DOI: 10.2478/pepsi-2018-0006

Albert Sesé*

Discussing the Scientific Evidence Generation and Psychological Research Methods in Postmodern Societies

ABSTRACT

Evidence generation by current Social and Health Sciences is coping with some important barriers that difficult credibility of scientific products. Information and communication technologies have a strong impact over social relationships in our postmodern societies. The incidence of post-truth in our context is generating a pernicious relativism, far from contrasting the information veracity. The aim of this paper is to analyze and discuss the challenges of research methods and statistical models, more specifically for Psychological research, taking into account the impact of novel techniques as big data and virtual reality. Special attention is also devoted to the discussion about statistical shortcomings of psychological research and to the reproducibility problem. Finally, some potential solutions are proposed to be applied in order to improve the quality of scientific evidence.

KEYWORDS: psychological research methods; statistical models; postmodern societies; big data; virtual reality.

* Correspondence regarding the paper should be sent to: Albert Sesé, Department of Psychology, Balearic Islands University, Carretera de Valldemossa, km. 7'5, 07122 Palma de Mallorca, Spain, e-mail: albert.sese@uib.es

DISCUSSING THE SCIENTIFIC EVIDENCE GENERATION AND PSYCHOLOGICAL RESEARCH METHODS IN POSTMODERN SOCIETIES

A new scenario

The irruption of new Information and Communication Technologies (ICT) has undoubtedly brought about a revolution that is transforming our postindustrial societies. It is a transformation that transcends not only the questions of the modernization of our daily work, but also the way in which we try to understand our reality. New technologies are changing our relationship and communication patterns, and can be considered a double-edged sword: on the one hand, they facilitate the flow of information, immediate communication without geographical limits, and the handling of huge amounts of information (big data); but on the other hand, they are a fertile ground for manipulation and social control. Obviously, a good result depends, then, on the appropriate use or not of such technological advances. Far from establishing an assessment of convenience or not of ICT, since they have come to stay and are already part of our lives, it is important to critically analyze their impact in the context. It is essential to find patterns of use of ICT that ultimately assist in achieving a happier humanity.

From, above all, the fall of the Berlin Wall, and the end of the Cold War, we live in a Volatile, Uncertain, Complex, and Ambiguous (VUCA) context. Volatility refers to the idea that situational occurrences cause certain social categorization that can speedy be changed and misunderstood; Uncertainty happens when the availability or predictability of information in events is unknown; Complexity refers to how social categories impacted the process of social cognition and perception, and to the difficulties associated with these psychological processes; and Ambiguity is defined as the scenario where relevant information is available but the overall meaning is still unknown. This acronym, originally formulated

by military intelligence, describes a complicate context, where ICT can play a crucial role. In general, the meaning of VUCA often relates to how people view the conditions under which they perceive and process information, make decisions, plan forward, manage risks, foster change and solve problems. For this reason, a VUCA context is a challenge itself for the educational models of the XXI century. It is not easy to prepare people for coping with a volatile, uncertain, complex and ambiguous world, where probably more than 50% of future jobs are not yet invented.

The so-called "Information Society" (InSoc), which has derived from the industrial society, and its "digital citizens", requires reliable and valid criteria for the management of information and decision-making. And Internet has become a kind of oracle where we can find any type of information easily without any filter other than your own perception. We can say that if the amount of information was competitive in the past, the quality of information is crucial today. One of the worst menaces for this new InSoc is the "post-truth" (World international word of 2016), that is, a deliberate distortion of a reality, which manipulates beliefs and emotions in order to influence public opinion and social attitudes. Internet is plenty of post-truths for influencing people to vote for some political parties, to consume some products, to inoculate ideologies… A post-truth always includes some part of true information together with false information. This way, we tend to read from reality only that which fits what we previously believe, and we do that even knowing that we are undermining objectivity. It is the growing primacy of our emotional response over fact and evidence, the replacement of verification with social media technical mechanisms that tell us what we want to hear. And the worst thing of this phenomenon is that we do not care about not to check the evidence. So, the success of post-truth depends to a large extent on the psychological effect of the self-affirmation of one's own prejudices by reading their manipulated content biasedly. In consequence, people are giving truth value

to the mere opinion and, in consequence, legitimate skepticism is yielding place to pernicious relativism. In fact, we can say that healthy skepticism is agonizing in the intensive care unit. And the icing on the cake is, therefore, to confuse the freedom of expression with believing that any opinion has the value of truth. This attitude dangerously opens the way to dogmatism, opinion polarization, and sectarianism. This new scenario can be considered even more worrisome than the classical "panem et circenses" of Roman societies.

