogy help learners understand ownership *at all*?

RQ2. Does our pedagogy help learners understand ownership *better than before*?

For RQ1, we asked participants simple comprehension questions about ownership, presented immediately following the book content that is relevant to a given question. These questions determine whether participants can transfer their ownership knowledge to situations similar to the text. Because the questions make references to the permissions model, we cannot establish a score baseline. Therefore we judge the scores in absolute rather than relative terms.

For RQ2, we gave participants a multiple-choice version of the Ownership Inventory. We inserted these questions later in the book after covering the essential prerequisites for a given program. To compare the baseline TRPL pedagogy against ours, we ran a kind of temporal A/B test. Participants answered Inventory questions after reading the original TRPL content for a few weeks. We then deployed the intervention and continued receiving responses to the Inventory. We quantified the pedagogy's effect based on the resulting change in scores.

5.1 Methodology

In designing our methodology, we traded off between minimizing the amount of infrastructure required, and maximizing the statistical power of our inferences from data. For example, we did not gather any demographic information about participants. We did not want to dissuade privacy-sensitive people from participating in the experiment (reducing the sample size). Moreover, we did not want to implement the infrastructure necessary to securely manage PII at scale. Nevertheless, given that our participants came over from a popular Rust textbook, we believe that people visiting TRPL are reasonably representative of the average Rust learner.

Additionally, the temporal A/B test is logistically simpler but statistically weaker than a traditional population-randomizing A/B test. The traditional setup requires a centralized user database to ensure a reader would not see condition A on their desktop and then accidentally enter condition B on their phone. Instead, our statistical inferences assume that each new participant is sampled from an i.i.d. stream of Rust learners. We discuss this and other trade-offs further in Section 5.4. Our methodology was evaluated by Brown's IRB, which determined that the project did not require institutional review due to the study's purpose and safeguards to ensure anonymity.

5.1.1 Participants. We recruited participants by advertising in the title page of the official web version of TRPL, courtesy of the authors. The advertisement read: "Want a more interactive learning experience? Try out a different version of the Rust Book, featuring: quizzes, highlighting, visualizations, and more." Since this advertisement was put up on November 1, 2022, our TRPL fork has received an average of 450 visitors per day, as measured by unique session IDs stored via cookies.

5.1.2 Materials. We developed 11 comprehension questions to cover the content of the ownership chapter. Appendix A.3 contains the full text of each question, and Table 2 contains a short description of each question. Figure 10a shows one example — to test understanding of permission diagrams, we asked participants to infer the permissions for a path at a given point.

Consider the permissions in the following program:

```

let mut s = String::from("Hello");
let t = &mut s;
/* here */
t.push_str(" world");
println!("{}", s);
    
```

At the point marked `/* here */`, what are the permissions on the path `s`? Select each permission below, or select "no permissions" if the path has no permissions.

R

O

W

No permissions

```

1 fn make_separator(user_str: &str) -> &str {
2   if user_str == "" {
3     let default = ".".repeat(10);
4     &default
5   } else {
6     user_str
7   }
8 }
    
```

Assume that the compiler did NOT reject this function. Which (if any) of the following programs would (1) pass the compiler, and (2) possibly cause undefined behavior if executed? Check each program that satisfies both criteria, OR check "None of these programs" if none are satisfying.

```
let s = make_separator("");
println!("{}", s);
```

```
let s =
make_separator("");
```

```
println!("{}",
make_separator("Hello
world"));
```

None of these programs

(a) A comprehension question ("Analysis state in permissions diagram") that tests whether a person can correctly interpret a permission diagram.

(b) A multiple-choice Ownership Inventory question (Q2a for the `make_separator` program). The distractors are drawn from common incorrect answers given for the open-ended version of the same question.

Fig. 10. Two examples of questions used in the evaluation.

We developed 24 Ownership Inventory questions based on the 8 program in Table 1. For each program, we created a close-ended version of Q1, Q2a, and Q3a in Figure 2 (the justification questions Q2b/Q3b did not translate to the multiple-choice setting). Following the concept inventory methodology, we selected distractors from common misconceptions about the Inventory programs. Figure 10b shows an example of a multiple-choice Ownership Inventory question. The `make_separator` task is an instance of a dangling stack reference. A common incorrect counterexample provided by participants in Section 2 was the snippet `let s = make_separator("")` which creates a dangling pointer, but does not use it. By using that incorrect answer as a distractor, the multiple-choice question is more likely to test for the presence of this misconception.

We spaced out the 24 questions into 4 sets of 6 questions, using the same order as in Table 1. Each pair of tasks was embedded into the end of the appropriate chapter such that the cumulative preceding content would cover all necessary features — Chapter 6, 8, 10, and 17, respectively. Additionally, to ensure participants had access to the relevant documentation on standard library constructs, all the code snippets in the Inventory questions used the same embedded language server technology as described in Section 2.3.2. As an example, the Chapter 6 Inventory questions can be viewed here: <https://rust-book.cs.brown.edu/ch06-04-inventory.html>

The quiz widget permits participants to take a quiz multiple times if they answer any question incorrectly. We eliminate any repeat attempts (as determined by the participants' session IDs) from the dataset and only evaluate based on each participant's first attempt on a given quiz.

