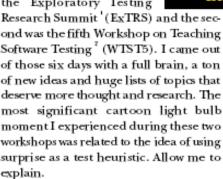# 'Hmm...That's Odd': Embracing Surprise And Curiosity

I recently spent six consecutive days attending two invitation-only workshops that, partly by design and partly by coincidence, focused on the educational theory of teaching software testing and developing, identifying heuristics and training others to use them in their testing. The first was the Exploratory Testing Research Summit[1] (ExTRS) and the second was the fifth Workshop on Teaching Software Testing[2] (WTST5). I came out of those six days with a full brain, a ton of new ideas and huge lists of topics that deserve more thought and research. The most significant cartoon light bulb moment I experienced during these two workshops was related to the idea of using surprise as a test heuristic. Allow me to explain.

Long before I had ever thought about testing heuristics, even before I really had a firm grasp on what I was doing as a performance test analyst, I trusted my instincts to dig deeper whenever something in the results I was analyzing caused surprise or piqued my curiosity, which always seemed to make me mumble, "Hmm...that's odd."

Or maybe it wasn't my instincts at all. Maybe it was the fact that my office buddy, Chris Walters, always seemed to hear me mumble and ask, "What's odd?" Of course, I never really had an answer for him, so I would say, "I dunno, come take a look at this." With a roll of his eyes, he would come to my desk and we'd look at the results together.

Sometimes he would glance at my screen and instantly have an answer for me. Other times we'd export the data to Excel, put the screen on the overhead projector and start charting, graphing and manipulating the data until the results started making sense. Yet other times we'd design and execute new tests using the approach that I'd later come to refer to as conducting performance experiments, to figure out what the "oddness" indicated.

Whether it was my instinct or Chris' sharp ears that caused me to pay attention to my mumbling, over time I realized that mumbling was my single most accurate indicator of performance issues that deserved further attention. Very rarely was the mumble related to a failure of any stated requirement. I remember mumbling about a particularly complex page showing response times faster than some relatively simple pages—it turned out that the complex page wasn't actually returning at all; rather the server was returning a cleverly masked error page that my tool didn't recognize as an error. I also remember mumbling about a site returning faster responses as the load increased—today, I know to check for things like server-side scripts compiling and database caching that make the beginning of my test an exception case rather than a performance anomaly, but at the time I'd never actually given that much thought. If I had not learned to look into those observations that make me mumble, I'd have never found most of the performance issues that, if left unresolved, would have caused monumental problems for my clients when the applications were moved into production.

As it turns out, what I was really doing was using a testing heuristic based on surprise and curiosity. Even when I didn't know what to expect the results to be, some results, or patterns in results, would surprise me. My natural, or as some would say, excessive, curiosity led me to try to figure out why I was surprised by the results. I came to call this my "Hmm...that's odd" heuristic that I use whenever I test. I don't mumble as much anymore, but I do a lot of mental design of future tests whenever a result or observation seems odd or out of place.

As it turns out, I'm not the only person with a "Hmm...that's odd" heuristic. Near the end of WTST5 James Bach described his "Surprise Heuristic" this way:

"I make an *observation* and experience *surprise* associated with a pattern within that observation. The surprise triggers *reflection* about *plausibility*. Often when I am surprised by an observation, it is because the pattern seems implausible relative to my current physical or mental model of the phenomenon. That realization leads to *reflection* about *risk*, so I bring to mind a risk associated with that pattern. If I am able to imagine a risk associated with the pattern, that triggers *reflection* about *magnitude of risk*. If I feel that the risk is important, that triggers *test redesign*."

This description is fantastic in both what it says directly and what it implies. The first thing that it highlights that isn't often discussed when evaluating test results is the idea of plausibility. Too often testers get caught up in the predefined pass/fail criteria of evaluating results and forget to

Scott Barber is the CTO at PerfTestPlus. His specialty is context-driven performance testing and analysis for distributed multi-user systems. Contact him at sbarber @perftestplus.com.

think about the plausibility or reasonableness of the results. We are all used to the specifications we are testing against being incomplete or flawed, especially when it comes to specifications related to system performance, yet we rarely talk about raising flags because results seem unreasonable or implausible. The funny thing is that when you ask a performance tester to tell you about the "best bug" he or she has found recently, you will almost certainly hear about a bug found outside of specification based testing.