A few weeks ago when I was preparing a conference on educational assessment, I wanted to use the phrase: "I fear the day when technology surpasses our humanity; the world will only have a generation of idiots". This phrase is attributed to Albert Einstein. But what were my surprise and my stupor when I checked that Einstein never left these words in writing and that the phrase had appeared on the internet more or less only since 2012. Although I can agree with the content of the phrase, the name of such an illustrious scientist cannot be used to endow such phrase with a greater emotional charge of truthfulness. What Einstein really wrote in 1948 was literally: "I believe that the abominable deterioration of ethical standards stems primarily from the mechanization and depersonalization of our lives". This is a "light" example of post-truth on the Internet but it serves to illustrate the silent inoculation of a social virus that treats to relativize any information, it is the war of fakes. Living in a VUCA globalized world, and grappling with the threat of post-truth, science and education are the most powerful weapons with which to try to equip people with resources so that they can be able to build a better and truer world.

Denying the evidence. The need for an Evidence-Based Practice (EBP)

As a result of some information retrieved from social media, some families decide not to vaccinate their children, or to use ho-

meopathic remedies, or to invert a lot of money in virtual coins… If reading this phrase we start looking for information on the Internet that reinforces these practices, we will find pages and pages that justify them, and even provide supposed evidence of its good result. Far from wanting to create more controversy with these issues, the incidence of post-truth on the Internet may be having a very negative impact on the attitude of searching for contrasted scientific evidence. This scenario also contributes to extreme VUCA context that generates a lot of uncertainty and hinders decision making processes. In my opinion, one of the best principles of the scientific method is to keep always alive a reasonable doubt about any theory or model under a continuous improvement principle. That is, science defends at all costs that evidence that has been rigorously contrasted until then, but will defend with that same force alternative evidence, even contrary to the previous one, in the case that it will show a better fit to reality. Empirical contrast, rigor, precision and parsimony are the key elements that sustain the eternal scientific curiosity. The inaccurate, untested, ambiguous and biased mechanism of post-truth may represent the antithesis of the scientific method. The lack of verification of information as an attitude may be generating a security breach in the waterline of the scientific method and also facilitating the emergence of so-called pseudo-sciences. This pernicious relativity also feeds the most bizarre conspiracy theories against science itself. Denying the evidence and installing a new relativist order, any model can be falsely recognized as truthful without testing.

So much so, that even scientists themselves and the transfer of evidence to professional practice are being affected. In what sense? Professionals begin to show significant resistance to change in their practice when new scientific evidence appears that modifies the existing action guidelines. Although it is true that there are other organizational barriers and lack of resources to implement scientific evidence to practice, an important part of the negative

disposition to change falls on the beliefs and attitudes of professionals. It is precisely in beliefs and attitudes where the general tendency to not verify the facts is increasing the resistance to change. A symptomatic effect of that resistance to change with respect to implementing scientific evidence is the emergence of back-to-method movements such as Evidence-Based Practice (EBP). In a simple way, EBP means returning to the path of the scientific method, diving into its findings and implementing them in professional practice. There is nothing new under the sun, only a claim to return to give credit only to those theories and models that have been empirically tested, and whose impact on our reality be cost-efficient. Many evidence-based practice guides are being developed to try to standardize procedures and interventions, mainly in the fields of Social and Health Sciences. Implementation of EBP means to improve the quality of the professionals, the intervention programs, the effect on patients/clients, and in sum, to improve the quality of our societies. But it is necessary to adopt a pragmatic attitude and apply high doses of patience because a lot of time is required for the generation of an organizational and individual change of such magnitude. The fight for being effective applying validated theoretical models can be also helped by the establishment of a culture of public accountability and sustainability.

