# The Irrational Tester

James Lyndsay, Workroom Productions Ltd.

## Preamble...

People are built to be irrational. Irrationality affects our judgement and the way we make decisions. Even if we make the right decisions, misunderstanding our reasons may make it harder to learn, to repeat, to communicate and to act together. This paper attempts to give a testing perspective to common models of irrationality.

Readers of Dan Ariely's *Predictably Irrational* [1], Malcolm Gladwell's *Blink* [2], and Naseem Taleb's *Fooled by Randomness* [3] will recognise plenty of the ideas in this paper. Those ideas in turn have come from experiments in behavioural economics, and studies of cognitive bias. I am no psychologist. Nor am I a behavioural economist. This is by no means an exhaustive study. However, as a tester, I will illustrate these ideas with testing stories. I will translate some 'de-biasing strategies' into testing terms. I hope to direct the interested reader to primary academic sources (see the references for links), and look forward to other testers' interpretations of this material.

## Confirmation Bias

We seem to be wired to find what we expect to find. This tendency is called ***Confirmation Bias***. Confirmation bias is clearly relevant to software testers – the story below describes one such encounter with irrationality.

> *I recently found a bug in my own code. Software that had worked perfectly with Flash Players 5 through 9 announced that it would not work with Flash Player 10. I immediately (and with hindsight ludicrously) suspected that a comparison operator was behaving as an alpha comparison (9 comes after 1x) rather than a numeric one (9 comes before 10). I had spent some time researching type casting in ActionScript 2.0 before I bothered to check the code.*
>
> *The code revealed an entirely different bug. My code only paid attention to the first character – a basic error, and one I had ignored in chasing after my more esoteric model.*

While the source problem may have been technical, the larger problem was my bias. I was looking for something I expected to find, and in doing so, I had fallen prey to confirmation bias. With each step in my exploration of solutions, I had moved further from the source of my problem.

Confirmation bias covers a multitude of sins, with many different names but the same characteristic behaviour. At the time of writing, http://en.wikipedia.org/wiki/Confirmation_bias listed over a dozen named similar concepts.

Other facets and related ideas include:

- We seek out information that supports our expectations, and avoid exploring ideas that might reject our favoured model. This facet is sometimes called ***Congruence Bias***. (Wason [18] and Barron [4] p. 174 – chapter on Hypothesis Testing).

- When presented with information, we discount that which does not support us more readily than we discount that which does. As a result, two people with entrenched positions will use the same data to justify their opposing beliefs. (Lord et al. [14], Taber & Lodge [23]). This is called the ***Polarisation Effect*** or ***Assimilation Bias***.

- Our judgement of the quality of something is not only based on our interaction with that thing, but also on our prior expectations. Ariely calls this ***The Effect of Expectations*** (Ariely [1], Ch. 9).

- ***Inattentional bias*** is the label for looking so hard for something else that we don't see an unusual thing. In a fine variant [21], Paul Carvalho, a blogging tester, describes how his expectations of *not* finding something led him to miss it while looking straight at it.

- ***The Clustering Illusion*** labels the phenomenon of seeing groupings or patterns when none exist. It is often illustrated by the tale of the Texas Sharpshooter, who shoots a bunch of bullets into his barn, then paints a target over the holes. Bugs do cluster – displaying similar causes,

or similar behaviours – but not all clusters are meaningful. The problem for testers lies not in seeing a cluster, but in seeing a cluster and concentrating testing only on that cluster. [1]

When trying to avoid confirmation bias in test design and test approaches, it may help to keep the following questions [2] in mind: *With this test, might I see the same behaviour if something else was the cause? What tests can I invent to distinguish between plausible alternatives?*

The urge to avoid confirmation bias is, perhaps, the primary reason to keep testing teams independent of coding teams. However, diversity, depersonalisation, and discussion are also useful tools to reduce the impact of (and potential for) confirmation bias, and can be particularly powerful in groups who do not choose to rely on separation between team members.