The next thing that Bach mentions that is often overlooked in testing is the idea of results that differ from our current model of the application. We all make models of the application we are testing. Sometimes they are formal, documented and go through a review process; sometimes they are subconscious mental models; and many times they are somewhere in between. But whether we realize it or not, we are constantly making and revising our models of the application we are testing. Of course, there is a funny thing about models: By definition, they are all wrong. Since models are simplifications that we use to aid in our understanding and no simplification can predict every possible behavior of what we have modeled, we get used to observations differing from what our model would have predicted. What Bach implies in his description of his surprise heuristic is that we shouldn't dismiss observations that differ from our model so easily. Rather than simply assuming that our models are wrong, we should consider the implications of what we've observed, investigate further, and only then decide whether we have discovered a problem with our model or a problem with the application.

"Models and modeling" is a topic that deserves a lot more discussion, but let's leave that discussion for another day in favor of completing our discussion about surprise. Bach goes on to discuss risk in his surprise heuristic. This is a critical component of all testing. Whether we've been tasked with evaluating compliance with a specification or estimating the end user's eventual happiness with the application, every decision we make about our observations is based on some kind of risk analysis. The challenge of trying to eval-

uate the risk of an observation that causes surprise is that surprise often simply indicates that the results are outside of some conscious or subconscious expectation rather than necessarily indicating a fault with the application. Unfortunately, many testers have been trained to report only observations that contradict stated expectations…which leads us to the final part of Bach's description: test redesign.

What he refers to as test redesign is one component of the investigation activity that we previously discussed in my November 2005 column titled "Investigation vs. Validation." Surprise, it seems, is one key indicator that further investigation is probably a good idea, if for no other reason than to gather enough information to do a reasonable risk analysis of the importance of our observations.

Another key indicator that further investigation is probably a good idea is curiosity. For example, maybe the test results meet the expectations that have been set (making them not surprising), but after interacting with the application you find that the expectation that has been set doesn't seem to make sense—that is, you become curious. Maybe the expectations aren't consistent with one another. Maybe the specification doesn't match your experience with similar applications. Maybe the expectation just isn't intuitive. Additional investigation can go a long way to helping us express to stakeholders what led to our curiosity in the first place.

As I did some cursory research on surprise, I was surprised to find that surprise and curiosity are often addressed in the same academic research papers. One paper that I found particularly interesting described the relationship between surprise and curiosity this way:

"Humans are surprised when they perceive something that they did not expect;

humans feel desire to learn more (i.e., feel curiosity) about novel (possibly unexpected) objects, usually manifested by focusing the senses on those objects in order to study and analyze them."[3]

So it seems to me that as testers, whether our internal or external mumbling of "Hmm…that's odd" is inspired by surprise or curiosity really doesn't matter as much as what we do with the observation that led to the mumble in the first place.

From where I sit, it seems obvious that we should embrace any surprise or curiosity caused by our test results by redesigning our tests and investigating further.

Near the end of WTST5, it occurred to me that I was hearing my father's voice in my head. It turns out that my subconscious was replaying a thinking lesson from my father, who once said to me, "Sometimes it is more important (or valuable) to question the answer than to to answer the question."

For all that I often annoyed my parents and teachers by interpreting "sometimes" much too liberally, this lesson has proven extremely valuable to me.

Reflecting on why my subconscious chose that moment to replay this lesson, I realized that a slight revision of the lesson effectively summarized my thoughts on surprise as a testing heuristic. "When surprised by or curious about an answer or result, it may be more valuable or important to question the answer, result or question than to simply answer the question." ☒

> *'Sometimes it is more important to question the answer than to answer the question.'*

**ENDNOTES**

1. ExTRS was attended by James Bach, Jon Bach, Scott Barber, Michael Bolton, Elisabeth Hendrickson, Cem Kaner, Mike Kelly, Jonathan Kohl, James Lyndsay and Rob Sabourin.
2. WTST5 was attended by James Bach, Jon Bach, Scott Barber, Michael Bolton, Steve Condly, Becky Fiedler, Arthur Goldberg, Morven Gentleman, Doug Hoffman, Allen Johnson, Paul Jorgensen, David Klappholz, Cem Kaner, Panos Linos, Chang Liu, James Lyndsay, Pat McGee, Grigori Melnik, Rob Sabourin and Andy Tinkham.
3. Macedo, Luis and Cardoso, Amilcar, "Towards Artificial Forms of Surprise and Curiosity," eden.dei.uc.pt/~amilcar/publ/ECCS99.pdf, 1999.