The replicability problem

Going back a few steps back, prior to the implementation of scientific evidence to practice, it is also important to stop to analyze the quality of the process of generating the evidence itself. Other but not less important principle of Science is replicability. Evidence is not evidence if it has not gone through the rigorous filter of systematic replication, and even if it has not demonstrated its capacity to be generalized to other contexts, situations, people, etc. And the question is how many effective replications are needed to consolidate the scientific validity of evidence? The answer is

not simple but strictly speaking evidence never is totally consolidated because a reasonable doubt about the existence of a better alternative must be always kept in mind. But being pragmatic, obviously we should say that the more replications we carry out, the greater degree of security we will have about the validity of any evidence. In fact, scientific replicability can be compared with a normal distribution that has theoretical infinity but practical finitude. At this point, scientists in general are facing a growing problem with the possibility of disseminating the results of their research. Given that the usual way to disseminate scientific evidence is to publish articles in journals, the role of the editorial management of scientific journals is therefore very important. What is published or what is not published depends in the first instance on the decision of the editors, who consider sending or not to peer review the articles they receive. This first filter can be justified by the high number of articles that authors send to journals, but there is a risk that selection biases will operate, even non-conscious ones. Although personal biases in the selection of articles can pose a serious danger to the process of disclosure of evidence, I would like to focus on the pressure that the impact indexes by which journals are evaluated exert on the initial filtering carried out by the editors. Researchers have a lot of experience about how the Impact Factor (IF) of journals operates, and how to publish or not to publish articles in "impacted" journals affects their careers. A journal with high IF implies that the articles published in there have received a high number of citations. That is, it can be considered that the evidence contributed by those articles has been taking into account by many researchers of this field. In this way, the editors of the journals try to filter those articles with a greater degree of novelty, pioneers, with a high level of unique contribution. In short, they try to publish those studies that can potentially receive a high number of citations. Nothing, shall we say, illicit, there is in this editorial practice. If an article shows new evidence, a great number of citations can be expected to be

received. And if I as editor can publish many of those articles, I will put my journal at the top of the list of journals with high impact. Where is the problem, really? The problem is that the desire to publish only very novel articles prevents the publication of replication studies. If there are significant difficulties for the publication of replications, the power of the evidence can drastically decrease. It can be called the problem of "single studies". Only one study but published in a high impacted journal can be dangerously considered as powerful evidence. The problem isn't necessarily that this single study was poorly designed, but without further replications, without a meta-analysis of the totality of the evidence generated on the question to reach a conclusion, the initial study can remain under suspicion. Any meta-analytic approach, rather than any one study in isolation, is likely to get closer to true evidence.

An article recently published in the prestigious journal "Science" by the Open Science Collaboration (2015) carried out replications of 100 experimental and correlational studies previously published in 3 Psychology Journals: "Psychological Science", "Journal of Personality and Social Psychology" and "Journal of Experimental Psychology: Learning, Memory and Cognition" using high-powered designs and original materials when available. The Open Science Collaboration is a group of 270 Psychology researchers from around the world interested in reproducibility of scientific evidence. Results indicated that replication effects were half the magnitude of original effects, representing a substantial decline because only 36% of replicated studies were statistically significant ($p < .05$) in original direction. Also the mean of original studies effect size was .403 while the mean for replicated studies dramatically decreased to .197. These results must not be considered as a defeat for Psychology, or Science more generally, but on the contrary to real value the need for doing replications and for disseminating its results: "Accumulating evidence is the scientific community's method

of self-correction and is the best available option for achieving that ultimate goal: truth" (Open Science Collaboration, 2015, p.7). What I am sure about it is the scientific method is not and never must be ideological, and we have to clearly see that Science does not always provide us the answers we realize, but it is an instrument to confront us with reality.

