JERZY MARIAN BRZEZIŃSKI[1]

CONTEXTS OF EMPIRICAL RESEARCH IN THE SOCIAL SCIENCES

ABSTRACT. Jerzy Marian Brzeziński, *Contexts of empirical research in the social sciences*, edited by Z. Drozdowicz and S. Sztajer, "Człowiek i Społeczeństwo" vol. XLI, Poznań 2016, pp. 211-246, Adam Mickiewicz University Press. ISSN 0239-3271.

The article analyses "immersing" empirical research conducted in the field of social sciences (psychology, sociology, education etc.) in four contexts: psychological, ethical, cultural and the context of methodological awareness. It focuses in particular on the following issues, crucial for the condition of social sciences:

– research competences of those who undertake to design and conduct empirical research (the state of social and individual methodological awareness);

– specificity of psychological research carried out with the participation of human subjects – the perception of empirical research as an interaction "researcher – participant of the study";

– cultural specificity of the environment in which research takes place;

– attitude of researchers towards research participants (the state of ethical awareness);

– accepted patterns of disseminating research results (the policy of editors of scientific journals to approve only of such papers which contain statistically significant results i.e. with significance at the level of $p<0,05$;;

– assessment practices applied in institutions employing researchers: bibliometric assessment of academic units and the evaluation of individual scholars' research performance;

– succumbing to the pressure exerted by supervisors on researchers to publish "at all costs."

[1] Jerzy Marian Brzeziński, ordinary member of PAS (Polish Academy of Sciences); Professor of Psychology; AMU (Adam Mickiewicz University in Poznań) Institute of Psychology; ul. Szamarzewskiego 89AB, 60-568 Poznań, e-mail: brzezuma@amu.edu.pl

Introduction

In scientific disciplines such as psychology or sociology (classed in the broader field of the social sciences) the corpus of scientific knowledge is acknowledged as scientific – at a given stage of their development and within a defined paradigm (in the sense of Thomas S. Kuhn, 1996/2001) – when theorems aspiring for such methodological status, or let's say candidates for scientific theorems, are confronted beforehand with empirical data. And, needless to say, the kind of data and how such a "confrontation" takes place does matter.

Empirical data of the highest methodological quality derive from **experimentation** based on Fisher's **principle of randomisation**. This, for example, constitutes the methodological foundation – known as the "golden standard" – of medical trials exploring the efficacy of new drugs or other forms of medical intervention. And ideally this should be conducted in the conditions of a laboratory experiment, ensuring a high degree of control over independent variables (main, concomitant and confounding) and an appropriately high level of the dependent variable measurement. This fundamental condition of randomisation applies also to less methodologically sophisticated **field experiments**, which are mainly conducted by social psychologists and sociologists, but also by clinical psychologists. The latter, mainly active in the area of social practice (diagnostics, psychotherapy) thereby check empirically (which should be treated as the desired methodological state) the efficacy of psychotherapy in the *efficacy* type of research model (see Jaworska, 2006). In more general terms, clinicians consider research referring to experimentation based on the *principle of randomisation* as the most methodologically desirable model of empirical studies (in source literature empirical research of this type is called a *randomised clinical trial, RCT* – see APA presidential task force on evidence-based practice, 2006). A research model significantly weaker methodologically is the **correlation model** (see Brzeziński, 2016c; Nisbett, 2016 – parts 3 and 4) one willingly used by sociologists, social psychologists but also clinicians (studies of the *effectiveness* kind).

Empirical studies conducted by representatives of scientific disciplines classed within the field of the empirical sciences (from physics and chemistry to biology, medicine, psychology and sociology) are also subject to diverse external conditioning: economic (the percentage of GNP allocated by the government for scientific research and the development of higher education), political (promotion of defined scientific disciplines by the government), or

ideological-religious. However, these will not be analysed in this paper. We shall be focusing in this article on those determinants that, when not noticed sufficiently early (above all by the academic community itself) and when effective attempts at their elimination or at least minimisation are not made, will have a destructive impact on the development of science and – which I consider particularly threatening – on professional practice referring to its achievements (see Brzeziński, 2016b – and the comprehensive model described there: *Scientific Research and Professional Practice in Psychology*, SRPPP).

