**Importance of Repeatable Results:**

**Once is not enough**

Natalia Juristo

Imagine you land on an uncharted planet where no human has gone before. For one reason or another, you are unable to leave the spaceship to explore the unknown, and you will only be able to gather information about that planet through the spaceship windows. You look out and see a strange being: an alien of a shape, size and colour that you have never seen or could have imagined before. You are extremely surprised, and you carry on looking at it in order to unlock the mystery and form an idea of what sort of thing you have in your sights. Are you sure that what your eyes are seeing (and your brain is perceiving) through the window really is a true likeness of the outside world? Could it be that the glass in the window is so thick that it is deforming the alien’s profile? And what if the glass is not perfectly transparent and is altering the colour that you can see? Besides, what if the food you had to eat last night had gone off? What if you are seeing things, and there really is nothing out there? What would you do to make sure?

One good strategy for ruling out the possibility that you are seeing things is to ask one of your travelling companions to look out of the window as well and tell you if he or she can see the same thing as you. Your friend Tim looks out of the window next to yours and cannot see anything at all. Is that so? You go round to the window where he is stationed and find that you cannot see the alien from there either, while Tim takes his place at your window. Aha! There is a light of some kind shining into Tim’s window, and it makes the alien invisible from there. Tim confirms that he, too, can see the alien from your window. “It does exist then; it’s out there,” you say. But Tim remembers that he had the same thing as you to eat the night before, and he could be seeing things, too. Or he might perhaps have been led by what you told him when you went to fetch him. “Let’s call Laurie,” he suggests, “We’ll not tell her anything about what we have seen and see what she makes out.” You lead Laurie to the window and ask her what she can see, taking care not to manipulate her in any way or show any signs of nervousness. Laurie looks out of the window and looks back at you dumbfounded. “There is a living being out there, watching us,” Laurie says astonished. It looks like you were right after all; there does appear to be something out there. “Watching us, you say?” you ask Laurie with surprise, “I can’t see any eyes.” You, Laurie and Tim each take it in turns to look out of the window, and you discuss and compare what you see. What Laurie referred to as an eye looks to you and Tim like a blotch on its skin (if you can call the alien’s outer covering skin…)

The fourth crew member has just woken up and comes over to join in the fun. Tom takes his station at another window from where he too can observe Al (by this time you have given the creature a pet name). The information that Tom proffers from the other window is crucial for settling the questions on which you, Laurie and Tim do not agree. From the other window, Tom can see Al from a different angle and make out some particulars that were unclear from your window.

The crew gets quite excited about the progress they are making, and Laurie decides to climb up to the observatory on top of the spaceship to study Al using the onboard telescope. Great! Thanks to this new instrument, we can now distinguish details that were unappreciable by just looking through the windows.

Finally, you write up a report which you send back to Earth informing that there is life on the planet SE and describing your observations.

This imaginary scenario has a lot more in common with empirical research work than one might think at first. We empirical software engineering researchers are every inch the spaceship’s crew on our way to an unexplored planet. We are on a voyage into the unknown: software development. Unfortunately, we cannot *penetrate* the unknown. We cannot travel to the place where the strings of software development are pulled; the place where the development variables causing the external behaviour that we observe in software projects are cooked up. Software *backstage* where hidden variables are ruling development behaviour is equally as impenetrable to our senses as gravity is. We human beings cannot perceive the gravitational forces that are at work behind the behaviours that we observe (for example, when we spill our coffee or when the Earth obediently moves in its orbit around the Sun). Likewise, we cannot directly observe the relationships between the variables causing the behaviours that we observe in software development. The only option open to us is to gather information about software development indirectly through empirical studies.

Any empirical study is a window onto the software development backstage where strings are pulled. But the prospect from that window gives us only one view of the reality that we are scrutinizing. Unfortunately, the window is not open; it contains a piece of glass that has a bearing on what we see. Thus, we have to take the same precautions as our friends on the spaceship that just landed on the unexplored planet SE. First, more than one researcher should make the observations, emulating Natalia when she told Tim and then Laurie to look out of the same window to check that neither her eyes nor her brain were deceiving her. Note that, even though they were all looking out of the same window, Natalia, Laurie and Tim still did not agree on everything about Al, as each of their brains might interpret what they were seeing differently (based on their previous experience, knowledge, preferences, etc).

Besides, one window, no matter how well designed it is, is not enough, because it offers only one view of the development phenomenon under observation. Tom’s observations of Al from another window provided a new perspective. This rounded out all the observations from the first window and cleared up some of the misunderstandings caused by the prospect from the baseline window. Similarly, because there is a relationship between reality and the observation instrument, no one individual experiment (or any other type of empirical study) can yield definitive results, as the observation instrument itself (the study setting) may be affecting some aspects of the results.

In order to make our evidence more reliable, we have to go a step further and observe the phenomenon using another type of instrument. Irrespective of their type or prospect, it may be the build of the windows, for example, the glass of which they are made, that is interfering with what we see. Our intrepid crew use the telescope as an alternative instrument to uncover other particulars. Likewise, different types of empirical studies (experiments, observational studies, historical studies, case studies, surveys…) and empirical paradigms (qualitative and quantitative approaches) provide complementary views of the reality that we are studying, and we can piece together a more accurate picture of the development phenomenon under study by synthesizing their results.

What is the moral of this story? A single result from one empirical study is just a preliminary piece of information; we should even consider it just an anecdote. Our studies have to be repeated, replicated and triangulated in order to form a reliable idea of any SE phenomenon that we investigate. Building a reliable piece of knowledge out of an empirical study requires:

* Looking through the same window: other researchers replicate the study in the same setting and using the same protocol as the baseline study. This strategy confirms that the results are independent of the researcher, site and sample; in other words avoids researcher, sample and site bias.
* Looking through another window: studies (of the same type) should be performed using different protocols, that is, the same type of empirical study but a different setting or protocol: operationalize differently the variables, use other study design, etc. This strategy guarantees that the results are independent of the protocol used; in other words avoids instrument bias.
* Looking through the telescope: alternative studies should be conducted. Observe the same reality through fully different studies provides triangulation of results so new information (or information’s view) is fetched that cannot be gathered from the empirical study (or its type) used as the baseline.

Noticed that none of these steps can be left out. If you skip steps one and two and move to step three, and results you get are different (which there is high probability to happen), you will be unable to trace back the source of variations since there are so many. The new information and the one you had will not fit together like pieces of a puzzle and the bigger picture will emerge. We need to go by baby steps.

To know more:

O. Gómez, N. Juristo, S. Vegas. *Understanding Replication of Experiments in Software Engineering: A Classification*. **Information and Software Technology**. 56, 8. 1033-1048. 2014

N. Juristo, S. Vegas. *The Role of Non-Exact Replications in Software Engineering Experiments*. **Empirical Software Engineering**. 16, 3. 295–324. 2011