Table 2. Readers’ accuracy on simple comprehension questions about the new ownership pedagogy. Accuracies are average correctness with the number of data points in parentheses. Questions are presented in the order encountered by readers. The full text of the questions is in Appendix A.3.

Description	Accuracy	N
Difference of stack and heap	64%	2060
Aliasing in runtime diagram	93%	2060
Moves in runtime diagram	83%	2060
Defined vs. undefined behavior	58%	1958
Compiler error due to move	84%	1958
Dereferencing multiple layers of indirection	47%	1683
How moves affect deallocation	95%	1683
Analysis state in permissions diagram	81%	1281
Why borrows change permissions	59%	1281
How references can cause undefined behavior	44%	1281
Compiler error due to overlapping borrows	87%	1180
Pooled accuracy	72%	

5.1.3 *Procedure.* On December 13, 2022, we deployed the Inventory questions to our book. After fixing bugs for a few weeks, on January 3, 2023 we froze the question content and started gathering pre-intervention data on performance with the baseline TRPL pedagogy. After 45 days, on February 17, 2023 we deployed the initial draft of our new ownership chapter.

For the next few months, we iterated on the text. This process consisted of reviewing quiz data and user feedback (readers had the option to report broken or confusing text, diagrams, and questions). We did not substantively change the pedagogy during this time, but rather fixed or clarified small issues. Here are two examples of the several changes made in this time:

- The original version of our new ownership chapter illustrated borrowing with an example involving vectors. One user alerted us to the fact that TRPL does not explain vectors until a later chapter, making the example difficult to understand. In response, we added context explaining the aspects of vectors that are needed for the example.
- The original version of the runtime diagram did not visually represent a variable being invalidated upon move, since technically move-invalidation is not part of the Rust runtime (a move is just a copy). But readers frequently complained that they expected moves to be visible in the diagram — that is, their mental model did not match our conceptual model. In this case, we ultimately agreed with readers that the point was too technical, and added invalidation to the runtime diagram.

A complete list of the changes can be found in the commit log of the GitHub repository for our book ([cognitive-engineering-lab/rust-book](https://github.com/cognitive-engineering-lab/rust-book)). On June 15, 2023 we froze the text and began gathering post-intervention data. Data collection continued for 45 days until July 30, 2023.

5.2 Results

5.2.1 *RQ1: Does our pedagogy help learners understand ownership at all?* Table 2 shows readers’ accuracies on the comprehension questions. Overall, the pooled accuracy of 72% shows that readers could mostly understand the basic concepts within our pedagogy. Readers were able to successfully interpret both runtime and compile-time diagrams (“Aliasing in runtime diagram” at 93%, “Moves in runtime diagram” at 83%, “Analysis state in permissions diagram” at 81%). Readers could also

Table 3. Effects of the permissions pedagogy for readers’ accuracy on Ownership Inventory questions. Questions are presented in the order encountered by readers. Only effects with $p < 0.15$ are included here, with $p < 0.05$ in bold.

Task	Q.	Before	N	After	N	Effect	d	p
make_separator	Q2	33%	1120	40%	660	+7%	0.14	0.007
make_separator	Q3	57%	1120	62%	660	+5%	0.11	0.024
get_or_default	Q1	56%	1120	66%	660	+10%	0.21	<0.001
get_or_default	Q2	10%	1120	16%	660	+6%	0.19	<0.001
remove_zeros	Q3	35%	629	52%	470	+17%	0.34	<0.001
reverse	Q2	28%	629	40%	470	+13%	0.27	<0.001
reverse	Q3	21%	629	33%	470	+13%	0.29	<0.001
find_nth	Q1	86%	314	90%	374	+4%	0.12	0.099
find_nth	Q2	16%	314	23%	374	+7%	0.18	0.018
find_nth	Q3	27%	316	34%	374	+7%	0.15	0.041
apply_curve	Q2	41%	452	58%	374	+17%	0.34	<0.001
Pooled significant effect:						+10%	0.22	

identify when the compiler was going to reject a program (“Compiler error due to move” at 84%, “Compiler error due to overlapping borrows” at 87%).

However, readers’ mediocre performance on a few of the comprehension questions suggests that their understanding may be somewhat shallow. For example, the ownership chapter provides an example program containing a variable `x : &Box<i32>`, and explains that two dereferences like `**x` are needed to access the inner integer. The comprehension question “Dereferencing multiple layers of indirection” presents a program that constructs an expression of type `Box<&Box<i32>>` (including a runtime diagram), and asks respondents to determine the number of dereference operations needed to access the inner `i32`. Only 47% of respondents correctly answer three, suggesting that readers still leave with a somewhat fragile understanding of an essential concept like pointers.

