Management and Coordination Plan

Key personnel for this work will include the two co-PIs, four PhD students with varying backgrounds and interests, one staff engineer, and 10 funded undergraduate interns hosted through the summer REPL program (see the BPC supplement). We will include other undergraduate and masters students in research for course credit as opportunities present themselves. Figure 5 sketches our envisaged mapping of tasks onto PhD dissertations, with a rough timeline. Figure 6 gives more detail on supporting effort and supervision.

Since all team members will have offices in the same building on Penn's campus, we do not envisage any significant costs specifically associated with management and coordination.

Project Roles

Principal Investigators The PIs are well positioned to co-lead an effort whose success requires deep contributions from both PL and HCI. Andrew Head recently co-founded a new HCI group at the University of Pennsylvania and specializes in interactive programming environments, while Benjamin Pierce has published widely on PL topics—including, for the past several years, the theory and practice of PBT. Head and Pierce already have an established collaboration, with two jointly supervised PhD students.

The proposed project will be a main focus for both PIs during the whole five years. Two summer months are allocated to each PI for this reason. Formally, PI Head will supervise the two HCI-focused PhD students, while PI Pierce will supervise the two PL-related students and the engineer. The PIs will jointly lead educational efforts at the undergraduate and masters level.

PhD 1, Advisor Head. Project: *Foundations*. The first PhD student will develop foundational knowledge to support the human-centered design and development of PBT tools. This work will begin with defining a preliminary theory of PBT on the basis of our studies to date, with a set of falsifiable hypotheses about how tooling supports PBT. Then, the student will conduct a survey to expand and generalize our understanding of barriers, conduct critical comparative studies of existing PBT solutions, and assist in the conduct of formative observational studies related to tool-building efforts.

PhD 2, Advisor Pierce. Project: *Generation*. The second PhD student will develop the PL theory necessary to enable new classes of powerful PBT generation tools. Central to this PhD will be the advancement of work on reflective generators, a powerful abstraction for producing random data, with potential to support tunability of distributions and sophisticated shrinking techniques. This student will also develop the PBT benchmark suite. Further efforts will involve the assessment of the usability of reflective output with the assistance of HCI PhD student 1, and the improvement of language design for generator automation.

PhD 3, Advisor Pierce. Project: *Specification*. The third PhD student will advance programming languages for specification. Their goal will be to enhance the expressiveness and understandability of specifications. Their work will begin by making specification languages more powerful by supporting formulation of properties across module boundaries, and the more efficient specification of model-based properties. They will then explore how to improve the walk-up-and-use experience of specification languages by building tools for mixed-initiative property specification and understanding the meaning of properties. Their work will conclude with developing technology for converting between property-based specifications into singular test-case specifications to serve as regression tests.

PhD 4, Advisor Head. Project: *Interaction*. The fourth PhD student will design, implement, and evaluate interactive views that provide programmers with powerful controls for influencing testing effectiveness. They will first develop tools that support the live viewing and analysis of complex distributions of generated data. Then, drawing on the reflective generators developed by PhD 2, they will introduce controls for programmers to influence the behavior of generators through direct manipulation. The PhD student will also design tools that support the understanding of individual inputs with novel features for interactive shrinking and debugging. Select views will be brought together and refined to make testing dashboards suitable for monitoring the outcomes of random PBT tests in professional software development settings.

Research Engineer, Supervisor Pierce. Many of the aforementioned projects—particularly those listed under the Diffusion theme (§7)—include technology transfer activities that bring research ideas into practice. Our staff engineer will lead these tech transfer activities, including re-implementing reflective generators

