Faster or Slower? Performance Mystery of Python Idioms Unveiled with Empirical Evidence

APPENDIX A
FORMATIVE STUDY

Table I shows representative discussions about performance impact of nine Python idioms on Stack Overflow. #N represents the number of questions we find for each Python idiom and blue text represents the view times that more than 1k times indicates the questions are popular.

(1) Developers are concerned with the performance of Python idioms. First, developers ask whether idiomatic code is faster or slower than non-idiomatic code. For example, for the star-in-func-call in Table I, developers are interested in whether the star operator affects the performance. Furthermore, developers would like to know how many times idiomatic code may be faster or slower than the non-idiomatic code. For example, for the list-comprehension in Table I, a developer states that the list comprehension is 50% faster than appending to a list with for loop. Finally, developers want to know the reasons why Python idioms causes the performance change. For example, for the for-multiple-targets in Table I, a developer ask why accessing by index slow things down compared to for-multiple-targets.

(2) Many questions lack clear evidence of the performance impact of Python idioms. Only 49.5% questions list code fragments and corresponding execution time to illustrate the performance of Python idioms. For example, for the chain-comparison in Table I, developers write a toy code pair of `x<y<z` and `x<y and y<z` to compare their execution time. Furthermore, for the two questions regarding loop-else and star-in-func-call, they lack code fragments or execution time, but provide only natural language descriptions for the performance of the idioms. For example, for the loop-else in Table I, a developer say "I was introduced to a wonderful idiom in which you can use a for/break/else scheme with an iterator to save both time and lines of code".

(3) Developers are often confused by the controversial descriptions or evidences of the performance of Python idioms. 33 out of 101 question threads present some contradictory performance results. Among the nine examples in Table I, six questions discuss the contradictory performance results, annotated with a *. For example, for the assign-multi-targets in Table I, a developer states that multiple assignment perform at least 30% better than the individual assignment. However, the developer shows that the Python official wiki states that multiple assignment is slower than of individual assignment. Similarly, for the chain-comparison in Table I, a developer finds the chain comparison is slower, but another developer states that the Python official wiki claims that chained comparisons are faster than using "and" operator. As another example, for the truth-value-test in Table I, it is generally understood that this idiom does not impact the performance largely. However, the developer encounter a scenario that the truth-value-test can shave 30% off runtime.

TABLE I: Python Idiom Performance Related SO Questions

Idiom	#N	Question
List-Comprehension*	26	**Question:** How to speed up list comprehension? my understanding is that for-loop is faster than list comprehension. **Comments:** In all cases I've measured the time a list comprehension was always faster than a standard for loop. (2k times)
Set-Comprehension*	12	**Question:** How do python Set Comprehensions work? I tried timeit for speed comparisons, there is quite some difference. **Answer:** List/Dict/Set comprehensions tend to be faster than anything else. (2k times)
Dict-Comprehension*	15	**Question:** Why is this loop faster than a dictionary comprehension? **Comment:** ... I do this with a dictionary with 1000 random keys and values, the dictcomp is marginally slightly faster. (9k times)
Chain-Comparison*	6	**Question:** Is "x < y < z" faster than "x < y and y < z"? In this page, we know that chained comparisons are faster than using the "and" operator. However, I got a different result... It seems that `x < y` and `y < z` is faster than `x < y < z`. (11k times)
Truth-Value-Test*	17	**Question:** bool value of a list in Python. **Answer:** 99.9% of the time, performance doesn't matter as suggested Keith. I only mention this because I once had a scenario, using implicit truthiness testing shaved 30% off the runtime. (31k times)
Loop-Else	7	**Question:** Pythonic ways to use 'else' in a for loop. **Answer:** I was introduced to a wonderful idiom in which you can use a for/break/else scheme with an iterator to save both time and LOC. (1k times)
Assign-Multi-Targets*	8	**Question:** Python assigning two variables on one line I've been looking to squeeze a little more performance out of my code; While browsing this Python wiki page, I found this claim: Multiple assignment is slower than individual assignment. I repeated several times, but the multiple assignment snippet performed at least 30% better than the individual assignment. (3k times)
Star-in-Fun-Call	5	**Question:** What does the star mean in a function call? Does it affect performance at all? Is it fast or slow? (245k times)
For-Mul-Targets	5	**Question:** How come unpacking is faster than accessing by index? (3k times)

