EUT EDIZIONI UNIVERSITÀ DI TRIESTE

Futuribili – Rivista di studi sul futuro e di previsione sociale vol. XXVIII, nn.1/2, 2023
ISSN (print) 1971-0720
DOI: 10.13137/1971-0720/37596
https://www.openstarts.units.it/handle/10077/15705

TACKLING COMPLEX SYSTEMS: WE NEED NON-GALILEAN SCIENCE

Ugo Bardi, member of The Club of Rome and The World Academy for Arts and Sciences, former professor of Chemistry at the University of Florence

Abstract: *While Galileo studied simple linear systems, modern science attempts to study complex, multi-parameter systems, where the rigid Galilean method just cannot work. The problem is that, while it is obvious that we can measure only what we can measure, that's not necessarily what we want, or need, to measure. Tests based on "surrogate endpoints" may well be the best we can do in medicine and other fields, but we should understand that the results are not, and cannot be, a source of absolute scientific truth. We need a new scientific method to work on these systems.*

Keywords: *Galileo, non-linear science, linear science, complex systems, surrogate endpoints.*

Sommario: *Mentre Galileo studiava sistemi lineari semplici, la scienza moderna cerca di studiare sistemi complessi e multiparametrici, dove il rigido metodo galileiano non può funzionare. Il problema è che, sebbene sia ovvio che possiamo misurare solo ciò che possiamo misurare, non è necessariamente ciò che vogliamo o dobbiamo misurare. I test basati su "endpoint surrogati" possono essere il meglio che possiamo fare in medicina e in altri campi, ma dobbiamo capire che i risultati non sono, e non possono essere, una fonte di verità scientifica assoluta. Abbiamo bisogno di un nuovo metodo scientifico per lavorare su questi sistemi.*

Parole chiave: *Galileo, scienza non lineare, scienza lineare, sistemi complessi, punti finali surrogati.*

The Scientific Method

Galileo Galilei is rightly remembered as the father of modern science because he invented what we call today the "scientific method," sometimes still called the "Galilean method." It is supposed to be the basis of modern science; the feature that makes it able to be called "Science" with a capital S, as we were told over and over during the COVID pandemic. But what is really this scientific method that's supposed to lead us to the truth?

Galileo's paradigmatic idea was an experiment about the speed of falling objects. It is said that he took two solid metal balls of different weights and dropped them from the top of the Tower of Pisa. He then noted that they arrived at the ground at about the same time. That allowed him to lampoon an ancient authority such as Aristotle for having said that heavier objects fall faster than lighter ones (not true). There followed an avalanche of insults to Aristotle that continues to this day. Even Bertrand Russell fell into the trap of poking fun at Aristotle, accused of having said that women have fewer teeth than men. Too bad that Aristotle never said anything like that.

It may well be that Galileo was not the first to perform this experiment, and it is not even clear that he performed it, but that's a detail. The point is that the result was evident, clear-cut, and irrefutable. Later, Newton started from this result to arrive at the notion that the same force that acted on an apple falling from a tree in his garden was acting on the Moon and the planets. From then on, science was supposed to be largely based on laboratory experiments, or at any rate experiments performed in tightly controlled conditions. It was a major change of paradigm: the basis of the scientific method as we understand it today.

The Tower of Pisa experiment succeeded in separating the two parameters that affect a falling body: the force of gravity and air drag. That was relatively easy, but what about systems that have many parameters affecting each other? Here, let me start with the case of health care, which is supposed to be a scientific field, but where the problem of separating the parameters is nearly impossible to overcome.

Surrogate endpoints

How can you apply the scientific method in medicine? Dropping a sick person and a healthy one from the top of the Tower of Pisa won't help you so much. The problem is the large number of parameters that affect the nebulous entity called "health" and the fact that they all strongly interact with each other.

So, imagine you were sick, and then you feel much better. Why exactly? Was it because you took some pills? Or would you have recovered anyway? And can you say that you wouldn't have recovered faster had you not taken the pill? A lot of quackery in medicine arises from these basic uncertainties: how do you determine the specific cause of a certain effect? In other words, is a certain medical treatment really curing people, or is it just their imagination that makes them think so?