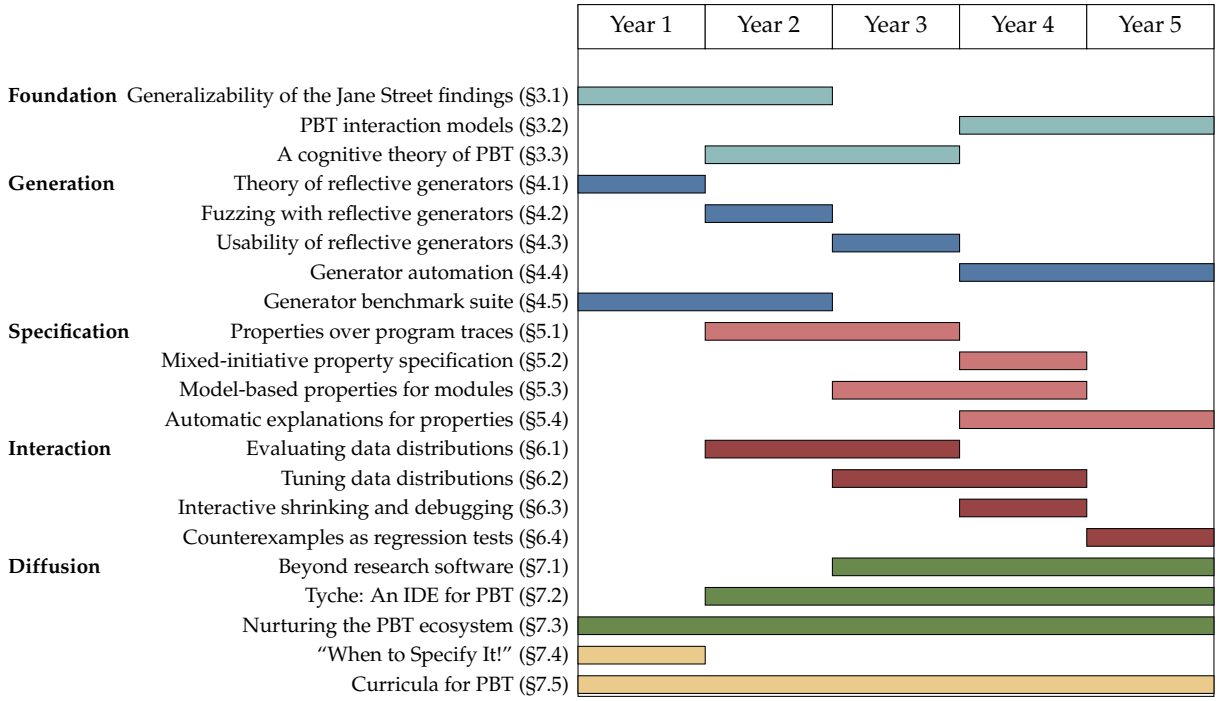

Figure 5: Project timeline, colored by the person responsible for the main effort on each task. = PhD 1, = PhD 2, = PhD 3, = PhD 4, = Research Engineer, = Faculty. We assume that PhD students 1 and 2 and the engineer join the project in its first year and that the other two PhD students join in year 2.

in existing PBT frameworks, improving library designs as per recommendations from PhD 2, maintaining the tools produced by PhD 3, etc. They will also assist with implementation-oriented research activities, in particular TYCHE. We consulted with an industry hiring manager to budget for a competitive salary for this engineer; they have agreed to assist us with best-practices for interviewing. We will look for a developer 2-3 years out of college with experience with open source and multiple programming languages.

Timeline

The work of the project will take place over the course of five years. Figure 5 offers best guesses for natural starting and completion time for the various project tasks, colored according to which PhD student (or, in a few cases, the engineer or faculty) should lead each task. Two PhD students and the engineer will be recruited during the first year, and two more students in the second year. Work at the beginning of the project will focus on understanding opportunities and challenges for PBT through further user studies, on extending our past research work on generators, and on building and starting to disseminate tools based on this work. In the second year, we will spin up efforts in specification tools and user-interface experiments and begin constructing the TYCHE IDE. The latter years of the project will introduce further foundational studies and infrastructure-building projects; they will emphasize more and more the knowledge- and technology-transfer elements of the project. Undergraduate education is a continuous thread.