Python developers are concerned with the performance of Python idioms. However, their evidences of performance improvement or regression are generally anecdotal based on either toy code or individual project experience. This leads to many controversies about if and when developers should or should not use Python idioms.

APPENDIX B
EMPIRICAL STUDY SETUP

A. Data Collection

Table II summarizes the number of code pairs (non-idiomatic versus idiomatic) in the synthetic dataset and the real-project dataset. For synthetic list/set/dict-comprehension code pairs, 1600 is computed by multiplying 4 numFor, 5 numIf, 5 numIfElse, 2 and 8 values (local or global) of variable scope and `size`. For synthetic chained comparison code, 11968 is computed by multiplying 2992 combinations of 2-5 comparisons from `CompopSet`, 2 and 2 values of variable scope and `isTrue`. For synthetic assign-multiple-targets, when the values are constants, the `isSwap` can only

TABLE II: The statistics of Synthetic and Real-Project Dataset

Idiom	Synthetic	Real-Project
List Comprehension	1600	734
Set Comprehension	1600	282
Dict Comprehension	1600	194
Chain Comparison	11968	2268
Truth Value Test	336	40116
Loop Else	128	198
Assign Multiple Targets	174	10583
Star in Func Call	1920	170
For Multiple Targets	4800	334
Total	24126	54879

be 0, so there are only 174 code pairs instead of 232. For synthetic truth-value-test, 336 is computed by multiplying 3, 2, 14, 2 and 2 values of `TestSet`, `EqSet`, `EmptySet`, scope and `IsTrue`. For synthetic loop-else, the 128 is computed by multiplying the 2, 2, 2, 8 and 2 values of `LoopSet`, `ConditionSet`, variable scope, `size` and `isBreak`. For synthetic star-in-fun-call, the 1920 is computed by multiplying the 30 numSubscript, 2, 2, 2, 2, 2 and 2 values of `hasSubscript`, `hasStep`, `hasLower`, `hasUpper`, variable scope and `isConst`. For synthetic for-multi-targets, the 4800 is computed by multiplying 30 numSubscript, 5 numTarget, 2, 2, 8 values of `hasStarred`, scope and `size`. For real-project code, the number of code pairs are reported by the refactoring tool (accompanied by at least one test case to execute the before- and after-refactoring code).

APPENDIX C
EMPIRICAL ANALYSIS

A. RQ2: How well can code features explain the performance differences caused by Python idioms?

Data size: Fig. 1 shows the relationship between `size` and performance speedup for set-comprehension and dict-comprehension. To see the trend clearly, we take the log of data size. For the set-comprehension, the speedup increases fast when the number of elements increases from 1 to 7(e^2). However, the speedup flattens when the number of elements increases over about 2981 (e^8). For the dict-comprehension, the speedup increases fast when the number of elements increases from 1 to 20 (e^3). However, the speedup flattens when the number of elements increases over about 2981 (e^8).

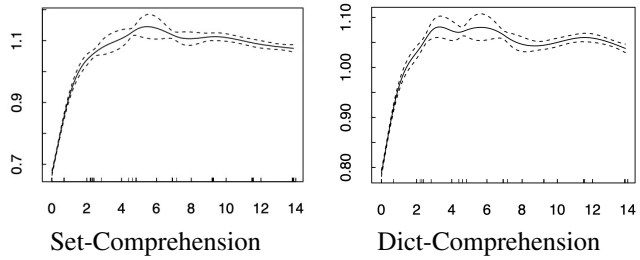

Fig. 1: The Relationship Between `size` and Performance Speedup for Set Comprehension and Dict Comprehension.

