# Project 3 Results

## Method

In our setup, we gave each participant a few repositories in languages they said they were familiar with (by "familiar", we meant having had used the language in a nontrivial project; a rough guideline of 1,000 LOC was given for "nontrivial"). They were allowed to peruse the repositories as much or as little as they wished, and were asked to provide a score between 1 (worst) and 5 (highest) for contributability. Specifically, they were asked to imagine contributing to the repository, and how hard they might find it to get started and get a PR accepted.

We compared these results with the scores generated by our system. These are below.

## Results

Most of our testers went through the following aspects:

- GitHub issues
- PRs
- Checking for readme files and `CONTRIBUTING.md` files
- Checking for comments
- Checking for tests

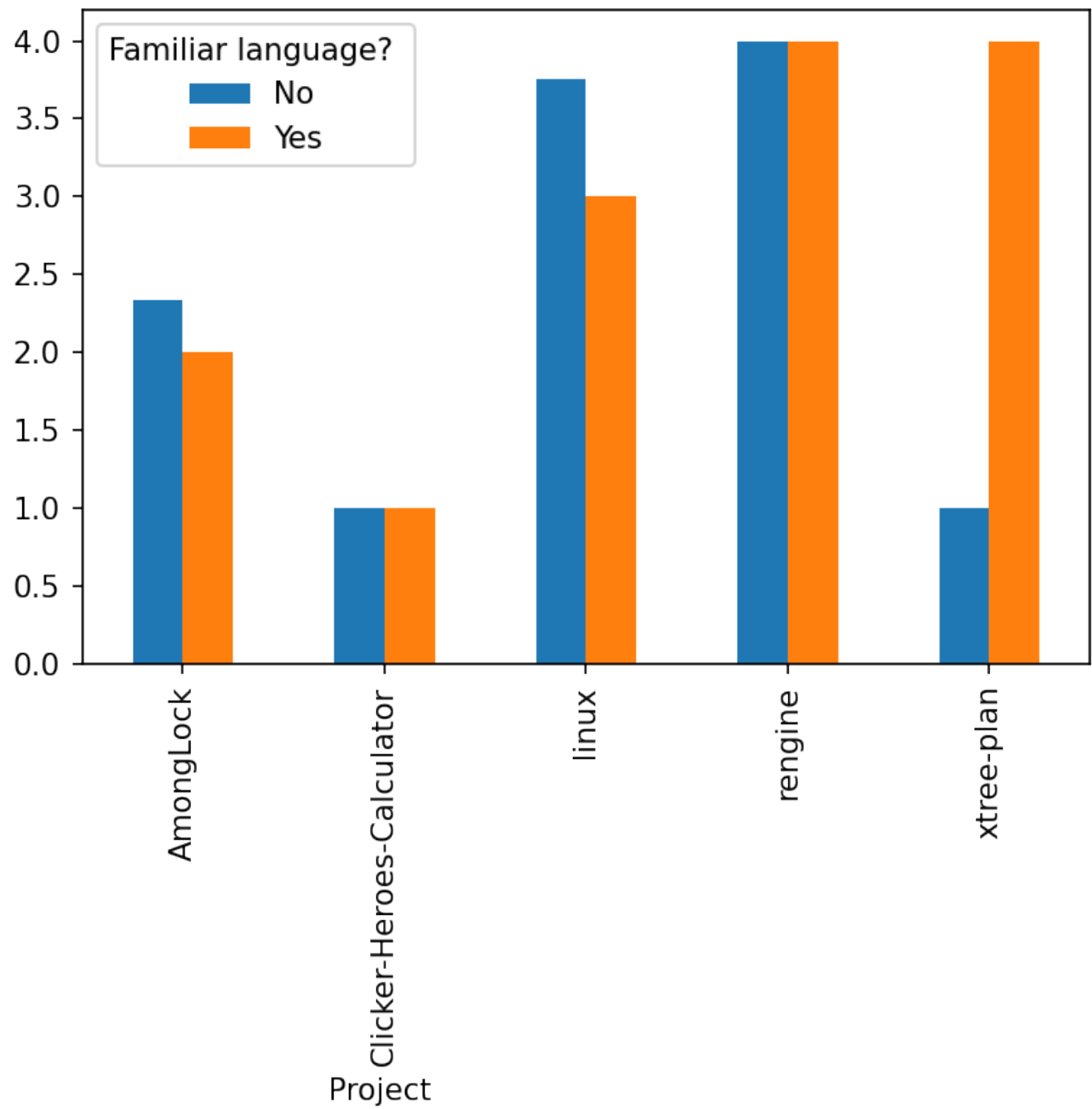The table of ratings is given below.