**De-biasing strategies**

- Actively seek out disconfirming hypotheses
- Seek alternate explanations from independent judges
- Promote independence of thought as opposed to ignorance of context
- Give information about the product only *after* the subject has used the product
- Avoid tiredness and keep one's mind engaged

Confirmation bias is also called the ***Tolstoy Syndrome***, and the final word can be left to this quotation from Tolstoy's *The Kingdom of God Is Within You (1894)*

*The most difficult subjects can be explained to the most slow-witted man if he has not formed any idea of them already; but the simplest thing cannot be made clear to the most intelligent man if he is firmly persuaded that he knows already, without a shadow of doubt, what is laid before him.*

## Illusion of control

There are, of course, a host of other ways in which we can fool ourselves. Here is an illustration of the ***Illusion of Control***.

> *I was teaching a class. Everybody started testing the same thing at the same time. One person swiftly announced they had 'broken it'. I asked how. They said 'I clicked seventeen times, and it stopped working'. I asked whether they could reproduce the bug. The tester reset the software and clicked carefully, seventeen times. The system under test clanked to a halt and became unresponsive. 'See?' said the tester, with reasonable pride, "I've broken it again".*
>
> *The change in behaviour was not, however, caused by the tester. The change happened when an on-screen dial reached the end of its travel. The system would have stopped responding, whether the tester was clicking, or not. The tester had neglected to wonder if anything else was exerting control over the system they were testing.*

When testing, we often control elements of the input, the data, the software and the hardware. In order to diagnose problems and write good bug reports, we seek reproducible experiments, and refine our tests until every step is necessary and sufficient. It is easy to feel in complete control of software – but we should guard against that feeling; the illusion can have a direct effect on our competence.

---

[1] A cluster may simply be another way of saying "I found lots of what I went looking for", or "Everywhere I looked, I found bugs"… Also see Taleb's "Ludic fantasy" [3] – the misconception that life is random in the same way games are.

[2] Based on Baron's heuristics in his chapter on *Hypothesis Testing* (p. 174) in *Thinking and Deciding* [4]; 1) Ask "How likely is a yes answer, if I assume that my hypothesis is false?" 2) "Try to think of alternative hypotheses; then choose a test most likely to distinguish them – a test that will probably give different results depending on which is true.". This has similarities to Mayo's "severe test" ([5] p. 178) ; a test that has an overwhelmingly good chance of revealing the presence of a specific error, if it exists – but *not* otherwise

In *Trading on Illusions* [8], Fenton-O'Creevy et a l. describe an experiment to measure how 'in control' traders felt. Putting them into situations over which they had varying – but unknown – influence, the experimenters asked the traders to assess to what extent they were in control of the situation. By comparing the trader's accuracy with their (performance-related) earnings, the experimenters found that less-able traders were typically more likely to feel in control.

Can we extrapolate their results to testers? Building on prior work[3], the paper proposes that traders are particularly susceptible to illusions of control because i) they use judgement and model risk ii) they develop models of causal relationships iii) their environments have lots of noise to obscure a signal iv) they are stressed v) they are in competition vi) they are goal-focussed vii) they are doing skilled work in a familiar domain. Many of these factors[4] apply equally well to testers.

**De-biasing strategies**

- Develop awareness of causes – especially those that are not in one's control
- Encourage a 'deliberative' mindset
- Seek feedback and reduce personal 'ownership' of strategies
- Recognise that one can feel 'in control' by controlling one's responses to the environment [5]. In practice, this means recognising a failing strategy, and cutting one's losses.

## Endowment effect

If you've ever found it hard to prune an over-long test suite because you remember how much effort it took to build, then you've noticed ***Loss Aversion***; the harder we've worked for something, the less we are willing to give it up.

This is a facet of the ***Endowment Effect***. It is summed up in *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias* [10], as "People often demand much more to give up an object than they would pay to acquire it".

The authors of that article, Kahneman, Knetsch and Thaler, detail their seminal experiments investigating the endowment effect in their paper *Experimental Test of the Endowment Effect and the Coase Theorem* [11]. In their experiments, students are encouraged to participate in various markets. The items in the market would be just as useful to the buyer as they are to the seller, but the markets are designed to be tools to examine situations where sellers are not prepared to sell at a price buyers are prepared to offer. Kahneman et al. put this down to ownership; ownership distorts markets.

