On the Law of Diminishing Returns in R&D and Science

Stanislav Sykora, Extra Byte, www.ebyte.it

First published in February 2014

The *law of diminishing returns*, well known from economics, actually extends to any sufficiently complex development, such as R&D of large software packages, complex mechanical and/or electronic devices, social and political institutions, and even Science itself. A simple example is provided to clarify why this is unavoidable. The illustration makes it also clear what are the basic rules to counteract, at least in part, the undesirable tendency, and what dangers they entail. I have found these realizations useful in my work and hope that somebody else might find them useful as well. As an after-thought, I have added also a reflection of the implications of the "law" to Science (intended as a complex product of Humanity) which might be of some epistemological interest.

Keywords: law, economy, software development, high-tech development, R&D, science, epistemology

Illustration of the thesis

It is one of the simple rules of Economics that returns from a specific technology (or, really, from any productive activity) keep diminishing with time [1,2].

Indeed, it makes sense that at the onset of any innovation cycle, there is the novelty factor which can be assigned a value of its own and, at the same time, there is little or no competition. Consequently, pricing is based on "how much is the market willing to pay" rather than on "what are the production costs". This can be maintained for a considerable time because the market, initially small and local, keeps expanding towards a global dimension. But once the technology/activity reaches maturity, there is probably a flourishing competition with diversified models, the market is saturated, and prices are based rigidly on production costs (and even those are being systematically decreased).

So far, everything is clear.

But why do we see *"diminishing returns on the same amount of effort and expenditure"* also in areas of pure research and development (R&D)? Returns intended in terms of technical performance of the final products, not in financial terms!

For example, if one starts a large software project such as verifying molecular structure on the basis of some kind of spectra (NMR, MS, IR, ...), one notices the phenomenon in a few months, and it becomes prominent after maybe 1-2 years. The same applies, I am sure, to any other high-tech developments, be it racing cars, communication technologies, 3D printers, weather forecasting, whatever. In my case, I am thinking and talking about Automatic Structure Verification (ASV) in NMR spectroscopy [3], but that is not really the point.

To understand the "why" of the phenomenon, one must first understand the fact that, at some practical, real-world level, everything is fuzzy and statistical. Nothing in the real life is perfect, everything is full of artifacts, and no rule is absolute and safe from exceptions. Hence, whatever modification one does to the object under development, it needs to be tested statistically for various replicas and under various conditions. Making it work perfectly for a single instance is of no use at all, because it would most likely not be robust – it might fail grotesquely in the next instance or, much worse, in an actual deployment!

Hence the need for testing benchmarks (large sets of instances) in any R&D process. Indeed, if you are an R&D manager, the benchmarks, and the testing methods themselves, should be one of your principal worries. Some people, like weather forecasters, are lucky because their benchmark data are supplied to them daily by the Nature itself; one just needs to keep a record and be patient. In other areas the benchmarks need to be build up, sometimes laboriously and not without controversy. Benchmarking and testing methods are a non-trivial branch of the R&D management science.

Now, assuming that we have a satisfactory testing set-up, what happens can be explained using a few simple examples:

First, suppose that we are in the initial stages of a project and we have a very bad "product", one that fails 50% of times. In weather forecasting, for example, that might be a plain "random guess method".

Now, somebody comes up with an idea (such as the "let me ask my granny" method), and a few preliminary tests show that it works just great. It reduces the incidence of failures by 40%. But, since nothing is perfect, it also ruins about 10% of the cases which were passing. Nevertheless, it is 40 to 10 in favor of the method. In fact, if we apply it to 100 cases, the statistical expectations are:

Of the 50 cases that were failing, now 30 fail and 20 pass.
Of the 50 cases that were passing, now 5 fail and 45 pass.

In total, we expect 65 cases passing and 35 failing. Great! 65:35 is a great improvement over 50:50.

Now suppose that we are more advanced and our refined "product" has 80% success rate. If we implement the new "plank", this is what we expect:

Of the 20 cases that were failing, now 12 fail and 8 pass.
Of the 80 cases that were passing, now 8 fail and 72 pass.

In total, we expect 80 cases passing and 20 failing, which means no gain at all !?! A useless modification even though, used alone, it has some merit.

Indeed, if our "product" were still better, reaching 95% success, we would get these expectations:

Of the 5 cases that were failing, now 3 fail and 2 pass.
Of the 95 cases that were passing, now 9.5 fail and 90.5 pass.

The failure rate went up from 5% to 12.5% !!! Since the quality of products is generally measured by their failure rates rather than success rates, it means that in this case, implementing the modification, the quality is EXPECTED to get worse by a factor of 2.5! Hence, the modification is too low-quality and is must not be adopted.

Any sophisticated "product", be it an ASV software or a Ferrari car, went necessarily through a long development history in which there were many implemented changes and improvements and added items[1]. Consequently, because of the principle we have just illustrated, the progress kept getting slower and more laborious.

That much to understand **the basic phenomenon of ever diminishing returns**.

Of course one might also simply say that the more sophisticated and highly-tuned is a system, the easier it is to spoil it by adding anything that is not comparably sophisticated. You do not stick a screwdriver into a running quantum computer, right?!

Or one might cast it in terms of entropy: highly sophisticated and useful objects are unusual and unlikely to happen spontaneously. In other words, they have high entropy and, as we all know, entropy does have the nefarious tendency to find ways to drop down.