| Project | Familiar language? | Human rating | System rating |
|---------|--------------------|--------------|---------------|
| https://github.com/mlohbihler/sero-timer | Yes | 1 | 1 |
| https://github.com/elastic/elasticsearch | Yes | 4 | 5 |
| https://github.com/Litteeen/AmongLock | Yes | 1 | 1 |
| https://github.com/torvalds/linux | Yes | 3 | 1 |
| https://github.com/Marak/faker.js | No | 5 | 5 |
| https://github.com/yogeshojha/rengine | No | 4 | 2 |
| https://github.com/mlohbihler/sero-timer | Yes | 2 | 1 |
| https://github.com/elastic/elasticsearch | Yes | 5 | 5 |

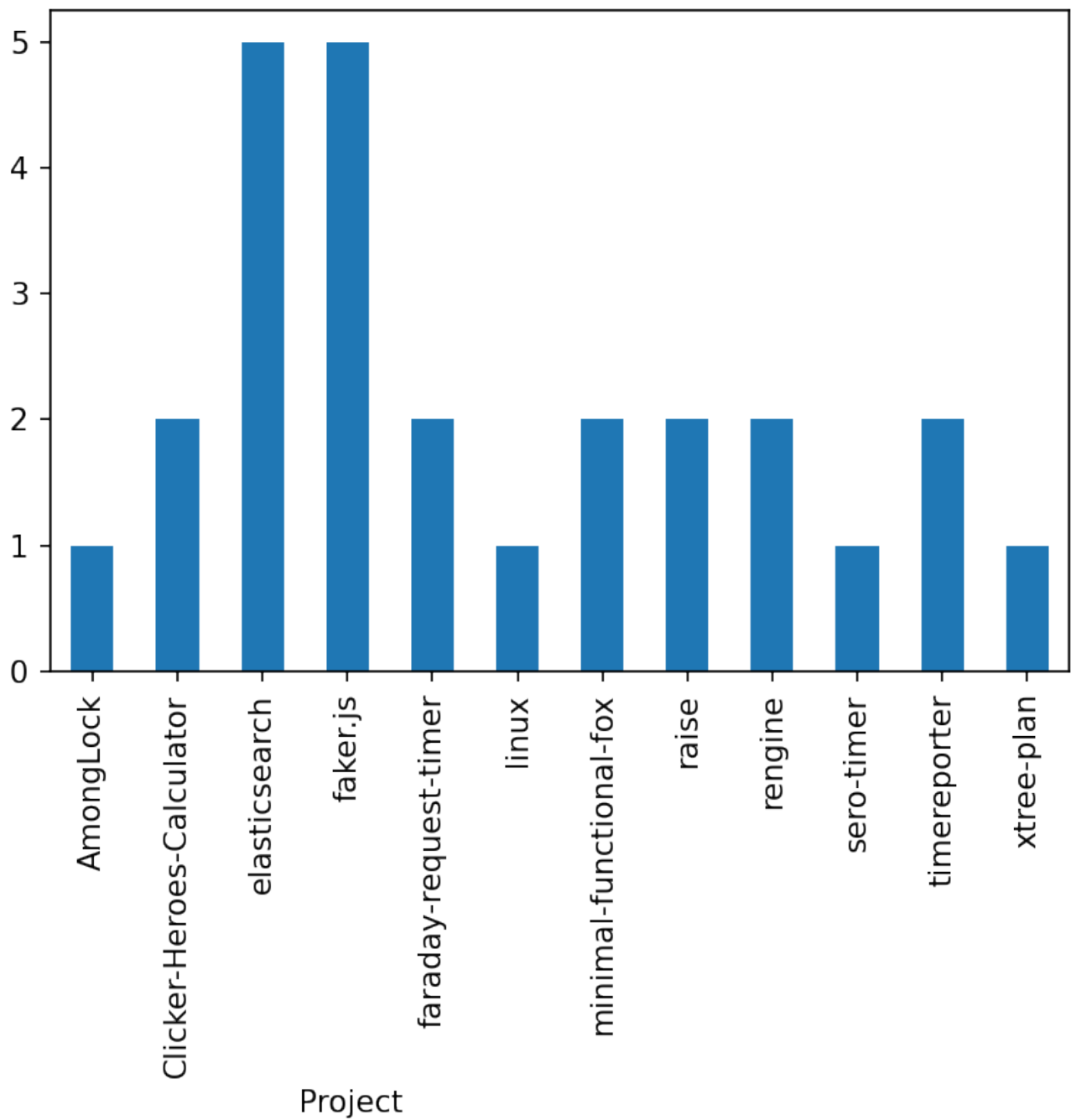| | | | |
|---|---|---|---|
| https://github.com/Litteeen/AmongLock | Yes | 3 | 1 |
| https://github.com/torvalds/linux | Yes | 3 | 1 |
| https://github.com/ai-se/xtree-plan | No | 1 | 1 |
| https://github.com/yrahul3910/raise | No | 2 | 2 |
| https://github.com/mlohbihler/sero-timer | Yes | 1 | 1 |
| https://github.com/elastic/elasticsearch | Yes | 5 | 5 |
| https://github.com/ai-se/xtree-plan | Yes | 4 | 1 |
| https://github.com/Litteeen/AmongLock | No | 2 | 1 |
| https://github.com/torvalds/linux | No | 5 | 1 |
| https://github.com/elastic/elasticsearch | Yes | 4 | 5 |
| https://github.com/mlohbihler/sero-timer | Yes | 1 | 1 |
| https://github.com/Dan2552/timereporter | No | 1 | 2 |
| https://github.com/brandonweiss/faraday-request-timer | No | 3 | 2 |
| https://github.com/Marak/faker.js | No | 5 | 5 |
| https://github.com/VikeStep/Clicker-Heroes-Calculator | No | 1 | 2 |
| https://github.com/yogeshojha/rengine | Yes | 4 | 2 |
| https://github.com/VikeStep/Clicker-Heroes-Calculator | Yes | 1 | 2 |
| https://github.com/Dan2552/timereporter | No | 2 | 2 |
| https://github.com/brandonweiss/faraday-request-timer | No | 4 | 2 |
| https://github.com/mut-ex/minimal-functional-fox | No | 3 | 2 |
| https://github.com/Litteeen/AmongLock | No | 2 | 1 |
| https://github.com/mlohbihler/sero-timer | Yes | 1 | 1 |
| https://github.com/elastic/elasticsearch | Yes | 5 | 5 |
| | | | |