5.2.2 RQ2: Does our pedagogy help learners understand ownership better than before? We focus on the first 18 Inventory questions, as those questions received enough responses to make statistical inferences. Many readers don’t read to the end of the book – pre-intervention, we only collected 77 responses to the last 6 questions versus 1,120 for the first 6. This dropout rate is comparable to the 90%+ dropout rates seen in MOOCs [Jordan 2015]. Additionally, participants answered the first Inventory question on average 4 days after answering the last comprehension question. Considerable time elapsed between reading the ownership chapter and taking the Inventory.

First, we analyze the overall Inventory scores for the $N = 177$ (pre) / 165 (post) participants who completed the first 18 questions. The average pre-intervention score was 48% ($\sigma = 16\%$). Notably, the average score on the open-ended Inventory questions in Section 2 was 41% (which should be more difficult than equivalent multiple-choice questions), showing that the quantitative results of the formative study reasonably generalized to a larger sample. The average post-intervention score was 57% ($\sigma = 15\%$). Using a two-tailed t -test, the difference is statistically significant ($p < 0.001$). The normalized effect size as measured by Cohen’s d is 0.56. Therefore, the pedagogy had a statistically significant positive effect (+9%) on overall Inventory performance. Additionally, the results confirm that Inventory questions are substantially harder than the comprehension questions.

Second, we analyze the intervention’s effect on each Inventory question individually. The intervention had a statistically significant effect on 10/18 questions. Table 3 shows the size of

these effects, including almost-significant effects. Overall, the pooled significant effect was 10% or $d = 0.22$ (note that the question-level effect size is smaller than the quiz-level effect size due to the higher per-question variance). Between the different types of questions, the intervention primarily affected performance on questions about undefined behavior (Q2) and fixing a type error (Q3) more so than identifying a type error (Q1). For example with `make_separator` Q2, the +7% effect corresponds to an 8% decrease in the incorrect response of “does not have counterexamples”. Conversely, for `reverse` and `apply_curve` Q2, the +13%/+17% effects correspond to participants answering correctly that these functions are safe and do not have counterexamples.

5.3 Discussion

The results on the comprehension questions show that our pedagogy is comprehensible to the average Rust learner. That is notable *per se*, as we have no control over the learner population, many of whom come with no experience in relevant areas like systems or functional programming.

The results on the Inventory questions show that the effect of our pedagogy is statistically significant with an effect size of $d = 0.56$. For reference, according to the meta-analysis of education research by Hattie [2008], the average effect of educational interventions on learning outcomes is $d = 0.40$. Hattie [2008, p. 17] argues that “the effect size of 0.40 sets a level where the effects of innovation enhance achievement in such a way that we can notice real-world differences, and this should be a benchmark of such real-world change.” Therefore, we interpret our results as saying that the new ownership pedagogy is a substantive step in the right direction. But a post-intervention average of 57% clearly demonstrates that we have not “closed the book” on the challenge of teaching ownership types.

5.4 Threats to Validity

Given the large scope of this experiment, we considered several threats to validity in its design.

5.4.1 Construct Validity. This experiment assumes that the Ownership Inventory is a valid instrument to measure a person’s understanding of ownership. To that end, we designed the Inventory such that the situations reflect common ownership problems (by weighting based on StackOverflow), and such that the questions reflect each stage of reasoning about ownership (based on our formative study). However, future work should validate the extent to which performance on the Inventory correlates to performance in solving ownership problems in practice.

5.4.2 Internal Validity. The setting of an online textbook provides the benefit of scale, but it also poses methodological challenges due to lack of controls. One such threat is the uncontrolled quizzing environment. A reader could augment their problem-solving with external aids like a friend, a compiler, a Google search, a large language model, and so on. Participants could also be influenced by learning material outside of the book, such as the official TRPL or Rust-related YouTube videos. To combat this threat, we explicitly instructed participants to not use external resources while solving quiz problems, and the quiz widget takes over the browser tab while taking a quiz. Moreover, we assume that the average participant will be a good actor — our readers are taking these quizzes for their own edification, not to get paid by us or to get a good grade. Gathering enough data should turn bad actors into noise.

Another threat is the uncontrolled assignment to experimental conditions. We chose not to perform a randomized-controlled trial for the reasons discussed in Section 5.1. However, it is possible that temporal correlations in readership could have affected our results. For example, if all the C++ engineers at one company decided to start learning Rust at the same time, then average scores would likely go up in that window of time compared to the average in the limit.

A final threat is teaching to the test. Unlike us, the TRPL authors were not aware of the Ownership Inventory when they wrote the book. At the extreme, if our pedagogy taught the exact answers to Inventory questions, then Inventory scores would not be a useful measure of ownership understanding. At the same time, part of the point of our experiment is exactly to teach to the test! For example, the Inventory is intentionally designed to measure understanding of undefined behavior, and in turn we intentionally designed our new pedagogy to explain undefined behavior. Like any well-meaning educator, we sought a balance. The Inventory materials do not appear anywhere in the revised text. But we do, for instance, walk through an example of how iterator invalidation causes undefined behavior, which is similar to the `remove_zeros` problem.