Focusing not on scientific journals but on newspapers, another interesting article has been recently published in the journal PLoS One (Dumas-Mallet, Smith, Boraud, & Gonon, 2017). The authors of this paper assembled a huge database of studies in biomedical science, follow-ups to those studies, and meta-analyses on those follow-ups. After that they searched the Dow Jones Factiva newspaper database to see how often each type of study was covered. Results indicated that initial studies were around five times more likely to be reported on than follow-up studies, and that meta-analyses were barely covered at all by newspapers. According with these results, many biomedical findings reported by newspapers are disconfirmed by subsequent studies. Authors conclude that this is partly due to the fact that newspapers preferentially cover "positive" initial studies rather than subsequent observations, in particular those reporting null findings. By the way, null findings that would not be easily published by journals with impact index, perhaps not even by journals without an impact factor.

Once it seems clear that there is a very important pressure on researchers to publish novel articles and, of course, with statistically significant effects, it is no less evident that we can become *p*-hacking slaves. Researchers can become slaves of the torture, massage or cooking of our data to make them reach the barrier of statistical significance, this $p < .05$ value. Although the responsibility for good practices rests fully with researchers, we cannot deny that the pressure of the process of publishing articles in journals with impact factor, as I pointed out before, can turn us into victims. This should not be interpreted as an excuse but as

a possible explanation of how some part of the current scientific production works. Many young and even not so young researchers live the data analysis situations as very stressful events. If preliminary tests obtain non-significant results, some researchers start a complex strategy that finally can lead to reach the statistical significance; for example, extracting different random or not random subsamples, implementing data imputation methods in presence of missing values, increasing the initial sample in order to obtain more statistical power, if it is possible. If in spite of applying such a massage to the data it is not possible to obtain significant results, the researchers self-impose the judgment of non-publication of the study. And the truth is that they are not wrong because the editorial policy of the journals is absolutely contrary to the publication of non-significant studies.

However, it is important to note that a non-significant result, in the case that the study does not present statistical power problems, therefore, having a good capacity to detect significant differences when in fact there are, can be as much or more relevant than a significant result. If the study was correctly designed, had a high degree of internal and external validity, controlling potential confounding variables, with a representative sample, with enough statistical power (.80 or higher), and a correct statistical model estimated, any non-significant result can lead to criticize a new hypothesis in novel studies, or previous significant studies in replication studies. Under these conditions both novel and replication studies with non-significant results should be published. A perverse effect can emerge from those studies with poor methodological conditions (little and non-representative samples, non-adequate designs, no control for confounding variables, low statistical power and validity, and inadequate or incorrect estimated models) in two ways: a) if the study obtain non-significant results, a true hypothesis can be incorrectly discarded forever; and b) if the study shows significant results, a bad hypothesis can be incorrectly assumed as true. In both cases, these studies

should not be published in any scientific journal. It does not mean that only perfect studies should be published, because perfection does not exist in science. It means that researchers have to fight for achieving the better conditions to carry on the better studies.

Be in a hurry is usually a very bad companion of scientific research. But many times the need to obtain funding, generally through competitive research projects, or the need to accredit merits to access a researcher position, can precipitate the development of studies and the publication of their results under inadequate conditions. In my opinion, the rush for publishing articles in journals with impact factor, caused in large part by the selection procedures of researchers by means of scoring, can turn scientific dissemination into a question of quantity and not of quality. That way, publishing articles can become a formal game in which it is more important to know the strategies to be able to pass the editorial filters of the journals than the quality of the evidence contributed by the articles. In order to avoid these inadequacies of science production it is extremely important that, on one hand, senior researchers ensure that young researchers learn not to sacrifice the quality of studies in favor of a greater number of published articles; and on the other hand, the editors of scientific journals ensure that peer review processes require strict adherence to the application of the scientific method in the articles reviewed. In short, it is a matter of scientific responsibility, of ethics in research. Finally, it is also crucial to point out that the real impact factor of any article is not only the mean average of cites obtained in the last two years but the social impact on people of its generated evidence.