Medical researchers have worked hard at developing reliable methods for drug testing, and you probably know that the "gold standard" in medicine is the "Randomized Controlled Test" (RCT). The idea of RCTs is that you test a drug or a treatment by keeping *all* the parameters constant except one: taking or not taking the drug. It is designed to avoid the effect called «placebo» (the patient gets better because she believes that the drug works, even though she is not receiving it) and the one called «nocebo» (the patient gets worse because he believes that the drug is harmful, even though he is not receiving it).

An RCT involves a complex procedure that starts with separating the patients into two similar groups, making sure that none of them knows to which group she belongs (the test is "blinded"). Then, the members of one of the two groups are given the drug, say, in the form of a pill. The others are given a sugar pill (the "placebo"). After a certain time, it is possible to examine if the treatment group did better than the control group. There are statistical methods used to determine whether the observed differences are significant or not. Then, if they are, and if you did everything well, you know if the treatment is effective, or does nothing, or maybe it causes bad effects.

For limited purposes, the RCT approach works, but it has enormous problems. A correctly performed RCT is expensive and complex, its results are often uncertain and sometimes turn out to be plain wrong. Do you remember the case of "Thalidomide"? It was tested, found to work as a tranquilizer, and approved for general use in the 1960s in Europe. It was later discovered that

it had teratogenic effects on fetuses, and some 10,000 babies in Europe were born without arms and legs before the drug was removed from the market. "*Tests on animals*" would have shown the problem, but they were not performed or were not performed correctly.

Of course, the rules have been considerably tightened since the Thalidomide disaster and nowadays testing on animals is required before a new drug is tested on humans. But let's note, in passing, that in the case of the mRNA COVID vaccines, tests on animals were performed in parallel (and not before) testing on humans. This procedure exposed volunteers to risks that normally would not be considered acceptable with drug testing. Fortunately, it does not appear that mRNA vaccines have teratogenic effects.

Even assuming that the tests are complete, and performed according to the rules, there is another gigantic problem with RCT: What do you measure during the test? Ideally, drugs are aimed at improving people's health, but how do you quantify "health"? There are definitions of health in terms of indices called QALY (quality-adjusted life years) or QoL (quality of life). But both are difficult to measure, and if you want long-term data you have to wait for a long time. So, in practice, "surrogate endpoints" are used in drug testing.

A surrogate endpoint aims at defining measurable parameters that approximate the true endpoint – a patient's health. A typical surrogate endpoint is, for instance, blood pressure as an indicator of cardiovascular health. The problem is that a surrogate endpoint is not necessarily related to a person's health and that you always face the possibility of negative effects. In the case of drugs used to treat hypertension, negative effects exist and are well known, but it is normally believed that the positive effects of the drug on the patient's health overcome the negative ones. But that's not always the case. A recent example is how, in 2008, the drug *bevacizumab* was approved in the US by the FDA for the treatment of breast cancer based on surrogate endpoint testing. It was withdrawn in 2011, when it was discovered that it was toxic and that it didn't lead to improvements in cancer progression (you can read the whole story in "*Malignant*" by Vinayak Prasad).

Consider now another basic problem. Not only is there *the number of parameters affecting people's health*, but the fact that they strongly interact with each other, as is typical of complex systems. The problem may take the form called "polydrug use," which especially affects old people who accumulate drugs on

their bedstands, just like old cars accumulate dents on their bodies. An RCT test that evaluates one drug is already expensive and lengthy; evaluating all the possible combinations of several drugs is a nightmare. If you have two drugs, A and B, you have to go through at least three tests: A alone, B alone, and the combination of A+B. If you have three drugs, you have seven tests to do (A, B, C, AB, BC, AC and ABC). And the numbers grow rapidly. In practice, nobody knows the effects of these multiple drugs uses, and probably nobody ever will. But a common observation is that when the elderly reduce the number of medicines they take, their health immediately improves (this effect is not validated by RCTs, but that does not mean it is not true. I noted it in my mother-in-law, who died at 101).