5.4.3 External Validity. Conditioned on construct and internal validity, our results should reasonably generalize to the larger population of Rust learners. TRPL is the official Rust textbook for the community, so its readers should be a representative cross-section of the broader Rust ecosystem.

6 RELATED WORK

In response to the reports of learners' struggles with ownership [Fulton et al. 2021; The Rust Survey Team 2020; Zeng and Crichton 2019; Zhu et al. 2022], researchers have explored several ways to help Rust users deal with ownership. For instance, Coblenz et al. [2022] showed that garbage collection can help users avoid ownership issues and thereby complete a coding task more quickly.

More directly relevant to our work are *ownership visualizations*. Dominik [2018] and Blaser [2019] developed a tool that visualizes a graph of the outlives-constraints generated by the Rust compiler. They did not evaluate the human factors of their tools, and we believe their visualization would be more appropriate for aiding compiler engineers than learners. Almeida et al. [2022] created RustViz, a visualization format for ownership annotations on a Rust program. In terms of pedagogy, RustViz's premise is that the key challenge with ownership is that "the user must learn to mentally simulate the logic of the borrow checker". Our pedagogy is based more on connecting Rust's static and dynamic semantics, which we show in our formative study is a more serious problem for Rust learners. In terms of implementation, RustViz diagrams are constructed by hand using a DSL, while we automatically generate our diagrams from the compiler. In terms of evaluation, Almeida et al. deployed RustViz in a classroom, finding that students responded to a Likert item that the visualizations were "helpful in terms of improving their understanding of ownership." Our evaluation goes further to quantify the effect of our pedagogy on learning outcomes.

Our runtime diagram is similar to program state visualizers in prior work — see Sorva et al. [2013] for a survey. In particular, our work is similar to C runtime visualizations [Egan and McDonald 2021; Ishizue et al. 2018; Taylor et al. 2023]. In the same vein, our findings about misconceptions of undefined behavior and memory safety are consistent with prior work on teaching C. For instance, Lam et al. [2022] found in a study of undergraduates who had taken a computer systems course that "many students displayed little knowledge or had misunderstandings about memory and memory layout" and would simply say "something bad" happens when unsafe operations occur.

Our work continues a line of CS education research about conceptual models. Bayman and Mayer [1988] first showed that an appropriate conceptual model for BASIC could help students "develop fewer misconceptions [...] and perform better on transfer tests." du Boulay [1986] coined the term "notional machine" for conceptual models specifically of a language's dynamic semantics, which has received renewed focus in recent years [Dickson et al. 2020]. Our work differs from most research on notional machines by focusing equally on a conceptual model of *static* semantics.

Our work also intersects with a line of programming language research on the human factors of type systems and functional languages. Most prior work has focused on algorithms for identifying the root cause of confusing type inference errors [Chital 2001; Wand 1986; Zhang and Myers 2014].

Recent work has broadened scope to develop theories about how programmers read functional programs [Marceau et al. 2011a], leverage the type system during development [Lubin and Chasins 2021], and solve problems with higher-order functions [Rivera and Krishnamurthi 2022].

This paper focuses on ownership types as they are implemented in Rust, but ownership types have taken many forms in prior work [Clarke et al. 2013]. For instance, early systems of ownership focused on ensuring uniqueness of access to data by checking for dominance in the alias graph [Clarke et al. 1998]. Later systems relaxed this constraint by permitting temporary borrowing of data, both mutably [Aldrich et al. 2002; Boyland 2001] and immutably [Dietl et al. 2012; Östlund et al. 2008]. The connection between ownership and permissions has been well-established within formal models such as fractional permissions [Boyland 2003] and F_{own} [Krishnaswami and Aldrich 2005].

7 GENERAL DISCUSSION

Future programming languages will undoubtedly have increasingly complex type systems. Rust is the language *du jour*, so this work focused on ownership types. But the next popular language could bring a renewed emphasis to any existing line of PL research: refinement types, session types, or even theorem proving. Effective transfer of these technologies will require pedagogies that do not expect learners to come equipped with Ph.D.-level knowledge of programming languages, mathematics, and Greek. While our immediate goal in this work was to make ownership types more understandable, our broader goal was to explore the viability of different techniques for improving PL pedagogy. In this section, we will briefly reflect on lessons learned.

First, to develop a metric for understanding of ownership, we created a concept inventory by combining data from StackOverflow with a formative study of Rust learners. StackOverflow works for popular languages like Rust, but is less useful for niche languages. Human factors research on niche languages can instead consider using telemetry from developer interactions as has been explored for Racket [Marceau et al. 2011b] and Coq [Ringer et al. 2020]. The concept inventory is an idea that could easily be reused in the context of other languages. Inventories can serve as communal benchmarks for progress in education research, like how datasets of programs serve as benchmarks for performance in compiler research.