The determinants of interest to us here originate from the following:

– the research competencies of the persons undertaking the designing and conducting of empirical research (the state of social and individual methodological awareness – understood as in: Brzeziński, 2012a, 2016b);

– the psychological peculiarities of scientific research conducted with the participation of people – a look at empirical research as at the interaction of: "researcher – participant in the research" (see Brzeziński, 2016c, chapter 4: *Wewnętrzne determinanty procesu badawczego (II) – badanie psychologiczne jako interakcja „badacz – osoba badana"*);

– the cultural peculiarities of the environment in which the research is conducted (see American Psychological Association, 2008);

– the researchers' attitude towards the people participating in the scientific research (state of ethical awareness – see Brzeziński, Chyrowicz, Poznaniak and Toeplitz-Winiewska, 2008);

– the adopted models of publicising research findings (the policy of scientific journal publishers only accepting articles speaking of statistically significant results, i.e. achieving a level of significance of: $p = 0.05$ – see e.g. Chambers, Feredoes, Muthukumaraswamy and Etchells, 2014);

– the application of evaluative practices in the institutions employing researchers: parametric assessment of the research institutions and appraisals of individual researchers' achievements (Brzeziński, 2015a, 2016a);

– how researchers succumb to the pressure to publish "whatever the cost."

This paper focuses on analysis of the above determinants.

1. How the empirical research process is grasped in the social sciences – the five phases of scientific research

Research studies conducted in disciplines classed among the empirical sciences (we are only concerned about such here) are carried out in a certain

standardised manner. As such it has become the norm to write in source literature (of a methodological profile) about the **structure of the research process** (e.g. Brzeziński, 20016c; Bunge, 1967; Franfort-Nachmias and Nachmias, 2001) – the process of the scientific research divided into smaller, homogeneous segments: stages, phases and links. My latest view of the structure of the research process in the social sciences (and in particular in psychology) is presented in Fig. 1. I shall now discuss this in brief.

One can isolate the following five phases in the research process, phases which are homogeneous in terms of content. The following is their necessarily brief description.

1.1. Phase I: problem, hypothesis, variables

Research begins with the formulation of a question regarding the determinants of a defined dependant variable, Y. On the whole, the researcher – Inspired by scientific achievements already documented (known are both the empirical theories and the data lending them credibility, obtained through empirical studies conducted in keeping with current methodological standards accepted by the researcher community) – searches for hypothetical independent variables, X, as those that may justifiably stand as determinants for variable Y. To put it into "technical" language, it is a matter of explaining the variance of the dependent variable through its origins – documented in empirical research controlled by the researcher – which are the variables X forming the picture (constructed by the researcher) of the space of variables significant for the variable Y in question.

If the "grammar" of the language of the social sciences is a set of methodological rules constituting social methodological awareness, then the variables defined by researchers form its "dictionary And the "sentences" constructed in this language enter ("for now" – see Krajewski, 1998, p. 93) the corpus of scientific knowledge of the given empirical discipline (e.g. psychology) when they acquire acceptable empirical evidence. Therefore, both the research problem and the hypothesis as the most sensible response (based on the researcher's current knowledge) to this question are formulated in the language of variables. And the variables are defined on the basis of empirical theories accepted by the researcher community.

From the variables selected and defined in the language of a particular empirical theory the researcher formulates hypotheses that are the most probable – based on his or her knowledge – answers to the questions posed at the beginning of the research (the problems). These in turn aspire – but

after being empirically verified and entered into social circulation – for scientific theorems. Confronted with the results of "external" empirical research studies (among other things thanks to the procedure of research replication), as well as with the state of professional practice (concerning its effectiveness) constructed under these studies' influence, they either maintain their scientific status or succumb to transformation, or are rejected in favour of better explanations of the variance of the dependant variable.

Some years back Hans Reichenbach (1938) wrote of two contexts of scientific research: the **context of discovery** and the **context of justification**. The first phase of scientific research – according to the diagram presented in Fig. 1 – may be identified with the first context. It is in this phase that the researcher has research ideas of some degree of originality, which he or she must then subject to empirical control. This in turn takes place in the second context, which is of a more "craft-like" character. By no means does this have to mean that the context of justification does not have a creative or "discovery-like" character. In keeping with this concept, these contexts should be treated separately. In particular analysis of the context of discovery would be the domain of psychology and sociology, while analysis of the context of justification would be the domain of the philosophy of science, of methodology. However, years later it was demonstrated that this dichotomy of contexts is unsustainable, and currently the thesis of their oneness is accepted. To put things very briefly, one cannot indicate at which "place" the first of these contexts comes to an end and the second begins. When performing research actions aimed at checking empirically some kind of hypothesis we are also making certain discoveries that may lead to the emergence of new hypotheses, etc. These contexts permeate one another. The specifics of a hypothesis mean that in order to effectively check it, one has to sometimes step beyond the run-of-the-mill schemata of the research procedure and design innovative methodological solutions. And in this sense these two contexts are interlaced.