In *Predictably Irrational* [1], Ariely makes the case that ownership has an effect on our beliefs and our opinions; our view of the world and our place within it. We place greater value on positions that have cost us something, and so are more unwilling to give up those positions. When challenged with inertia or ideology, it may be time to think about the endowment effect.

**De-biasing strategies**

- Although Ariely draws a blank on a universal approach to avoiding endowment bias, he recommends disinterest, viewing the situation from a position of non-ownership.
- Kahneman et al., dealing with economic theory, advise in [10] that more precise responses to change may be achieved by separating, rather than aggregating, the effects of favourable and

[3] cf. Langer 1975 [13] "when an individual is actually in the situation, the more similar the chance situation is to a skill situation in outcome-independent ways, the greater will be the illusion of control. This illusion may be induced by introducing competition, choice, stimulus or response familiarity, or passive or active involvement into a chance situation. When these factors are present, people are more confident and are more likely to take risks."

[4] in my opinion; i, ii, iii, vii (testers use skill and judgment to model cause and effect in a noisy but familiar environment)

[5] Control of one's responses is "secondary control". Control of one's environment is "primary control". Secondary control promotes adaptation and is perhaps simpler to assess more accurately.

unfavourable change. In [11], they note that disputes are simpler to resolve by asking people to give up future gains, than by asking them to give up what they already hold.

## Failure to commit

In *Predictably Irrational* ([1] Ch. 6. and [6]), Ariely describes an experiment conducted with three undergraduate classes. The experiment was designed to reveal information about how deadlines affect our performance. He set each class three pieces of work. For one class, he set equally spaced deadlines. For the second, he set a single, end-of-term deadline. He allowed each student in the third class to set their own deadlines. Performance was judged on the mark given to a piece of work. Work which missed a deadline was not marked.

Students in the first class (spaced deadline) performed better than students in the second (one end-of-term deadline). In the third class, students who set well-spaced deadlines did better than those who grouped their deadlines closer to the end of term. Although those with the latest deadlines had the most time to do the work, they performed least well.

In a different experiment ([1] Ch. 8. and [16]), Ariely engineered a situation where the subjects earned money for a trivial task. They were able to influence what they were paid by selecting from limited options. For some subjects, the options become more limited with each selection. The experiment was designed to reveal information about the ways in which options (and the loss of options) affect decision-making.

Ariely found that subjects with unchanging options earned more money than those whose unused options receded. People tended to keep all their options open – even when the cost of keeping those options was more than any benefit that the low-value option could bring. This seems particularly plausible where options were similar, or where their ultimate consequences were hard to estimate.

If we can legitimately extend Ariely's results to the familiar world of software projects, his experiments provide novel illumination. Testers who are keen to navigate all available options (I'm one) may be at a disadvantage without conscious, swift commitment to a course of action. Swift and clear commitment has the benefit of reducing decision fatigue, and it is plausible that – for some circumstances – the benefit of making a decision is more than any projected potential loss of making the wrong decision. Perhaps there is room for a coin flip rule; "If we're dithering, we should consider tossing a coin so we can start moving".

As tactics to avoid the irrationality his experiments show so clearly, Ariely recommends using spaced deadlines, and taking swift choices. With this in mind, two common software development practices stand out: Session-based timeboxes in exploratory testing, and sprints on agile projects. Although sprints and timeboxes work at different scales, and are set by a team on the one hand, and individuals on the other, both invoke regular, hard deadlines. Both make people focus on available options, by deferring their options until a debriefing period after the sprint or session. Sprints and sessions help us commit our focus towards fixed, regular deadlines.