Coordination Mechanisms

The main coordination mechanism for the project will be weekly group meetings for the entire project team—PIs, PhD students, staff engineer, undergraduates and masters students. Each PI will also meet weekly with the two PhD students working under their primary supervision (Figure 6). Slack and email will be used for daily team communications. Git will be used for software development and writing.

Themes and Tasks	Main effort	Support	Supervisor
Foundation			
Generalizability of the Jane Street findings (§3.1)	PhD 1	REPL 1	Head
PBT interaction models (§3.2)	PhD 1	REPL 2	Head
A cognitive theory of PBT (§3.3)	PhD 1	Everyone	Head
Generation			
Theory of reflective generators (§4.1)	PhD 2	REPL 3	Pierce
Fuzzing with reflective generators (§4.2)	PhD 2	Engineer	Pierce
Usability of reflective generators (§4.3)	PhD 2	PhD 4	Pierce
Generator automation (§4.4)	PhD 2	REPL 4	Pierce
Generator benchmark suite (§4.5)	PhD 2	PhD 1	Pierce
Specification			
Properties over program traces (§5.1)	PhD 3	REPL 5	Pierce
Mixed-initiative property specification (§5.2)	PhD 3	PhD 4	Pierce
Model-based properties for modules (§5.3)	PhD 3	REPL 6	Pierce
Automatic explanations for properties (§5.4)	PhD 3	REPL 7	Pierce
Interaction			
Evaluating data distributions (§6.1)	PhD 4	REPL 8	Head
Tuning data distributions (§6.2)	PhD 4	REPL 9	Head
Interactive shrinking and debugging (§6.3)	PhD 4	PhD 2	Head
Counterexamples as regression tests (§6.4)	PhD 4	REPL 10	Head
Diffusion			
Beyond research software (§7.1)	Engineer		Pierce
Tyche: An IDE for PBT (§7.2)	Engineer	PhD 4	Pierce
Nurturing the PBT ecosystem (§7.3)	Engineer	PhD 1	Pierce
"When to Specify It!" (§7.4)	Faculty	Everyone else	
Curricula for PBT (§7.5)	Faculty	Engineer	

Figure 6: Project themes, tasks, and personnel in detail.

Given the cross-disciplinary nature of the project agenda, it will be important for all team members to have an appreciation of key methods and conceptual tools from both PL and HCI. The four PhD students will be encouraged to take both Penn's core PL theory course (CIS 5000, *Software Foundations*), and our new PhD-level HCI course (CIS 5120, *Introduction to Human-Computer Interaction*).

Four unfunded external collaborators—John Hughes, Hila Peleg, Leonidas Lampropoulos, and Zac Hatfield-Dodds—will assist with various project activities, as described in the Project Description. Letters of collaboration can be found in the Supplemental Documents.

One measure of success in team coordination will be co-ownership of project tasks and co-authorship on publications—concretely, the majority of papers should include 1+ "PL authors" and 1+ "HCI authors."

Success Metrics

One major benefit of HCI methodologies is that they are self-validating—tools are built with users in the loop and evaluated in user studies before being shipped. Each tool-building project will include design justifications and evaluation based on user feedback, and the more theoretical projects will be measured by their ability to support user-facing tools. We also plan to verify that our improvements to the ecosystem for PBT actually drive increased adoption, measured by downloads of our TYCHE IDE as described in §7.2. Success of our educational efforts will also be measured by uptake: We will advertise our PBT course materials in educational circles (directly to colleagues and indirectly through a paper submitted to the CS in Education conference), aiming to see them used by at least 10 universities by the end of the project. Finally, the success of especially the more foundational threads will, as usual, be measured partly in publications. We will target top venues in PL (e.g., ICFP, PLDI, OOPSLA, and POPL), HCI (e.g., CHI, UIST, and VL/HCC), Software Engineering (e.g., ICSE and FSE), and CS Education (Koli Calling).