If it's not good testing, it's not good regression testing either.

22 March 2011 · exploratory testing, pre-scripted testing, regression testing, testing

Pick a coin from your pocket, and hold it at arms length. Take a good look. Now take another one, of the same denomination and hold it out at arms length as before. Based on your observations alone - can you say they are the identical?

Lets go a step further. If someone had given you one coin to look at, then exchanged it for another, could you have determined whether they are the same or different coins? Maybe, yes? If the differences had been large enough e.g. one coin was heavily tarnished or scratched, then the different coins would be identifiable. Or if you'd been given the opportunity to examine the coin using magnifying equipment, you probably could of found differences.

But lets assume our only test was a standard set of checks i.e.: viewing at arms length and comparing what we see with our notes/records. It's better than nothing, I would see some differences, some might be important ones. For example if my next coin was blank: I might have suspected an issue with my coin supply, and investigated.

What about my next coin… it is blank on one side. Unfortunately it's not the side I check when I hold it at arms length. So as far as my checks are concerned there has been no regression in the quality of the coins being produced by my pocket. So until I go 'live' and try and spend my coins out in the real world of shopkeepers, I'm none the wiser.

Do you see the flaw in our logic here? If we noticed a degradation in coin quality the testing is good. If the testing does not find an issue, it still must be good, because previously those checks found a different issue. Because I was only performing one test or one set of tests I was blind to issues that I can't see with that one test.

If we'd been testing the coins independently, we probably would of been more critical. We might of thought: sure it looks good in the arm length test, what about the weight: maybe thats wrong. We'd try a number of different tests trying to find an issue. We'd ask other people about coins, learn about their two sided nature and perform tests for it.

But as soon as we enter 'regression testing' mode, we often start to disregard this behaviour and start to mindlessly run the same tests. We avoid exploration, sometimes without noticing. Sometimes people actively avoid exploration during regression testing thinking it's inappropriate. This approach would assume that the test you have been running is some kind of super-observer, capable of helping you to see all problems.

If the system has changed significantly, with the addition or removal of complex behaviours, surely the tests might not also need to adapt? The assumption that the same test will somehow catch a change in functionality, reliability etc is based on the premise that our super-test was testing everything -before- and still is. As testers we know it didn’t, doesn’t and never will be that super-test. We need to adapt to each new release in an attempt to find new issues. If our tests aren’t finding an issue, it’s just as possible that the tests are ineffective as it is that the system isn’t defective.