The case of Face Masks

Some medical interventions have specific problems that make RCTs especially difficult. An example is that of face masks to prevent the spreading of an airborne pathogen. Evidently, there is no way to perform a blind test with face masks, but the real problem is what to use as a surrogate endpoint. At the beginning of the COVID pandemic, several studies were performed using cameras to detect liquid droplets emitted by people breathing or sneezing with or without face masks. That was a typical "Galilean", laboratory approach, but what does it demonstrate? Assuming that you can determine if and how much a mask reduces the emission of droplets, is this relevant in terms of stopping the transmission of an airborne pathogen? As a surrogate endpoint, droplets are at best poor, at worst misleading.

A much better endpoint is the PCR (polymerase chain reaction) test that can directly detect an infection. But even here, there are many problems. As an example, consider an *often touted study* performed in Pakistan that claimed to have demonstrated the effectiveness of face masks. Let's assume that the results of the study are statistically significant (*really?*) and that nobody tampered with the data (and we can never be sure of that in such a heavily politicized matter). Then, the best you can say is that *if* you live in a village in Pakistan, *if* there is a COVID wave ongoing, *if* the PCR tests are reliable, *if* the people who wore masks behave exactly like those who don't, and *if* random noise

didn't affect the study too much, *then* by wearing a mask you can delay being infected for some time, and maybe even avoid infection altogether. Does the same result apply to you if you live in New York? Maybe. Is it valid for different conditions of viral diffusion and epidemic intensity? Almost certainly not. Does it ensure that you don't suffer adverse effects from wearing face masks? Duh! Would that make you healthier in the long run? We have no idea.

The Pakistani study is just one example of a series of studies on face masks that were found to be ill-conceived, poorly performed, inconclusive, or useless in a recent rigorous review published in the Cochrane Network. The result is that no one has been able to detect a significant effect of face masks on the diffusion of an airborne disease, although we cannot say that the effect is zero.

The confusion about face masks reached stellar levels during the COVID-19 pandemic. In 2020, Tony Fauci, director of the NIAID, first advised against wearing masks, then he reversed his position and publicly declared that face masks are effective and even that two masks are better than just one. Additionally, he declared that the effectiveness of masks is "science" and, therefore, cannot be doubted. But, nowadays, Fauci has *reversed his position,* at least in terms of mask effectiveness at the population level. He still maintains that they can be useful for an individual "*who religiously wears a mask*." Now, imagine an RCT dedicated to demonstrating the different results of "religiously" and "non-religiously" wearing a mask. So much for science as a pillar of certainty.

Conclusion

Many fields of science suffer from similar problems of method. Climate science, ecosystem science, biological evolution, economics, management, policies, and others are cases in which you cannot reproduce the main features of the system in a laboratory and which, at the same time, involve many parameters interacting with each other in a non-linear manner. You could say, for instance, that the purpose of politics is to improve people's well-being. But how could that be measured? In general, it is believed that Gross Domestic Product (GDP) is a measure of the well-being of the economy and, hence, of all citizens. Then, it is concluded that economic growth is always good, and

that it should be stimulated by all possible interventions. But is it true? GDP growth is another kind of surrogate endpoint used simply because we know how to measure it. But people's well-being is something that we don't know how to measure.

Is *non*-Galilean science possible? We *must* start considering this possibility without turning to discarding *the need for good data and good measurements. But, for complex systems, we *must* move away from the rigid Galilean method and use dynamic models. We are moving in that direction, but we still have to learn a lot about how to use models and, incidentally, the COVID-19 pandemic showed us how models can be misused and *lead to various kinds of disasters*. We need to work much harder to set modern science on a solid footing, as Galileo did. Will we ever be able to do that? Time will tell.

Acknowledgement – The Author thanks Dr Alberto Gasparini for his support for this article. Disclaimer: this article was not supported by any external agency. No part of it was written by an AI.