Open source inevitably good

I've been reading an excellent book this week called “The Wisdom of Crowds: Why the Many are smarter than the Few” by James Surowiecki. The reason I am writing about it is that it highlights something which it occurred to me was highly relevant to the great proprietary v. open source debate.

I have been struggling for some time to understand fully why a bunch of enthusiasts using a software process officially deemed chaotic, have managed to produce arguably the most reliable complex application in history (the Linux kernel) and many other comparably high quality applications such as Apache. Certainly there are some obvious things such as the “many eyeballs” theory eloquently described by Eric Raymond in his compelling essay “The Cathedral and the Bazaar” which we know as large-scale code inspections, and there are also some much less obvious things such as the unusually high levels of perfective maintenance, (maintenance which seeks to improve in some way without changing functionality). However, what I have read so far in Surowiecki so far suggests there may be yet another layer which I hadn't consciously anticipated.

Surowiecki's book itself is implicitly based around a couple of equations which communications engineers have known about for years. Let me illustrate. I would like you to imagine a thousand people listening for a very distant sound so quiet that some of them will not even be able to hear it. All they have to do is say “Yes” if they hear it and “No” if they don't. The astonishing thing is that by counting up the votes, you can estimate very accurately exactly how loud it is although no one person would be able to do this. Surowiecki illustrates this phenomenon with a wide variety of examples, some of which are frankly spooky, although unsurprising to a mathematician. His opening example is wonderful and describes how 19th century British scientist Francis Galton happened to attend a country fair at which a crowd had to guess the weight of an ox. He expected that one or two good judges of this would easily out-perform a crowd of amateurs. However, when he averaged the crowd's guesses, the average was only a pound out. This is NOT a coincidence. Even more compelling is the account of how the lost US submarine Scorpion which disappeared in 1968 was found. Expert advice turned out to be useless - the submarine could not be found, the area was simply too large. However a naval officer called John Craven combined lots of expert's evidence using something called Bayes theorem. The resulting position coincided with none of the individual expert's but turned out to be within 220 yards of the lost submarine.

The requirements for crowds acting together to have this seemingly astonishing prescience are basically that they all have some knowledge and that they act effectively independently. One ideal forum described in the book is when there is a regular supply of new ideas which the 'crowd' votes on by either supporting or not supporting.

You can probably see where this is going because this is exactly the Darwinian breeding ground through which open source software goes and what you see has been unconsciously but unerringly selected by the anonymous crowd's combination of knowledge and independence. High quality open source isn't a surprise, its inevitable.