



[Proposals](#)

[Reviews & Meetings](#)

[Awards & Reporting](#)

[Fellowships](#)

[Manage Financials](#)

[Administration](#)

Proposal Review 2 : 2321227

[Back to Proposal](#)

Agency Name: National Science Foundation

Agency Tracking Number: **2321227**

Organization:

NSF Program: CISE Core: Large Projects

PI/PD: Pierce, Benjamin

Application Title: CISE: Large: Property-Based Testing for the People

Rating: Very Good

Review

Summary

In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to intellectual merit.

Summary

In property-based testing (PBT), users specify properties that should be satisfied about the program under development and testing framework generates random inputs and tests to see if the properties are held for all the generated inputs. A counterexample is found if the property fails for some inputs. PBT is perceived as effective on finding software defects but suffers from several usability issues, such as the generated inputs may not satisfy precondition for the property, properties are ineffective, counterexamples are hard to interpret, etc. This proposal aims to improve the usability of PBT by bringing together two researchers from programming languages and human-computer interactions, 4 PhD students, several undergraduate students, and one staff engineer.

***INTELLECTUAL MERIT

Strengths

- +1. Property-based testing has good potential of catching software defects but suffers from several usability issues. The proposed activities will make significant contributions to both theoretical and usability issues.
- +2. The proposed activities comprehensively cover well-known issues with PBT. Some ideas are quite innovative, such as reflective generators.
- +3. The proposal laid out a well-thought plan for carrying out the proposed activities. The activities are appropriately assigned to involved personnel.
- +4. The team is well qualified for conducting the proposed research. PI Pierce has done intensive work in programming languages and property-based testing. PI Head has a lot of experience in human-computer interactions.
- +5. The team has adequate resources to conduct proposed research. The University has space to hold all involved personnel in the same building, which facilitates communications. The project does not need intensive computational power. PhD students and a staff engineer are critical to the success to the project, and there should not be problems recruiting them.

Weaknesses

- 1. The proposed activities are mostly based on the insights from a single study conducted at Jane Street, where OCaml is the prevalent programming language. However, it is unclear if the results from that study are representative. In particular, the usability issues discovered from there seem to well match what the community already knew. This is both a positive and negative sign about the effectiveness of the study. On the positive side, it implies that no other challenges exist for applying PBT. On the negative side, the study did not capture extra challenges of PBT.
- 2. The proposed activities focus on PBT. However, it would be informative if a comparison with other testing or bug detecting methods is presented. In particular, what is benefits/costs of PBT compared to other techniques, what is its capability of bug detection, what is its potential synergy with other techniques, such as symbolic execution for finding inputs to cover certain execution paths, what is its relation with differential testing when you are two versions of a program? How is it compared to formal methods in terms of costs and benefits?
- 3. The proposal does not discuss where PBT is mostly useful. There are some scenarios that PBT is effective, such as generic data structures and algorithms, complex systems that the user is able to capture general input/output relations, etc. However, PBT does not seem to work well for highly-interactive systems such as GUI and stateful, complex systems whose behaviors are hard to capture.
- 4. The proposal does not focus on a specific language or project as its study subject. The proposal mentioned several times about developing support for Python, Java, and Rust. However, it would be better to choose one language and have coherent development for that language. It would also be insightful to choose a few projects as study subjects to evaluate the effectiveness of proposed approaches.

Results of prior NSF support (if applicable):

PI Pierce received a prior grant (NSF [1955565](#)) "Collaborative Research: SHF: Medium: Bringing Python Up to Speed". That funded project, among many other things, developed a property-based testing framework for Python. That project led to 5 publications and a mutation testing tool for Python.

In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to broader impacts.

***BROADER IMPACTS, including enhancing diversity and integrating research and education:

+1. The proposed project has remarkable potential on bringing the research and practices of property-based testing to a new level. As more thorough testing often improves software reliability. In that sense, this project will greatly benefit the society.

+2. This project will create open-source tools, documentations, and educational materials, which will benefit the open-source community, industry, education, and research communities.

+3. The team is well qualified to perform all the activities that have broad impacts. In particular, the PI Pierce has extensive experiences on writing excellent textbooks, papers, and tutorials. It is expected that the team will create easily accessible materials for industry, academia, and students. The recruited staff engineer should be able to develop various prototypes to hand them over to the open source community by the end of the project.

+4. The team has adequate resources to conduct proposed activities on integrating research and education.

Weaknesses:

-1. The team may be further consolidated by adding a research psychologist. The proposed activities involve many surveys and user studies. Having a research psychologist can avoid many pitfalls in designing surveys and studies.

--Adequacy of Data Management Plan:

The data management plan is adequate. For software artifacts, the team plan to release them under an open-source license, host them on github, and will maintain for at least 3 years after the project ends. Technical papers will be available through conference proceedings, which are often open-access nowadays. Educational materials will be posted publicly and raw data will be stored follow IRB rules.

--Adequacy of Post-doctoral Mentoring Plan (if applicable):

N/A

Please evaluate the strengths and weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable

-- Does the proposal identify a computer and information science and engineering grand challenge and an agenda to tackle such a challenge?