[1] The nature of the items varies widely from one application field to another. In software they can be algorithms (or fragments thereof) pieces of code, in mechanics design tricks, in electronics circuits, etc. Above all, everywhere, concepts. In an attempts at abstractions, I shall henceforth call them ***planks***.

Counter-actions

However, the above line of reasoning can be exploited further. There is little doubt that some of the *planks* implemented in the early stages of the development of any product were of the low-quality type. At the moment when they were adopted, they helped a lot and therefore were appreciated, even cherished! But it is quite possible that they are no longer helpful in the present, more sophisticated stage. The presence of such obsolete items in any complex product[2] is extremely likely, and they may clutter the development pipelines and hinder further progress.

This is a valid lesson and a precious insight:

Don't get hooked on something just because it used to work nicely in past stages of a project.

In general, constant additions of concepts, tools, methods, refinements and other *planks* are important, but merciless pruning of the old ones is just as important, too! If an old, rustic *plank* is ruining the scores, adding anything sophisticated will never help, until the offender is removed. In my ASV project, I have seen it over and over[3].

Removing ageing *planks* from a project is not a demerit, it is a necessity!
It is the only way to mitigate the Law of Diminishing Returns!

And it does not mean that the removed *planks* are "bad" or "useless".
They are just not useful enough, which is something very different.

Of course, once we realize these principles, there is also the opposite danger of being hasty and removing prematurely still useful *planks*. This is hard to avoid (in my ASV project it happened several times). I have not yet found an objective method how decide when to prune an old item.

For example, a drop in performance upon the removal of a *plank* does not necessarily mean much. It happens almost always; it is extremely rare that, upon removal of a *plank*, there is an immediate surge in performance. This is because in a sophisticated product there are many items which were adjusted (tuned) to work together as a system and removing any one disturbs this system equilibrium. Consequently, the normal behavior upon removal of a low quality *plank* is an initial drop in performance, followed by a recovery – and eventually an actual improvement – upon readjusting (retuning) the whole system, provided the latter is done competently.

But, as I said, these are just hunches; I have no idea how to decide the proper pruning moment objectively.

Application to epistemology

It strikes me that the above must apply also to Science, intended as Humanity's sophisticated R&D project. Here, at the basic level, the *planks* are the various paradigms of science, and the "natural laws". We of course do not have any access to "real" natural laws, what we can do are just more or less successful hypotheses, mostly embodied in mathematical models, and tested against reality in much the same way as the weather forecasts.

On the average Science (intended as a product) is becoming more and more reliable and sophisticated, though it is still very far from what we would like it to be. And therefore it is subject to the same R&D rule stated above: while it is evolving, some old *planks* need to be abandoned because they are detrimental to the overall performance and hinder further progress even though, considered individually, they are still potentially useful.

---

[2] Even the "products" of natural evolution, such as the Homo Sapiens species, can not escape this rule. There are numerous obsolete leftover items in the human body and mind which are no doubt a hindrance today.
[3] Sorry to say it, but many of the chemical rules supplied as potential *planks* by my chemist friends [4] turned out to be of the grand-nanny's weather forecasts category. Incredible what a poor programmer has to live with ☺ .

One question that this raises is *what, in the context of Science, is the benchmark*.

Some people think that it is our *"understanding"* of The Reality and/or of The Universe [5,6]. They say, or imply, that Science has some kind of "duty" to explain and make us understand the objective Reality.

Other's think that it is the *elegance and internal coherence of our theories*. A more self-contained and shiny theory is considered to be more likely to be "true", regardless of anything else.

I strongly oppose all of that as irrational and outright foolish, on the following grounds:

1) We can't, and never will, really "understand" anything. Why things exist, and why specific relations between them exist are, at some deep enough level, impenetrable ontological mysteries. For example, we can understand why friction exists, describing it as the result of molecular interactions. But we will never understands why interactions exist, to start with. They just do! We can only describe them, never understand their "being". That, I think, is quite obvious. Consequently, Science is not about understanding, it is "just" about describing.

2) Internal elegance of the Science edifice is only in the eyes of the beholder. What looks simple to me might appear complicated to you, and even more so to an inhabitant of the Beetle Juice planetary system. Religious people, for example, will find it preposterous to think that an almighty Creator might find *anything* complex or, for that matter, rank things according to their complexity. An atheist, on the other hand, realizes that any such ranking is necessarily purely subjective and there is no reason why Nature should heed it. For example, Nature solves most crazy systems of differential equations in real time, so to say, it does not need any computers! In this respect, Nature is *au pair* with Gods.

This, the way I see it, leaves as a Science Benchmark only the **capability to predict**, to foresee. Contrary to common belief, it is quite possible to predict without complete understanding. And the outcomes of predictions (just like in weather forecasting) can be used as a statistical benchmark test.

But this is another story ...

References

[1] Wikipedia, Diminishing returns.
[2] Case K.E., Fair R.C., Principles of Economics (10th ed.). Prentice-Hall, 2011.
[3] Cobas C., Seoane F., Sykora S., *Automatic Structure Verification (ASV) as an AI wizard*, Poster at 54th ENC, Asilomar, 14 Apr 2013, DOI: 10.3247/SL4Nmr13.002.
[4] Manuel Perez, Carlos Cobas, Michael Bernstein, Chen Pen, Gonzalo Hernandez, and many others: private communications.
[5] Wiki Answers, What is the goal of science.
[6] University of Berkeley, Understanding Science 101.

History of this document

9 Feb 2014: Assigned a DOI (*10.3247/SL5Epis14.001*) and uploaded online as Stan's Library entry.