Magical thought on statistical models

Many researchers but especially PhD and MD students have the false belief that a study achieves more quality depending on the more advanced statistical model that has been implemented. Although it is true that the use of advanced statistical models al-

lows the contrast of complex hypotheses, however these models will not be useful if the methodological design of the research is not adequate. Another false belief is to think that a statistical relationship between two variables implies causality. This false statement reinforces the magical thought about statistical methods. But causality only can be tested using a rigorous methodological design. This false belief about the use of complex statistical models reinforces the magical thought about its power for increasing the quality of scientific evidence. Applying an adequate statistical model is a necessary but not sufficient condition for achieving good scientific evidence. Research design is the necessary framework where to obtain data with enough quality to subsequently be able to correctly apply statistical models for the contrast of the proposed hypotheses. Even though this is true, unfortunately it seems that the weight of research is more focused on statistical models and their complexity than on the quality of research designs. In fact, some journals with an impact factor may also make the mistake of overestimating the use of complex statistical models without taking into account the quality of the data obtained through an adequate methodological design. But this is an old question and a skeleton in the closet of Social and Health Sciences. I like to discuss with my students one phrase said by one of the fathers of Modern Statistics, Sir Ronald Aylmer Fisher: "To consult the statistician after an experiment is finished is often merely to ask him/her to conduct a post mortem examination. He/She can perhaps say what the experiment died of". This sentence formed part of the Presidential Address that Fisher conducted in the First Indian Statistical Congress in 1938. According to this idea, a large part of the worktime of any research team should be devoted to trying to find the best methodological approach to operationalize the hypotheses to be tested. And of course, if possible, the research groups should include methodologists in the team. Once a design is implemented and the data has

been obtained to be analyzed, no statistical model will be able to fix serious flaws of the research design used.

Despite the existence of a magical thought about the capacity of advanced statistical models, it is paradoxical to see how the use of statistical tools has important shortcomings. Therefore, the problem lies not only in believing that statistical models are more important than research designs, but also that the use of such models is not carried out with the correctness and rigor that should be done. Loftus (1996) clearly points out that Psychology will be a better science in as much as it improves its way if analyzing data. But at the same time he literally stated: "Sometimes I feel that what we do in research in Psychology is like trying to build a violin with a stone mallet and a chainsaw. The tools we apply to the task are not the appropriate ones and, as a result, we end up building a large quantity of bad quality violins". Researchers don't have also to put into practice automatic sets of steps about statistics, on a way of a "cook book" or a "cheating sheet", which is probably condemned to failure. I call this practice "the bad musician syndrome", and it is referred to a person who is able to play almost perfectly a musical piece, but she/he is completely unable to play that same piece if there has been any change in its score.

At the root of this problem probably underlies a lack of appropriate statistical background that can make researchers misunderstand data analysis procedures or outputs. In this sense, the appropriate use of both optimal research designs and statistical techniques may help us to discover complex relationships between the variables under study and it can make scientific progress be more fruitful. I conducted a review of 623 articles published during 2010 in 8 journals with high impact factor belonging to the category "Psychology, Clinical" of the Subject Category Listing of the Journal Citation Reports database (Sesé & Palmer, 2012). This article presented a panoramic view of the degree of use of statistical methodology and its level of diversity and complexity

in Clinical and Health Psychology. The study aimed to analyze whether a greater use of statistical techniques corresponded with the journals with a greater impact factor value. Results effectively showed a non-parametric positive correlation of .62, but quantity should not be confused with adequacy, that is, it is not the number of techniques used which should be relevant, but rather the use of the most appropriate technique, in other words, the most powerful one, according to the nature of data. Regarding the possible improper use of statistical techniques, the study highlighted some important shortcomings concerning relevant statistical information that is not provided. A clear example of this shortcoming is the provision of effect size, which only takes place in 52.78% of the reviewed studies. Another alternative way of offering effect sizes is to provide the confidence intervals for the estimated parameters, which only appeared in 18.87% of articles. Another important informative shortcoming refers to statistical power analysis, as only 3% of the studies included the estimated value of a priori power, and 3.64% provided the observed or empirical power. Finally, as far as the assessment of the fulfilment of statistical assumptions corresponding to each statistical model, only 17.27% of articles stated carrying out an analysis of such assumptions. This low incidence is as or even more worrying than the shortcomings previously referred, as the application of techniques when faced with possible violation of assumptions may seriously compromise the veracity of the statistical conclusions obtained. With these results it is difficult to assert that the real situation regarding the quality of the generation of statistical inference is negative, but in the absence of information concerning assumption analysis, many of the statistical conclusions generated may be compromised. Hence it is extremely important to implement and reference statistical assumption analysis and their result in each study. For this reasons, researchers and journal editors must be self-responsible to provide this essential statistical information and demand it during the review process if it is not