Second, to develop a conceptual model for ownership, we carefully selected a level of abstraction that was concrete enough to explain relevant phenomena like undefined behavior, while abstract enough to avoid unnecessary details. We leveraged the rich prior work on distilling the Rust type system into a small, explainable set of mechanisms, especially the Oxide [Weiss et al. 2021] and Polonius [Matsakis 2018] models. However, PL research usually distills type systems to permit formal reasoning, such as a soundness proof. An open question is how to distill type systems for didactic reasoning, that is, to help learners acquire a conceptual model valid for common tasks. For example, one of our principles was that our model must be encodable in a concise visual representation, which is not a property usually expected of standard PL research. Future work can investigate the properties of semantics that make them more or less explainable.

Finally, to evaluate the efficacy of our pedagogy, we publicly deployed our textbook and compared pre/post-intervention scores on the Ownership Inventory. Collecting telemetry from quizzes in online learning resources is a readily applicable strategy for other contexts. Learners *want* to take quizzes to engage with the content they are reading. Temporal A/B testing offers a lightweight method for evaluating content changes without sophisticated infrastructure. We encourage anyone interested in programming language learning to try out our methodology. To that end, we have open-sourced our frontend quiz plugin and our backend telemetry system at: <https://github.com/cognitive-engineering-lab>

ACKNOWLEDGMENTS

The authors are immensely grateful to Niko Matsakis and Amazon. They provided both the encouragement and the funding to initiate this project, and supplied additional emergency funding when our first grant application fell through because we were studying Rust instead of C++. We thank Carol Nichols for taking a leap of faith in allowing us to advertise in TRPL; this was essential for driving traffic to the experiment. Later parts of this work are partially supported by the US NSF under Grant No. 2319014.

REFERENCES

- Jonathan Aldrich, Valentin Kostadinov, and Craig Chambers. 2002. Alias Annotations for Program Understanding. In *Proceedings of the 17th ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications* (Seattle, Washington, USA) (OOPSLA '02). Association for Computing Machinery, New York, NY, USA, 311–330. <https://doi.org/10.1145/582419.582448>
- Marcelo Almeida, Grant Cole, Ke Du, Gongming Luo, Shulin Pan, Yu Pan, Kai Qiu, Vishnu Reddy, Haochen Zhang, Yingying Zhu, and Cyrus Omar. 2022. RustViz: Interactively Visualizing Ownership and Borrowing. In *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, New York, USA, 1–10. <https://doi.org/10.1109/VL/HCC53370.2022.9833121>
- Piraye Bayman and Richard E. Mayer. 1988. Using conceptual models to teach BASIC computer programming. *Journal of Educational Psychology* 80 (1988), 291–298. <https://doi.org/10.1037/0022-0663.80.3.291> Place: US Publisher: American Psychological Association.
- David Blaser. 2019. *Simple Explanation of Complex Lifetime Errors in Rust*. Bachelor's Thesis. ETH Zürich.
- John Boyland. 2001. Alias burying: Unique variables without destructive reads. *Software: Practice and Experience* 31, 6 (2001), 533–553. <https://doi.org/10.1002/spe.370> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/spe.370>
- John Boyland. 2003. Checking Interference with Fractional Permissions. In *Static Analysis*, Radhia Cousot (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 55–72.
- Olaf Chitil. 2001. Compositional Explanation of Types and Algorithmic Debugging of Type Errors. In *Proceedings of the Sixth ACM SIGPLAN International Conference on Functional Programming* (Florence, Italy) (ICFP '01). Association for Computing Machinery, New York, NY, USA, 193–204. <https://doi.org/10.1145/507635.507659>
- Dave Clarke, Johan Östlund, Ilya Sergey, and Tobias Wrigstad. 2013. *Ownership Types: A Survey*. Springer Berlin Heidelberg, Berlin, Heidelberg, 15–58. https://doi.org/10.1007/978-3-642-36946-9_3
- David G. Clarke, John M. Potter, and James Noble. 1998. Ownership Types for Flexible Alias Protection. In *Proceedings of the 13th ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications* (Vancouver, British Columbia, Canada) (OOPSLA '98). Association for Computing Machinery, New York, NY, USA, 48–64. <https://doi.org/10.1145/286936.286947>
- Michael Coblenz, Michelle L. Mazurek, and Michael Hicks. 2022. Garbage Collection Makes Rust Easier to Use: A Randomized Controlled Trial of the Bronze Garbage Collector. In *Proceedings of the 44th International Conference on Software Engineering* (Pittsburgh, Pennsylvania) (ICSE '22). Association for Computing Machinery, New York, NY, USA, 1021–1032. <https://doi.org/10.1145/3510003.3510107>
- Will Crichton. 2020. The Usability of Ownership. In *Proceedings of the 1st Workshop on Human Aspects of Types and Reasoning Assistants (HATRA)*. arXiv:[arxiv:2011.06171](https://arxiv.org/abs/2011.06171)
- Paul E. Dickson, Neil C. C. Brown, and Brett A. Becker. 2020. Engage Against the Machine: Rise of the Notional Machines as Effective Pedagogical Devices. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education* (Trondheim, Norway) (ITiCSE '20). Association for Computing Machinery, New York, NY, USA, 159–165. <https://doi.org/10.1145/3341525.3387404>
- Werner Dietl, Sophia Drossopoulou, and Peter Müller. 2012. Separating Ownership Topology and Encapsulation with Generic Universe Types. *ACM Trans. Program. Lang. Syst.* 33, 6, Article 20 (jan 2012), 62 pages. <https://doi.org/10.1145/2049706.2049709>
- Dietler Dominik. 2018. *Visualization of Lifetime Constraints in Rust*. Bachelor's Thesis. ETH Zürich.
- Benedict du Boulay. 1986. Some Difficulties of Learning to Program. *Journal of Educational Computing Research* 2, 1 (1986), 57–73. <https://doi.org/10.2190/3LFX-9RRF-67T8-UVK9>
- Tristan Dyer, Tim Nelson, Kathi Fisler, and Shriram Krishnamurthi. 2022. Applying Cognitive Principles to Model-Finding Output: The Positive Value of Negative Information. *Proc. ACM Program. Lang.* 6, OOPSLA1, Article 79 (apr 2022), 29 pages. <https://doi.org/10.1145/3527323>
- Matthew Heinsen Egan and Chris McDonald. 2021. An evaluation of SeeC: a tool designed to assist novice C programmers with program understanding and debugging. *Computer Science Education* 31, 3 (2021), 340–373. <https://doi.org/10.1080/>