| | | | |
|---|---|---|---|
| https://github.com/Litteeen/AmongLock | No | 3 | 1 |
| https://github.com/torvalds/linux | No | 2.5 | 1 |
| https://github.com/yogeshojha/rengine | No | 4 | 2 |
| https://github.com/Marak/faker.js | No | 4.5 | 5 |

## Analysis

We see fair agreement between users and the system. Although our sample size (34 total) is small, we see that in general, humans tend to agree with each other (with the exception of one project). First, we look at the human ratings:

Next, we present the system-generated contributability scores:

In general, there is agreement between users and the system, except that the human ratings are more liberal. (Aside: some projects are missing in the human chart since, due to the random sets of repositories given to each user, some repositories did not have testers for either the "Familiar with language" or "Not familiar with language" groups).

## Threats to Validity

- **Internal validity:** We believe the humans may have been biased by the course to check only for certain aspects, such as a `CONTRIBUTING.md` file (which is not always indicative of the contributability); further, some people were also biased based on previous notions of a project (for example, some users rated Linux highly seeing only the number of PRs and issues closed and accepted, and the near 1M commits; however, as one of our testers pointed out, it lacks in documentation for a big-picture, even though individual files are well-commented).
- **External validity:** We believe the number of samples are not enough to make scientifically justifiable conclusions; nevertheless, within the constraints of the experimental setup, this is a good-faith effort in an unbiased analysis of the system and how it compares to human rankings.

## Discussion

Our analysis points to a rather successful, if limited, system. In the future, the system should incorporate more language-specific elements; for example, in JavaScript, it is fairly uncommon to include a `CONTRIBUTING.md` file, and in GNU projects, the `LICENSE` file is called `COPYING` instead.