present, respectively. Although the psychological literature still presents some deficiencies in the quality of the technical report of the statistical analysis, little by little both researchers and editors are increasing the rigor and the correction in this task.

The solution for this problem of poor statistical background is education. An example of some mistakes about the rudiments of statistics is provided by Badenes, Frías, Monterde, and Pascual (2015). These authors tested the potential misunderstanding of the *p*-value in a sample of 418 Psychology professors from all psychological subdisciplines including Methodology. The task was to assess as true or false a set of 5 interpretations in relation to the statistical results of a study, and specifically on the *p*-value obtained. Information provided was: "A research article indicates a *p*-value of .001 in the results section (alpha = .05)". And the five statements to be marked as True or False were:

1. The null hypothesis has been shown to be true
2. The null hypothesis has been shown to be false
3. The probability of the null hypothesis has been determined ($p = .001$)
4. The probability of the experimental hypothesis has been deduced ($p = .001$)
5. The probability that the null hypothesis is true, given the data obtained is .001

Please, I invite you to do this simple task.

What do you think if I tell you that none of the statements is true? Keep calm. The study showed that only 6.2% of all participants marked the five statements as False, and only 19.4% of methodologists. A true statement with a $p = .001$ can be: The probability of the results of the statistical test is known, assuming the null hypothesis is true; or for example, given that $p = .001$, the result obtained makes it possible to conclude that the differences are not due to chance.

In consequence, empirical evidence suggests that many academic psychologists do not know how to correctly interpret

something as basic as the *p*-value. The inverse probability fallacy presents the greatest comprehension problems: It is the false belief that the *p*-value indicates the probability that the null hypothesis (H0) is true, given certain data ($Pr(H_0 | Data)$). My point is: if even Methodology instructors also interpret the significance of the *p*-value erroneously, how many mistakes can be made with advanced statistical models? The inappropriate use of statistical methods may lead to distorted results, incorrect conclusions, and substantial waste of financial and other resources and is considered highly unethical (even when it is unintentional).

At this point, I dare to suggest a set of tips to avoid malpractice: To spend all the time it takes to choose the most efficient methodological design to test our hypotheses; to search for the optimal statistical model for our data, and to avoid automatized steps; to break a magical thought about statistical models, "the more complex the better", "don't use sledgehammers to crack nuts"; to improve serious misunderstanding about statistical inference interpretation and data reporting; and to strictly follow the guidelines of the APA "Task force on statistical inference" or any other actual standard for applying research methods in Psychology or other Social and Health Sciences. In sum, I believe that to improve the quality of scientific evidences in our fields depends on a triply sophisticated demand for researchers: 1) much more reflexive approach of theoretical models; 2) expertise in research designing; and 3) a deeper statistical background for data analysis, correct results interpretation, and adequate information report. And there are two not incompatible but complementary ways to make it possible: The active implication of methodologists in interdisciplinary research teams; and/or to provide the researchers a much better background and expertise in research methods.