08993408.2020.1777034

- Mohammed F. Farghally, Kyu Han Koh, Jeremy V. Ernst, and Clifford A. Shaffer. 2017. Towards a Concept Inventory for Algorithm Analysis Topics. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (Seattle, Washington, USA) (SIGCSE '17). Association for Computing Machinery, New York, NY, USA, 207–212. <https://doi.org/10.1145/3017680.3017756>
- Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. 2013. *Statistical Methods for Rates and Proportions* (3 ed.). Wiley-Interscience, New York.
- Kelsey R. Fulton, Anna Chan, Daniel Votipka, Michael Hicks, and Michelle L. Mazurek. 2021. Benefits and Drawbacks of Adopting a Secure Programming Language: Rust as a Case Study. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. USENIX Association, Berkeley, CA, 597–616. <https://www.usenix.org/conference/soups2021/presentation/fulton>
- Jean-Yves Girard. 1987. Linear logic. *Theoretical Computer Science* 50, 1 (1987), 1–101. [https://doi.org/10.1016/0304-3975\(87\)90045-4](https://doi.org/10.1016/0304-3975(87)90045-4)
- Dan Grossman, Greg Morrisett, Trevor Jim, Michael Hicks, Yanling Wang, and James Cheney. 2002. Region-Based Memory Management in Cyclone. In *Proceedings of the ACM SIGPLAN 2002 Conference on Programming Language Design and Implementation* (Berlin, Germany) (PLDI '02). Association for Computing Machinery, New York, NY, USA, 282–293. <https://doi.org/10.1145/512529.512563>
- Guide to Rustc Development. 2023. The MIR (Mid-level IR). <https://rustc-dev-guide.rust-lang.org/mir/index.html>.
- Philip J. Guo. 2013. Online Python Tutor: Embeddable Web-Based Program Visualization for Cs Education. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education* (Denver, Colorado, USA) (SIGCSE '13). Association for Computing Machinery, New York, NY, USA, 579–584. <https://doi.org/10.1145/2445196.2445368>
- Sally Hamouda, Stephen H. Edwards, Hicham G. Elmongui, Jeremy V. Ernst, and Clifford A. Shaffer. 2017. A basic recursion concept inventory. *Computer Science Education* 27, 2 (2017), 121–148. <https://doi.org/10.1080/08993408.2017.1414728> arXiv:<https://doi.org/10.1080/08993408.2017.1414728>
- John Hattie. 2008. *Visible learning*. Routledge, London, England.
- Geoffrey L. Herman, Michael C. Loui, and Craig Zilles. 2010. Creating the Digital Logic Concept Inventory. In *Proceedings of the 41st ACM Technical Symposium on Computer Science Education* (Milwaukee, Wisconsin, USA) (SIGCSE '10). Association for Computing Machinery, New York, NY, USA, 102–106. <https://doi.org/10.1145/1734263.1734298>
- David Hestenes, Malcolm Wells, and Gregg Swackhamer. 1992. Force concept inventory. *The Physics Teacher* 30, 3 (1992), 141–158. <https://doi.org/10.1119/1.2343497>
- Ryosuke Ishizue, Kazunori Sakamoto, Hironori Washizaki, and Yoshiaki Fukazawa. 2018. PVC: Visualizing C Programs on Web Browsers for Novices. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) (SIGCSE '18). Association for Computing Machinery, New York, NY, USA, 245–250. <https://doi.org/10.1145/3159450.3159566>
- Katy Jordan. 2015. MOOC Completion Rates. <http://www.katyjordan.com/MOOCproject.html>.
- Lisa C. Kaczmarczyk, Elizabeth R. Petrick, J. Philip East, and Geoffrey L. Herman. 2010. Identifying Student Misconceptions of Programming. In *Proceedings of the 41st ACM Technical Symposium on Computer Science Education* (Milwaukee, Wisconsin, USA) (SIGCSE '10). Association for Computing Machinery, New York, NY, USA, 107–111. <https://doi.org/10.1145/1734263.1734299>
- Steve Klabnik and Carol Nichols. 2022. The Rust Programming Language. <https://doc.rust-lang.org/book>.
- Neel Krishnaswami and Jonathan Aldrich. 2005. Permission-Based Ownership: Encapsulating State in Higher-Order Typed Languages. In *Proceedings of the 2005 ACM SIGPLAN Conference on Programming Language Design and Implementation* (Chicago, IL, USA) (PLDI '05). Association for Computing Machinery, New York, NY, USA, 96–106. <https://doi.org/10.1145/1065010.1065023>
- Jessica Lam, Elias Fang, Majed Almansoori, Rahul Chatterjee, and Adalbert Gerald Soosai Raj. 2022. Identifying Gaps in the Secure Programming Knowledge and Skills of Students. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1* (Providence, RI, USA) (SIGCSE 2022). Association for Computing Machinery, New York, NY, USA, 703–709. <https://doi.org/10.1145/3478431.3499391>
- J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. <http://www.jstor.org/stable/2529310>
- Rebecca S. Lindell, Elizabeth Peak, and Thomas M. Foster. 2007. Are They All Created Equal? A Comparison of Different Concept Inventory Development Methodologies. *AIP Conference Proceedings* 883, 1 (2007), 14–17. <https://doi.org/10.1063/1.2508680>
- Justin Lubin and Sarah E. Chasins. 2021. How Statically-Typed Functional Programmers Write Code. *Proc. ACM Program. Lang.* 5, OOPSLA, Article 155 (oct 2021), 30 pages. <https://doi.org/10.1145/3485532>
- Guillaume Marceau, Kathi Fisler, and Shriram Krishnamurthi. 2011a. Do Values Grow on Trees? Expression Integrity in Functional Programming. In *Proceedings of the Seventh International Workshop on Computing Education Research*

