Are We Overconfident in Our Understanding of Overconfidence?

Raymond R. Panko
Shidler College of Business
University of Hawai`i
2404 Maile Way
Honolulu, HI 96821
001.808.377.1149
Ray@Panko.com

ABSTRACT

In spreadsheet error research, there is a Grand Paradox. Although many studies have looked at spreadsheet errors, and have found, without exception, has error rates that are unacceptable in organizations, organizations continue to ignore spreadsheet risks. They do not see the need to apply software engineering disciplines long seen to be necessary in software development, in which error types and rates are similar to those in spreadsheet development..[1] Traditionally, this Great Paradox had been attributed to overconfidence. This paper introduces other possible approaches for understanding the Grand Paradox. It focuses on risk blindness, which is our unawareness of errors when they occur.

Categories and Subject Descriptors

K.8.1: Spreadsheets. D.2.5 Testing and Debugging.

General Terms

Experimentation, Verification.

Keywords

Methodology. Spreadsheet Experiments, Experiments, Inspection. Sampling, Statistics

1. INTRODUCTION

Despite overwhelming and unanimous evidence that spreadsheet errors are widespread and material, companies have continued to ignore spreadsheet error risks. In the past, this Great Paradox had been attributed to overconfidence. Human beings are overconfident in most things, from driving skills to their ability to create large error-free spreadsheets. In one of the earliest spreadsheet experiments, Brown and Gould [1] noted that developers were extremely confident in their spreadsheets' accuracy, although every participant made at least one undetected error during the development process. Later experimenters also remarked on overconfidence. Panko conducted an experiment to see if feedback would reduce overconfidence, as has been the case in some general overconfidence studies. The study found a statistically significant reduce in confidence and error rates, but the error rate reduction was minimal. Goo performed another experiment to see if feedback could reduce overconfidence and errors. There was some reduction in overconfidence but no statistical reduction in errors.

2. RISK BLINDNESS IN BEHAVIORAL STUDIES

This paper introduces other possible approaches for understanding the Grand Paradox. It focuses on risk blindness, which is our unawareness of errors when they occur.

Naatanen and Summala [9] first articulated the idea that humans are largely blind to risks. Expanding on this idea, Howarth [5] studied drivers who approached children wanting to cross at an intersection. Fewer than 10% of drivers took action, and those actions would have come too late if the children had started crossing the street. Svenson [14] studied drivers approaching blind bends in a road. Unfamiliar drivers slowed down. Familiar drivers did not, approaching at speeds that would have made accident avoidance impossible.

Fuller [2] suggested that risk blindness in experienced people stems from something like operant conditioning. If we speed in a dangerous area, we get to our destination faster. This positive feedback reinforces risky speeding behavior. In spreadsheet development, developers who do not do comprehensive error checking finish faster and avoid onerous testing work. In contrast, negative reinforcement in the form of accidents is uncertain and rare.

Even near misses may reinforce risky behavior rather than to reduce it. In a simulation study of ship handling, Habberley, Shaddick, and Taylor [4] observed that skilled watch officers consistently came hazardously close to other vessels. In addition, when risky behavior required error-avoiding actions, watch officers experienced a gain in confidence in their "skills" because they had successfully avoided accidents. Similarly, in spreadsheet development, if we catch some errors as we work, we may believe that we are skilled in catching errors and so have no need for formal post-development testing.

Another possible explanation comes from modern cognitive/neuroscience. Although we see comparatively little of what is in front of us well and pay attention to much less, our brain's constructed reality gives us the illusion what we see what is in front of us clearly [11]. To cope with limited cognitive processing power, the CR construction process includes the editing of anything irrelevant to the constructed vision. Part of this is not making us aware of the many errors we make [11]. Error editing makes sense for optimal performance, but it means that humans have very poor intuition about the error rates and ability to avoid errors [11]. For the CR process this is an acceptable tradeoff, but it makes us confident that what we are doing works well.

Another explanation from cognitive/neuroscience is System 1 thinking, which has been discussed in depth by Kahneman [7]. System 1 thinking uses parallel processing to generate conclusions it is fast and easy, but its working are opaque. If we are walking down a street and a dog on a leash snaps at us, we jump. This is fast or System 1 thinking. It is very effective and dominates nearly all of our actions, but it has drawbacks. First, it gives no indication that it may be wrong. Unless we actively turn on slow System 2 thinking, which we cannot do all the time, we will accept System 1 suggestions uncritically. One problem with doing so is that System 1 thinking, when faced with an impossible or at least very difficult task, may solve a simpler task and make a decision on that basis. For instance, if you are told that a bat and ball cost a dollar and ten cents and that the bat costs a dollar more than the ball, a typical System 1 thought response is that the ball costs ten cents. This is wrong, of course, but System 1 thinking tends to solve the simpler problem, $1.10 - $1.00. If we do not force ourselves to engage in slow and odious System 2 thinking, we are likely to accept the System 1 alternative problem solution.