B. RQ3: What are the root causes of performance differences caused by Python idioms and what cause the inconsistencies between synthetic and real-project code?

Fig. 2 shows the differences in bytecode instructions between non-idiomatic code and idiomatic code for nine Python idioms.

- **List-comprehension:** From the green box of List-Comprehension in Fig. 2, we could see that the *list-comprehension* needs to execute additional preparation instructions to load listcomp code object and then call it as function before loop iteration. From the blue box of List-Comprehension in Fig. 2, we could see the non-idiomatic code executes the LOAD_* instructions to push data into the stack and then calls the append function to append the element. In contrast, the idiomatic code only executes the LIST_APPEND instruction to append the element.
- **Set-comprehension:** Similar to the *list-comprehension*, *set-comprehension* needs to execute additional preparation instructions to load setcomp code object and then call it as function before loop iteration (green box of Set-Comprehension in Fig. 2). For adding elements, its non-idiomatic code needs to execute the LOAD_* instructions to push data into the stack and then call the add function to add the element, but the idiomatic code only needs to execute the SET_ADD instruction to append the element (blue box of Set-Comprehension in Fig. 2). Besides, non-idiomatic needs to call a function to initialize an empty set, but the *set-comprehension* does not need to do it (red box of Set-Comprehension in Fig. 2).
- **Dict-comprehension:** Similar to the *list-comprehension* and *set-comprehension*, the *dict-comprehension* needs to execute additional preparation instructions to load dictcomp code object and then call it as function before loop iteration (green area of Dict-Comprehension in Fig. 2). However, different from the *list-comprehension* and *set-comprehension*, for appending the element, the non-idiomatic code of the *dict-comprehension* does not need to call a function, but only needs to load the object and store a key-value pair. In contrast, the *dict-comprehension* directly executes MAP_ADD to add the element (blue box of Dict-Comprehension in Fig. 2).
- **Chain-comparison:** The *chain-comparison* replaces one instruction to load the chained comparator o of non-idiomatic code with the instructions to create a reference to the chained comparator o (DUP_TOP) and reference shuffling (ROT_THREE) for the three comparators (n, o, p) (blue box of Chain-Comparison in Fig. 2).
- **Truth-value-test:** From the red box of Truth-Value—Test in Fig. 2, we can see The *truth-value-test* removes instructions of loading the object 0 from `EmptySet` and the comparison operation !=.
- **Loop-else:** The non-idiomatic code determines whether the code executes a break statement by setting a different value to the flagging variable flag. After the for statement is executed, if the flag is True, code executes the body of the if statement. Operations flagging variables and comparison operation is not concise for developers, so the *loop-else* removes the instruc-