Some challenges, new trends and perspectives for research methods

The impact that ICTs, computer development, internet, and mobile telephony are having on human behavior is incontestable. The technological society announced decades ago by postmodernist researchers such as Alain Touraine or Daniel Bell has arrived and is modifying our behavioral patterns, both socially and individually. Social networks are ubiquitous in our lives since we can easily access them at any time through a smartphone, which is nothing more than a computer. Applications like WhatsApp, Facebook, Twitter, or Instagram, among many others, have produced a change in the way we communicate with each other, and also in the way we send and receive information. All these instruments can make life more comfortable, but also can promote undesirable social control and other effects like to be jonesing for having wifi coverage, or to be anxious or depressed according with how many "likes" we received in our virtual wall. In fact, there are some research projects about the anxiety level that people experiments when no internet coverage is available, or about the time people waste in useless tasks with the smartphone, or about the decreasing quality of the human relationships owing to these new communication channels. Undoubtedly, new forms of communication have to be studied in order to minimize pernicious effects, both social and individual, and try to integrate constructive learning about the responsible and healthy use of social networks and the internet.

Looking back a little, maybe only about 35 years, I remember myself analyzing data with an immensely large computer, with a ridiculous capacity, and fed by data punch cards. I remember when a simple exploratory factor analysis had to be developed by hand and rotation iterative procedures needed more than 3 days to be finished… Fortunately, at the present time we have the best statistical armament that we have ever had, and incredible laptops with almost unlimited capacities for analyzing data

provided by research. One of the most relevant tasks in research consists of obtaining large and representative samples; enough data for fitting models and generate useful evidence. Many studies cannot generate quality evidence precisely because they have not achieved a sufficient number of participants, of data. In this sense, ICT and the mobile technologies have opened an immaterial door for getting thousands of data in real time. Despite the privacy breach problems that can happen, and in fact are happening, due to our interaction with the internet, a good use of such channels can place us in an unknown dimension regarding the possibilities of massive data research. The term "Big Data" initially describes the large volume of data, both structured and unstructured, that inundates a business on a day-to-day basis. And the management of this big data requires complex storage systems. Managing massive data for a company can help to understand and even try to explain the behavior of markets, customers, suppliers, employees, and with this information, design strategies of maximum cost-efficiency. This trading concept has been adapted to social research in the sense that extremely large data sets obtained from virtual channels can be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions. And one important advantage of this new scenario is that this massive data can be obtained in real time. Making an analogy, these new data collecting procedures can act as a classical *Holter* for measuring blood pressure. There are a lot of psychological processes or phenomena that would be susceptible to be studied by means of this approach. For example, these real time data may be extremely useful for monitoring subjects with suicide ideation problems. In a few years, I am sure that sophisticated sensors installed in the smartphones or inserted in our clothes (wearables) will make possible to obtain physiological data, even biomarkers, also in real time. In that sense, other skeleton in the closet of psychological research, the lack of longitudinal designs, can be clearly improved

because these technologies make easier to keep people engaged to the measurements. Needless to say that all these processes require the informed consent of participants and the research must accomplish the corresponding ethical codes.

Big data also allows the analysis of a huge amount of variables in search of complex relationships, with an explanatory, predictive, or classificatory purpose. But traditional statistical models are not capable to handle many variables at the same time, and for this reason, Data Mining techniques are the best tools for analyzing big data. The most powerful advantage of this kind of advanced statistical models is its capacity for finding complex relationships or patterns among thousands of variables and millions of subjects, and that these techniques do not require the fulfilment of statistical assumptions. In this last sense, Data Mining models operate as an exploratory and non-parametric approach. You can find a lot of different statistical models that can be considered as a Data Mining technique: Artificial Neural Networks, Bayesian nets, Random Forests, Matching Learning, Association Rules, Regression Trees, among many others. I recommend the lecture of the paper titled "A Practical Guide to Big Data Research in Psychology" recently published in the prestigious journal Psychological Methods (Chen & Wojcik, 2016) as a primer for those readers interested on these models.