This may be why, when developers are asked whether a spreadsheet they have just completed has errors, they quickly say no, on the basis of something other than reasoned risk. Reithel, Nichols, and Robinson [13] had participants look at a small poorly formatted spreadsheet, a small nicely formatted spreadsheet, a large poorly formatted spreadsheet, and a large nicely formatted spreadsheet. Participants rated their confidence in the four spreadsheets. Confidence was modest for three of the four spreadsheets. It was much higher for the large well-formatted spreadsheet. Logically, this does not make sense. Larger spreadsheets are more likely to have errors than smaller spreadsheets. This sounds like System 1 alternative problem solving.

3. CONCLUSION

If we are to address the Great Paradox successfully and convince organizations and individuals that they need to create spreadsheets more carefully, we must understand its causes so that we can be persuasive. Beyond that, we must address the Spreadsheet Software Engineering Paradox—that computer scientists and information systems researchers have focused on spreadsheet creation aspects of software engineering, largely ignoring the importance and complexity of testing after the development of modules, functional units, and complete spreadsheets. In software engineering, it accepted that reducing errors during development is good but never gets close to success. Commercial software developers spend 30% to 50% of their development resources on testing [6,8], and this does not count rework costs after errors are found. Yet spreadsheet engineering discussions typically downplay or completely ignore this five-ton elephant in the room. It may be that spreadsheets are simply newer than software development, but spreadsheets have been use for a generation, and strong evidence of error risks have been around almost that long.

We have only looked at the situation at the individual level. Testing must be accepted by groups and even corporations. Even at the group level, this paper has not explored such theories as the diffusion of innovations. If spreadsheet testing is mandated, that will reduce risks. However, user developers must have the freedom to explore their problem spaces freely by modifying their spreadsheets as their understanding grows. Testing methods must reflect the real process of software development.

4. REFERENCES

[1] Brown, P. S. and Gould, J. D. 1987. An experimental study of people creating spreadsheets. *ACM Transactions on Office Information Systems.* 5, 3 (Nov. 1987), 258-272.

[2] Fuller, R. 1990. Learning to make errors: evidence from a driving simulation. *Ergonomics,* 33, 10/11 (Oct/Nov, 1993), 1241-1250.

[3] Goo, Justin M. W. 2002. *The effect of feedback on confidence calibration in spreadsheet development.* Doctoral Dissertation, University of Hawaii.

[4] Habberley, J. S., Shaddick, C. A., and Taylor, D. H. 1986. *A behavioural study of the collision avoidance task in bridge watchkeeping.* College of Marine Studies, Southampton, England. Cited in Reason (1990).

[5] Howarth, C. I. 1990. The relationship between objective risk, subjective risk, and behavior. *Ergonomics, 31,* 527-535. Cited in Wagenaar & Reason, 1990.

[6] Jones, T. C. 1998. *Estimating software costs*. McGraw-Hill, New York, NY.

[7] Kahneman, D. 2011. *Thinking, fast or slow.* Farrar, Strauss and Giroux, New York, NY.

[8] Kimberland, K. 2004. *Microsoft's pilot of TSP yields dramatic results*, news@sei, No. 2. http://www.sei.cmu.edu/news-at-sei/.

[9] Naatanen, R. and Summala, H. 1976. *Road user behavior and traffic accidents*. North-Holland, Amsterdam. Cited in Wagenaar & Reason, 1990.

[10] Panko, R. R. 2007. Two experiments in reducing overconfidence in spreadsheet development. *Journal of Organizational and End User Computing,* 19, 1 (January–March 2007), 1-23.

[11] Panko, R. R. 2013. The cognitive science of spreadsheet errors: Why thinking is bad. *Proceedings of the 46th Hawaii International Conference on System Sciences (Maui, Hawaii,* January 7-10, 2013).

[12] Reason, J. 1990. *Human error.* Cambridge University Press, Cambridge, England.

[13] Reithel, B. J., Nichols, D. L., and Robinson, R. K. 1996. An experimental investigation of the effects of size, format, and errors on spreadsheet reliability perception. *Journal of Computer Information Systems,* 54-64.

[14] Svensen, O. 1977. *Risks of road transportation from a psychological perspective: A pilot study*. Report 3-77, Project Risk Generation and Risk Assessment in a Social Perspective, Committee for Future-Oriented Research, Stockholm, Sweden, 1977. Cited in Fuller, 1990.

[15] Wagenaar, W. A. and Reason, J. T. 1990. Types and tokens in road accident causation. *Ergonomics, 33,* 10/11 (Nov. 1993), 1365-1375.