A further procrastination experiment, described in The Economist [19] and detailed in McCrea [15], reveals that "People act in a timely way when given concrete tasks but dawdle when they view them in abstract terms"[6]. As a clear illustration, "almost all the students who had been prompted to think in concrete terms completed their tasks by the deadline while up to 56% of students asked to think in abstract terms failed to respond at all". What is interesting here is that this was not to do with the tangibility of the goal, but the way that the students were *prompted to think* about it. Again, if this experiment can be extrapolated[7] then it is important not only to have a clear goal, but to have an environment that encourages team members to think in concrete terms.

[6] A position which will be familiar to readers of Dave Allen's well-known "Getting Things Done"

[7] It may be that this heuristic applies only to tasks that are, like those in the experiment, relatively easy and only moderately important. A reference in McCrea [15] ("Dewitte and Lens (2000) argued that chronic procrastinators focus on task details to such an extent that they feel overwhelmed") suggests that it may be a hard course to steer between concrete action, and just-enough detail.

**De-biasing strategies**

To avoid procrastination and fruitless keeping-open of options

- Commit
- Commit frequently
- Commit early
- Commit to something that you can do
- Think in concrete terms

## Broken Windows

When considering perceptions of quality and bug fixing priorities, it may be interesting to consider an idea that first came to prominence in issues of civic order in New York.

Kelling and Wilson, in their 1982 magazine article *Broken Windows* [20], put forward the intuitive but controversial idea that vandalism is encouraged by vandalism, and that swift fixes to small damage in public spaces are an important part of preventing larger damage. Their idea became the basis of the "zero-tolerance" approach to street crime in New York. These ideas are put to the test in experiments described in Keiser's *The Spreading of Disorder* [12]. The experimenters looked at people's behaviour in situations where rules had been visibly flouted.

In one experiment, they set up a barrier across a shortcut to a car park. The barrier had two notices on it; one to forbid people to trespass, the other to forbid people to lock their bicycles to the barrier. 82% of people pushed past the barrier when it had bicycles locked to it. 27% pushed past when the bicycles were locked up a meter away. Three times as many people were prepared to break the trespass rule when the bicycle rule had already been broken. Similar effects were observed for littering and for theft. It made no difference whether rules had been set legally, by social norms, or arbitrarily. People were more likely to deviate from 'normal' behaviour in environments where a rule – any rule – had been broken ostentatiously.

Imagine, then, a parallel with obvious yet minor bugs. If a cosmetic bug is left unfixed, those who see the bug may be less inclined to follow 'normal' behaviour. If the software is under construction and the bug is regularly seen by those building the software, could it act to discourage conscientious development? If the system is in use, might an obvious bug nudge users towards carelessness with their data? Both these questions are speculative, but if we can accept that pride in our work tilts us to do better work, we should also accept that we could allow ourselves to work less carefully on a sub-standard product.

## Caveat

A little knowledge is a dangerous thing [8]. Studies of bias are fascinating, but fickle, and a warning is in order. This last irrationality has no particular link to software testing, but every bit of relevance to this paper.

Recognising irrationality may simply confirm one's suspicions that everyone else is irrational. If you find yourself thinking "That's what's wrong with these people; they're biased", or "I knew I was surrounded by fools and charlatans from the very start" then my keen advice is that you pause for a moment.

The ***Dunning-Kruger effect*** [9] is the name given to a worrying and common trait: Most people think their abilities are above average. Even the least competent consider themselves to be really rather good[9]. Confidence in one's own competence is, by its nature, something to beware. Confidence arrived at without study, practice or an independent assessment all the more so.

In *Unskilled and Unaware of It* [9], Kruger and Dunning explore the idea that "Competence begets calibration". They describe experiments where those who were least competent were trained – and subsequently were able to revise their estimates of confidence to one which more accurately reflected their relative performance. In a later paper, *Why the Unskilled are Unaware* [7], Ehrlinger expands on this to say that the metacognitive understanding "intelligence is malleable" might also lead to more accurate self-assessments.

I hope this paper will encourage its readers to be more aware of their decisions, and less accepting of their defaults. The reader should take *themselves* as the first point of scrutiny. Adjustment for bias is a way of tempering one's *own* decisions and understanding.