Another relevant application of Big Data and Data Mining is the development of expert systems or Artificial Intelligence (AI). Getting a machine to make decisions and execute a series of actions autonomously has been the golden dream of many researchers and even novelists for decades. Today that dream has been made possible thanks to the robotic technology and the statistical procedures of pattern learning and decision making. The well-known Siri (Apple) and Alexa (Amazon) are cybernetic personalities who can interact with us by responding to our requests for information or even having a meaningful conversation. These intelligent systems can be fed back with new patterns and pro-

gressively adjusting a more contingent response to each situation. Taken to the extreme, allow me the license to say that we would find ourselves facing the scenario proposed in the cinema by the Matrix trilogy, where machines and computer programs come to subdue humans. Discarding this terrifying and utopian vision, AI can generate important changes in our daily domestic habits, in driving cars, and even in the health care we receive. But perhaps more important than the statistical algorithms that we must use to establish optimal AI models are the ethical criteria that must guide the response of intelligent machines to ambiguous situations, and obviously such decisions must be fully consensual. This issue is already being discussed for the development of autonomous vehicle driving systems. For example, in a traffic situation in which the car cannot avoid a hit, but can choose between two options, where one of them who is run over is a dog, and the other is a person, which of the two options should choose the vehicle? Who do we decide to run over the machine, the animal or the person? What if the options were a child and an elder? This is really a challenge, not only for science, but for humanity, which is the one who must determine the range of tolerable responses of an intelligent machine, especially in the presence of moral dilemmas. Again, from an optimistic perspective, if we manage to use the AI properly, we will be able to achieve great advances that allow us to enjoy a less hostile and more sustainable world.

Last but not least, we find the great potential of virtual reality techniques for social sciences and especially for psychology. If AI can collaborate to develop intelligent machines, virtual reality can create and recreate infinite scenarios where we can participate and interact for therapeutic, training, learning, or even preventive purposes. The quality of current software developments and programming is of such magnitude that the ecological validity of virtual reality can make us believe that we are living a real situation. Psychologists can perfectly imagine therapeutic sessions about arachnophobia, social skills, moral dilemmas, or emotional

intelligence. And at the same time that the virtual reality session is taking place, the patient can also be physiologically assessed. Psychometrics also can be positively influenced by virtual reality because the generated scenario can implement complex psychological measurement beyond self-reported measures that also includes non-verbal and physiological answers. It seems clear that the classical concept of test or even the adaptive testing can be subsumed inside the virtual reality. In consequence, Virtual reality can be another dream that comes true for psychological research; to create artificial experimental conditions but where it is difficult to distinguish between fiction and reality. It is highly probable that the virtual reality programmer will be one of the specialized job profiles of future methodologists and even a new academic position in our universities.

The real challenge of current science is not easy because it must stand firm in the defense of the principles of the scientific method, giving a preponderant place to replication and meta-analysis, and integrating the almost unlimited potential of big data and the virtual reality to research. All this must be possible without forgetting the need to demand of researchers an adequate level of statistical knowledge, free of magical thoughts about statistical techniques. And the challenge is also complicated because a strong relativism is installed in our social environment, where post-truth is favoring the appearance of dogmatic attitudes and away from the contrast of information. As a researcher and as a professor of researchers, despite the mentioned barriers, the opportunities to generate knowledge are so attractive that nothing can prevent the improvement of the quality of scientific evidence.

REFERENCES

Badenes, L., Frías, D., Monterde, H., & Pascual, M. (2015). Interpretation of the p value. A national survey study in academic psychologists from Spain. *Psicothema, 27*(3), 290-295. doi: 10.7334/psicothema2014.283

Chen, E., Wojcik, S. (2016). A Practical Guide to Big Data Research in Psychology. *Psychological Methods, 21*, 458-474. doi: 10.1037/met0000111.

Dumas-Mallet, E., Smith, A., Boraud, T., Gonon, F. (2017). Poor replication validity of biomedical association studies reported by newspapers. *PLoS One*. doi: 10.1371/journal.pone.0172650

Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current directions in psychological science, 5*, 161-171. doi: 10.1111/1467-8721.ep11512376

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 1-8. doi: 10.1126/science.aac4716

Sesé, A., Palmer, A. (2012). The current use of statistics in Clinical and Health Psychology under review. *Clinic & Health*, 23(1), 97-108. doi: 10.5093/c12012v-23n1a2