Largely yes. On one hand, the challenges about property-based testing mentioned in this proposal

are well known by the research community. After all, QuickCheck has been already for more than two decades. The study with Jane Street does not reveal new challenges. On the other hand, these are real challenges of making property-based testing usable. The proposal provides a coherent agenda to tackle these challenges.

-- Does the proposal explicitly identify the participating CISE core programs it covers and make the case for why the challenge is within the scope of one or more of these participating core programs?

Yes, the paper identifies itself as covering CISE-CCF's SHF and CISE-IIS's HCC programs. The paper also explains why the challenges the proposed project aims to solve are within the scope of these core programs.

-- Does the proposal define the roles of all members of the team and the synergies among them in a Management and Coordination plan?

The proposal clearly specifies the roles of the PI Pierce, the co-PI Head, the four PhD students, and the staff engineer. The two main personnel have strong expertise in programming languages and human-computer interactions, the two main research areas this proposal covers. The proposal discusses the main roles and supporting roles played by involved PIs and PhD students for all proposed activities. The proposal does not explicitly state milestones for the projects but does include metrics to measure the successes of proposed activities. The proposal discusses communication mechanism across the whole group and within the subgroup led by each PI. The proposal also justifies the budget requests when discussing the roles of its members.

BROADENING PARTICIPATION IN COMPUTING:

1. Goal and Context: Does the plan describe a goal and the data from your institution(s) or local community that justifies that goal?

Yes

2. Intended population(s): Does the plan identify the characteristics of participants from an underrepresented group, including school level?

Yes

3. Strategy: Does the plan describe activities that address the stated goal(s) and intended population(s)?

Yes, but the strategy for recruiting underrepresented participants in the REU plan was not clearly articulated.

4. Measurement: Is there a plan to measure the outcome(s) of the activities?

Yes

5. PI Engagement: Is there a clear role for each PI and co-PI? Does the plan describe how the PI is prepared (or will prepare or collaborate) to do the proposed work?

Yes

Summary Statement

This is a well-organized and well-thought proposal that identifies main research challenges and develops a coherent plan for tackling these challenges. If the identified challenges are representative and the proposed activities are conducted successfully, this project will have far-reaching impact on popularizing property-based testing,

which potentially increases the reliability and correctness of software that modern society relies on.

ADDITIONAL SUGGESTIONS:

- 1). In Project description: the argument of having only 2 PIs is not very convincing
- 2). In Section 1, the shorthand BST used without being introduced
- 3). In Section 1, paragraph 3 will be more convincing if some statistics are given about the effectiveness of PBT vs other testing techniques.
- 4). In Section 1, paragraph 5, feels like the sentence begins with "First and foremost..." is incomplete.
- 5). In Section 2, it is a good idea to use insights from a real-world study to motivate the usability issues of PBT. However, it is quite likely that all programmers at Jane Street have strong background on functional programming, where QuickCheck was invented for. It is possible that the usability challenges captured there are not representative. For example, one question is, what is the learning curve of property-based testing vs other testing strategies?
- 6) themes themes -> themes
- 7) will required -> will require
- 8) The sentence begins with "we will state and properties..." is not clear.
- 9) even logs -> event logs?
- 10) "generate generating", not clear?
- 11) In Section 6.1, the goal of real-time visualization may not be easy to achieve since property-based testing usually takes quite some time to finish, unless for rather simple properties
- 12) The sentence "where is no correct program to compare to" is not clear.

Justification of rating, including key strengths and critical weaknesses:

The proposed activities will contribute significantly to the theoretical and experimental developments of property-based testing. It will lay foundations of HCI in the context of property-based testing, which has largely been unexplored so far. The idea of reflective generators are particularly innovative, with many applications both within this project and beyond. The tool developed by this project will allow industrial, academic, and educational communities to benefit from the advances of property-based testing from this project.

The challenges the proposal aims to solve are extracted from the study with Jane Street, which may not very representative as it mainly uses the functional programming language OCaml. The project could be better informed about the challenges of property-based testing by broadening study subjects. The proposal may want to prioritize a programming language that Tyche will be developed for.



About Services

- [Account Management](#)
- [Award Cash Management Service \(ACMS\)](#)
- [Notifications & Requests](#)
- [Project Reports](#)
- [Proposal Status](#)
- [Public Access](#)



NSF Award Highlights

- [Research Spending & Results](#)

Contact

- [Contact Help Desk](#)



News & Discoveries

- [News](#)
- [Discoveries](#)
- [Multimedia Gallery](#)

Funding & Awards

- [Recently Announced Funding Opportunities](#)
- [Upcoming Funding Opportunity Due Dates](#)
- [A-Z Index of Funding Opportunities](#)
- [Find Funding](#)
- [Award Search](#)
- [Proposal & Award Policies & Procedures Guide \(PAPPG\)](#)

Publications & About NSF

- [Publications](#)
- [About the National Science Foundation](#)
- [Careers](#)
- [Staff Directory](#)

[Feedback ▶](#)

 [See all NSF social media ▶](#)