On the Proof Theory of Property-Based Testing of Coinductive Specifications, or:
PBT to Infinity and beyond

Roberto Blanco¹, Dale Miller², and Alberto Momigliano³

¹ INRIA Paris, France
² INRIA Saclay & LIX, École Polytechnique, France
³ DI, Università degli Studi di Milano, Italy

Reasoning about infinite computations via coinduction and corecursion has an ever increasing relevance in formal methods and, in particular, in the semantics of programming languages, starting from [16]; see also [13] for a compelling example — and, of course, coinduction underlies (the meta-theory of) process calculi. This was acknowledged by researchers in proof assistants, who promptly provided support for coinduction and corecursion from the early 90’s on, see [19, 10] for the beginning of the story concerning the most popular frameworks.

It also became apparent that tools that search for refutations/counter-examples of conjectures prior to attempting a formal proof are invaluable: this is particularly true in PL theory, where proofs tend to be shallow but may have hundreds of cases. One such approach is property-based testing (PBT), which employs automatic test data generation to try and refute executable specifications. Pioneered by QuickCheck for functional programming [7], it has now spread to most major proof assistants [4, 18].

In general, PBT does not extend well to coinductive specifications (an exception being Isabelle’s Nitpick, which is, however, a counter-model generator). A particular challenge, for example, for QuickChick is extending it to work with Coq’s notion of coinductive via guarded recursion (which is generally seen to be an unsatisfactory approach to coinduction). We are not aware of applications of PBT to other form of coinduction, such as co-patterns [1].

While PBT originated in the functional programming community, we have given in a previous paper ([5]) a reconstruction of some of its features (operational semantics, different flavors of generation, shrinking) in purely proof-theoretic terms employing the framework of Foundational Proof Certificates [6]: the latter, in its full generality, defines a range of proof structures used in various theorem provers such as resolution refutations, Herbrand disjuncts, tableaux, etc. In the context of PBT, the proof theory setup is much simpler. Consider an attempt to find counter-examples to a conjecture of the form \( \forall x[(\tau(x) \land P(x)) \supset Q(x)] \) where \( \tau \) is a typing predicate and \( P \) and \( Q \) are two other predicates defined using Horn clause specifications. By negating this conjecture, we attempt to find a (focused) proof of \( \exists x[(\tau(x) \land P(x)) \land \neg Q(x)] \). In the focused proof setting, the positive phase (where test cases are generated) is represented by \( \exists x \) and (\( \tau(x) \land P(x) \)). That phase is followed by the negative phase (where conjectured counter-examples are tested) and is represented by \( \neg Q(x) \). FPCs are simple logic programs that guide the search for potential counter-examples using different generation strategies; they further capture diverse features such as \( \delta \)-debugging, fault isolation, explanation, etc. Such a range of features can be programmed as the clerks and experts predicates that decorate the sequent rules used in a FPC proof checking kernel: the kernel is also able to do a limited amount of proof reconstruction.

As explained in [5], the standard PBT setup needs little more than Horn logic. However, when addressing infinite computations, we need richer specifications. While coinductive logic programming, see [21] and [3] for a much more principled and in depth treatment, may at first
seem to fit the bill, the need to model infinite behavior rather than infinite objects, that is (ir)rational terms on the domain of discourse, has lead us to adopt a much stronger logic (and associated proof theory) with explicit rules for induction and coinduction.

A natural choice for such a logic is the fixed point logic $\mathcal{G}$ [9] and its linear logic cousin $\mu$MALL [2], which are associated to the Abella proof assistant and the Bedwyr model-checker. In fact, the latter has already been used for related aims [11].

To make things more concrete, consider the usual rules for CBV evaluation in the $\lambda$-calculus with constants, but define it coinductively, following see [13]: using Bedwyr’s concrete syntax, this is written as:

```
Define coinductive coeval : tm -> tm -> prop by
  coeval (con C) (con C);
  coeval (fun R) (fun R);
  coeval (app M N) V :=
    exists R W, coeval M (fun R) /\ coeval N W /\ coeval (R W) V.
```

Is evaluation still deterministic? And if not, can we find terms $E$, $V_1$, and $V_2$ such that $\text{coeval } E \; V_1 \; V_2 \; (V_1 = V_2 \rightarrow \text{false})$? Indeed we can, since a divergent term such as $\Omega$ co-evaluates to anything. In fact, co-evaluation is not even type sound in its generality. Our PBT approach aims to find such counter-examples.

It can also be used to separate various notion of equivalences in lambda and process calculi: for example, separating applicative and ground similarity in PCFL [20], or analogous standard results in the $\pi$-calculus. While analogous goals have been achieved for labeled transition systems and for CCS (using, for example, the Concurrency Workbench), it is a remarkable feature of the proof-theoretic account that is easy to generalizes PBT from a system without bindings (say, CCS) to a system with bindings (say, the $\pi$-calculus). Such case is possible since proof theory accommodates the $\lambda$-tree syntax approach to treating bindings [14]: this approach includes the $\nabla$ quantifier [15] that appears in both Abella and Bedwyr.

In our current setup, we attempt to find counter-examples, using Bedwyr to execute both the generation of test cases (controlled by using specific FPCs [5]) and the testing phase. Such an implementation of PBT has the advantages of allowing us to piggyback on Bedwyr’s facilities for efficient proof search via tabling for (co)inductive predicates. There are a couple of treatments of the negation in the testing phase. One approach to eliminating negation from intuitionistic specification can be based on the techniques in [17]. Another approach identifies the proof theory behind model checking as the linear logic $\mu$MALL [12] and in that setting, negations can be eliminated by using De Morgan duality (and inequality).

References


---

1In this case equality is purely syntactical, since by construction terms will be ground when compared, but the logic implements a richer notion [8].