### De-biasing Strategies

- Prompt yourself to remember that if you don't have the skill, you probably *don't have the skill*. Then go get the skill.

- If you notice that you are using claims of bias to give you more ammunition with which to argue against something you don't like, then that is nature's way of telling you that you're suffering from confirmation bias.

[8] In his blog article Overcoming Bias: Knowing About Biases Can Hurt People [22], Eliezer Yudkowsky (who seems to know his stuff) offers advice to anyone who is considering bringing their understanding of bias to the notice of others: *Whether I do it on paper, or in speech, I now try to never mention calibration and overconfidence unless I have first talked about disconfirmation bias, motivated skepticism, sophisticated arguers, and dysrationalia in the mentally agile. First, do no harm!* In commentary to his posting, he later offers the pithy statement *the error is not absence of generalization, but imbalance of generalization, which is far deadlier.* My apologies for not being able to follow his advice to the letter in this paper.

[9] Oddly, the most competent consider themselves to be, comparatively, less able than they truly are. But don't get your hopes up…

## Conclusion

A tester's judgement is central to their job. If we can hope that rational decisions are a worthwhile end in themselves, then it seems crucial to appreciate that we may be subject to bias, and useful to be able to recognise common biases.

We can help ourselves work more swiftly and more effectively by focusing our minds on the choices we have made. We can help ourselves work more harmoniously by recognising that we hold our beliefs more strongly simply because they are ours. Perhaps recognising our biases will also help us recognise ways in which we can learn.

My one line heuristic for avoiding bias as a tester is to care less, and to commit sooner. You will have different ideas – and I hope you will share them with the community at large.

James Lyndsay, Workroom Productions Ltd.

London, February 2009

This paper may be found at www.workroom-productions.com

Contact: jdl@workroom-productions.com

Skype / iChat / twitter: workroomprds

© Workroom Productions 2009. Some rights reserved.

James Lyndsay has asserted his right under the Copyright, Designs and Patents Act 1988 to be identified as the author of this work.

This work is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 2.0 UK: England & Wales License. http://creativecommons.org/licenses/by-nc-sa/2.0/uk/

Under this license, you may copy, distribute, display, and perform the work, and make derivative works. However, you must give the original author credit, you may not use this work for commercial purposes without prior permission, and if you alter, transform, or build upon this work, you may distribute the resulting work only under a licence identical to this one.

***This is version 1.06 of this paper. All comments gratefully received – let's make 1.1 better by far.***

*Changed for v 1.06 from 1.0 - new preamble, adjusted wording in conclusion, fixed minor errors and ambiguities*

## Appendix: References

**Further reading (pop science)**

[1] Ariely, Dan; "Predictably Irrational" *HarperCollins (2008)*

[2] Gladwell, Malcolm; "Blink" *Little, Brown (2005)*

[3] Taleb, Naseem; "Fooled by randomness" *Random House (2004)*

Want more? James Surowiecki "The Wisdom of Crowds" (wonderful, and potentially debiasing all on its own); Malcolm Gladwell "The Tipping point" (interesting, but glib); Clay Shirky "Here Comes Everybody" (which seems to *ahem* ¿share? rather a lot of sources with…); Philip Ball "Critical Mass" (which will make anyone with a background in statistical mechanics mash their head into their desk in frustration).

**Reference Books**

[4] Baron, J; "Thinking and Deciding" *Cambridge University Press, 4th edition, 2008*. Excerpt of referenced pages available via Google books

[5] Mayo, D; "Error and the Growth of Experimental Knowledge" *The University of Chicago Press, 1996*. Excerpt of referenced pages available via Google books. Only tangentially related to this paper, but interesting for testers.