- (Providence, Rhode Island, USA) (*ICER '11*). Association for Computing Machinery, New York, NY, USA, 39–44. <https://doi.org/10.1145/2016911.2016921>
- Guillaume Marceau, Kathi Fisler, and Shriram Krishnamurthi. 2011b. Measuring the Effectiveness of Error Messages Designed for Novice Programmers. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education* (Dallas, TX, USA) (*SIGCSE '11*). Association for Computing Machinery, New York, NY, USA, 499–504. <https://doi.org/10.1145/1953163.1953308>
- Niko Matsakis. 2018. An alias-based formulation of the borrow checker. <http://smallcultfollowing.com/babysteps/blog/2018/04/27/an-alias-based-formulation-of-the-borrow-checker>
- Tim McNamara. 2021. *Rust in Action*. Manning Publications, New York, NY.
- MSRC Team. 2019. We need a safer systems programming language. <https://msrc-blog.microsoft.com/2019/07/18/we-need-a-safer-systems-programming-language/>
- Niko Matsakis. 2023. nikomatsakis/a-mir-formality: a model of MIR and the Rust type/trait system. <https://github.com/nikomatsakis/a-mir-formality>.
- Diego Ongaro and John Ousterhout. 2014. In Search of an Understandable Consensus Algorithm. In *2014 USENIX Annual Technical Conference (USENIX ATC 14)*. USENIX Association, Philadelphia, PA, 305–319. <https://www.usenix.org/conference/atc14/technical-sessions/presentation/ongaro>
- Johan Östlund, Tobias Wrigstad, Dave Clarke, and Beatrice Åkerblom. 2008. Ownership, Uniqueness, and Immutability. In *Objects, Components, Models and Patterns*, Richard F. Paige and Bertrand Meyer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 178–197.
- Leo Porter, Daniel Zingaro, Soohyun Nam Liao, Cynthia Taylor, Kevin C. Webb, Cynthia Lee, and Michael Clancy. 2019. BDSI: A Validated Concept Inventory for Basic Data Structures. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (Toronto ON, Canada) (*ICER '19*). Association for Computing Machinery, New York, NY, USA, 111–119. <https://doi.org/10.1145/3291279.3339404>
- Seth Poulsen, Geoffrey L. Herman, Peter A. H. Peterson, Enis Golaszewski, Akshita Gorti, Linda Oliva, Travis Scheponik, and Alan T. Sherman. 2021. Psychometric Evaluation of the Cybersecurity Concept Inventory. *ACM Trans. Comput. Educ.* 22, 1, Article 6 (oct 2021), 18 pages. <https://doi.org/10.1145/3451346>
- Ralf Jung. 2023. RalfJung/minirust: A precise specification for "Rust lite / MIR plus". <https://github.com/RalfJung/minirust>.
- Talia Ringer, Alex Sanchez-Stern, Dan Grossman, and Sorin Lerner. 2020. REPLICA: REPL Instrumentation for Coq Analysis. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs* (New Orleans, LA, USA) (*CPP 2020*). Association for Computing Machinery, New York, NY, USA, 99–113. <https://doi.org/10.1145/3372885.3373823>
- Elijah Rivera and Shriram Krishnamurthi. 2022. Structural versus Pipeline Composition of Higher-Order Functions (Experience Report). *Proc. ACM Program. Lang.* 6, ICFP, Article 102 (aug 2022), 14 pages. <https://doi.org/10.1145/3547633>
- Juha Sorva, Ville Karavirta, and Lauri Malmi. 2013. A Review of Generic Program Visualization Systems for Introductory Programming Education. *ACM Trans. Comput. Educ.* 13, 4, Article 15 (nov 2013), 64 pages. <https://doi.org/10.1145/2490822>
- Andrew Taylor, Jake Renzella, and Alexandra Vassar. 2023. Foundations First: Improving C's Viability in Introductory Programming Courses with the Debugging C Compiler. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (Toronto ON, Canada) (*SIGCSE 2023*). Association for Computing Machinery, New York, NY, USA, 346–352.
- The Rust Programming Language. 2023a. rust-lang/miri: An interpreter for Rust's mid-level intermediate representation. <https://github.com/rust-lang/miri/>.
- The Rust Programming Language. 2023b. rust-lang/polonius: Defines the rust borrow checker. <https://github.com/rust-lang/polonius>.
- The Rust Survey Team. 2020. Rust Survey 2020 Results. <https://blog.rust-lang.org/2020/12/16/rust-survey-2020.html>.
- Jeffrey Vander Stoep. 2022. Memory Safe Languages in Android 13. <https://security.googleblog.com/2022/12/memory-safe-languages-in-android-13.html>.
- Mitchell Wand. 1986. Finding the Source of Type Errors. In *Proceedings of the 13th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages* (St. Petersburg Beach, Florida) (*POPL '86*). Association for Computing Machinery, New York, NY, USA, 38–43. <https://doi.org/10.1145/512644.512648>
- Kevin C. Webb and Cynthia Taylor. 2014. Developing a Pre- and Post-Course Concept Inventory to Gauge Operating Systems Learning. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education* (Atlanta, Georgia, USA) (*SIGCSE '14*). Association for Computing Machinery, New York, NY, USA, 103–108. <https://doi.org/10.1145/2538862.2538886>
- Aaron Weiss, Olek Gierczak, Daniel Patterson, and Amal Ahmed. 2021. Oxide: The Essence of Rust. arXiv:arXiv:1903.00982
- Lea Wittie, Anastasia Kurdia, and Meriel Huggard. 2017. Developing a concept inventory for computer science 2. In *2017 IEEE Frontiers in Education Conference (FIE)*. IEEE, New York, USA, 1–4. <https://doi.org/10.1109/FIE.2017.8190459>
- Herbert Wolverson. 2021. *Hands-on rust*. Pragmatic Programmers, Raleigh, NC.