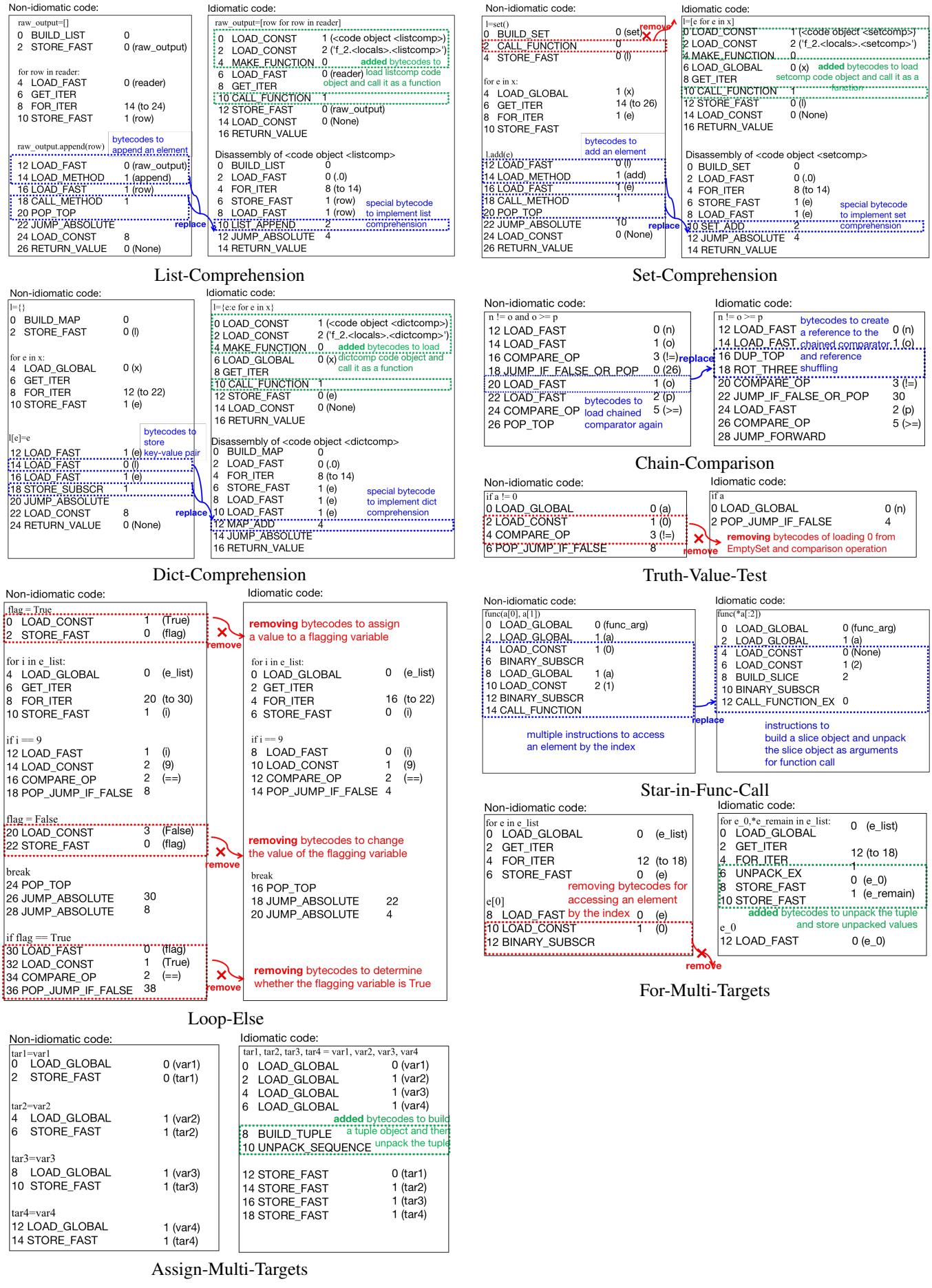

Fig. 2: Bytecode Instructions of Python Idioms (Right) and Corresponding Non-Idiomatic Codes (Left)

tions of setting flagging variables and comparison operation of non-idiomatic code (red box in Loop-Else in Fig. 2).

- **Assign-multi-targets:** Compared to the non-idiomatic code, the idiomatic code additionally executes the BUILD_TUPLE instruction to build a tuple and another UNPACK_SEQUENCE instruction to unpack the sequences to put values onto the stack right-to-left (green box of Assign-Multi-Targets in Fig. 2).
- **Star-in-func-call:** The *star-in-func-call* replaces multiple instructions to load an element by index (BINARY_SUBSCR) with instructions to build a slice object and unpack the slice object into arguments (blue box of Star-in-Func-Call in Fig. 2).
- **For-multi-targets:** In the body of the for statement, the *for-multi-targets* removes instructions to access an element by index (red box of For-Multi-TargetsFig. 2). However, the *for-multi-targets* additionally executes instructions to unpack the object and then store the unpacked values (green box of For-Multi-TargetsFig. 2).