**Papers (alphabetically, by author)**

*Links to .pdfs are given if they appear to be sanctioned by the publisher or author*

[6] Ariely & Wertenbroch, "Procrastination, Deadlines, and Performance: Self-Control by Precommitment", *Psychological Science No. 13 (3), 219-224 (2002)*. Available at http://www.predictablyirrational.com/pdfs/deadlines.pdf

[7] Ehrlinger, Johnson, Banner, Dunning, & Kruger, "Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent," *Organizational Behavior and Human Decision Processes, vol. 105, no. 1, pp. 98-121, January 2008*. Available from http://www.psy.fsu.edu/~ehrlinger/Self_&_Social_Judgment/Ehrlinger_et_al_2008.pdf

[8] Fenton-O'Creevy, et al. "Trading on illusions: unrealistic perceptions of control and trading performance." *Journal of Occupational and Organizational Psychology, 76(1), pp. 53-68 (2003)*.

[9] Kruger & Dunning; "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments" *Journal of Personality and Social Psychology, Vol. 77, No. 6. (1999)*. Version available at http://www.apa.org/journals/features/psp7761121.pdf

[10] Kahneman, Knetsch & Thaler."The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives 5(1), 193-206 (1991)*. Available from http://faculty.chicagobooth.edu/richard.thaler/research/articles/1-The_Endowment_Effect_Loss_Aversion_and_Status_Quo_Bias.pdf

[11] Kahneman, Knetsch, & Thaler "Experimental Test of the endowment effect and the Coase Theorem." *Journal of Political Economy 98(6), 1325-1348 (1990*). Available from http://faculty.chicagobooth.edu/richard.thaler/research/Experimental%20tests%20of%20the%20Endowment%20Effect%20and%20the%20Coase%20Theorem.pdf

[12] Keizer et al. "The Spreading of Disorder" *Science 12 December 2008: 1681-1685*. Available from http://www.ppsw.rug.nl/~lindenb/documents/articles/2008_Keizer_Lindenberg_Steg_The%20spreading_of_disorder_Science_Express.pdf

[13] Langer, E. J. "The illusion of control." *Journal of Personality and Social Psychology, 32(2), 311- 328, (1975)*.

[14] Lord, Ross & Lepper; "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence" *Journal of Personality and Social Psychology Vol. 37, No. 11 (1979)*. Available version at http://synapse.princeton.edu/~sam/lord_ross_lepper79_JPSP_biased-assimilation-and-attitude-polarization.pdf

[15] McCrea et al., "Construal Level and Procrastination" *Psychological Science, 19 (12), 1308-1314, (2008*). Version at: http://www.uni-konstanz.de/FuF/SozWiss/fg-psy/gollwitzer/PUBLICATIONS/McCreaetal.PsychSci09.pdf

[16] Shin & Ariely, "Keeping doors open: The effect of unavailability on incentives to keep options viable," *Management Science, vol. 50, no. 5, pp. 575-586, (2004)*. Available at http://www.predictablyirrational.com/pdfs/doors.pdf

[17] Taber & Lodge, "Motivated skepticism in the evaluation of political beliefs," *American Journal of Political Science, vol. 50, no. 3, pp. 755-769, July 2006*.

[18] Wason, P.C.; "On the failure to eliminate hypotheses in a conceptual task." *Quarterly Journal of Experimental Psychology, 12, 129-140 (1960)*.

**Magazine Articles**

[19] The Economist (no byline), "Motivating minds" *The Economist Jan 22nd 2009*. Available from http://www.economist.com/science/displaystory.cfm?story_id=12971028

[20] Kelling & Wilson, "Broken Windows", *The Atlantic Monthly, March 1982*. Available from http://www.theatlantic.com/doc/198203/broken-windows

**Blog postings**

[21] Paul Carvalho, "Lessons Learned by a Software Tester: Don't ask me to look.. I'm tired.", 2004. Available from http://swtester.blogspot.com/2004/12/dont-ask-me-to-look-im-tired.html

[22] Eliezer Yudkowsky, "Overcoming Bias: Knowing About Biases Can Hurt People", *overcomingbias.com, April 2007*. Available from: http://www.overcomingbias.com/2007/04/knowing_about_b.html

**Websites**

Kickoff with http://en.wikipedia.org/wiki/Cognitive_bias and http://www.google.com/search?q=cognitive+bias. When you get to http://www.overcomingbias.com/ , you're cooking.