- Joshua Yanovski, Hoang-Hai Dang, Ralf Jung, and Derek Dreyer. 2021. GhostCell: Separating Permissions from Data in Rust. *Proc. ACM Program. Lang.* 5, ICFP, Article 92 (aug 2021), 30 pages. <https://doi.org/10.1145/3473597>
- Anna Zeng and Will Crichton. 2019. Identifying Barriers to Adoption for Rust through Online Discourse. In *9th Workshop on Evaluation and Usability of Programming Languages and Tools (PLATEAU 2018) (OpenAccess Series in Informatics (OASlcs), Vol. 67)*, Titus Barik, Joshua Sunshine, and Sarah Chasins (Eds.). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 5:1–5:6. <https://doi.org/10.4230/OASlcs.PLATEAU.2018.5>
- Danfeng Zhang and Andrew C. Myers. 2014. Toward General Diagnosis of Static Errors. In *Proceedings of the 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (San Diego, California, USA) (POPL '14)*. Association for Computing Machinery, New York, NY, USA, 569–581. <https://doi.org/10.1145/2535838.2535870>
- Shuofei Zhu, Ziyi Zhang, Boqin Qin, Aiping Xiong, and Linhai Song. 2022. Learning and Programming Challenges of Rust: A Mixed-Methods Study. In *Proceedings of the 44th International Conference on Software Engineering (Pittsburgh, Pennsylvania) (ICSE '22)*. Association for Computing Machinery, New York, NY, USA, 1269–1281. <https://doi.org/10.1145/3510003.3510164>