1.2. Phase II: operationalisation of variables

Empirical research cannot be conducted if the variables – which have the methodological status of **theoretical terms** – are not granted **empirical sense**. This is a very important phase of scientific research in which the researcher, guided by a specified empirical theory (e.g. psychological) on the basis of which the variables were defined, assigns the human research subjects values accessible through observation. By proceeding thus, he or

she may, for example, translate the theoretical variable "intelligence" defined in the language of the given theory of intelligence into the language of metrological operations. This will enable the assigning of corresponding numerical values to the research subjects. For example, these could be *intelligence quotient* (IQ) values (as understood by David Wechsler) obtained using a psychometric tool such as *D. Wechsler's Intelligence Scales* (more regarding the genesis of the IQ measurement and contemporary solutions in this respect in: Brzeziński, 2015b).

The procedure itself of **operationalisation** has undergone a long evolution: from the simple *operationism* expounded in the classic work of the physicist Percy W. Bridgman (Bridgman, 1927), via the important work of Rudolf Carnap (Carnap, 1959) and Jerzy Kmita's conception (Kmita, 1973) of introducing theoretical terms into the empirical sciences, to the most complete solution to this problem proposed by Elżbieta Hornowska (Hornowska, 1989), referring to Leszek Nowak's idealising theory of science (Nowak, 1974; Brzeziński, 1976, 1978).

1.3. Phase III: research plan; conducting the research

A hypothesis worded in the language of operationalised variables is subject to **empirical verification** (positive: confirmation, or negative: falsification), which – such has research practice shaped over the decades – is conducted, let us recall, as experimental research or as correlational research. This phase of the research process embraces not only the devising of an empirical research plan (sometimes quite complex), but also the method of recruiting people, and – in the case of experimental research studies – their distribution into sub-groups, which will be subjected to experimental manipulation in various ways. From the point of view of possible future external replication (conducted by other researchers), it is important to describe the successive steps that the researcher must take in order to conduct given empirical research successfully, and – should such a need arise – to repeat it.

1.4. Phase IV: quantitative data analysis ↦ statistical conclusion

Contemporary planning for empirical research (see Phase III of the research process) also entails the planning of statistical analysis of the data, that will be carried out in order to confirm or reject the research hypothesis (see Brzeziński, 2012b). Wording it differently, the researcher must, within the accepted statistical paradigm, choose the statistical model within which he or

she will, using the tools of statistics, assess the hypothesis. A paradigm quite widely adopted in the social sciences is NHST (*Null Hypothesis Significance Testing*). To elaborate, this will be: in the case of experimental research, the ANOVA/MANOVA statistical model (see Kirk, 1995; Winer, Brown and Michels, 1991), and in the case of correlational research the regression/correlation model (see Pedhazur, 1997; Cohen, Cohen, West and Aiken, 2003).

Realising the importance of this phase of empirical research, publishers of the world's more important scientific journals drew up recommendations for authors of articles containing reports on research conducted. And so, for example, a few years ago the *American Psychological Association* (APA) appointed specialist working groups which compiled a set of recommendations for writers publishing in psychological journals – see Wilkinson L. & Task Force on Statistical Inference (1999), APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008), American Psychological Association (2010a), Harasimczuk and Cieciuch (2012). These standards change over the years, become more exacting. And so, for example, it is impossible to publish – in a good scientific journal in the field of psychology – an article presenting the results of empirical research when the author only provides the values of the test statistics and the achieved level of statistical significance, *p*. "Today" one also has to give the value of the **effect size** (see Wilkinson L. & Task Force on Statistical Inference; 1999; American Psychological Association, 2010a).

The result of an intersubjective procedure verifying a hypothesis (within the NHST paradigm) leads to the formulation of a **statistical conclusion**. The said verification is conducted with the assistance of the best and consciously selected (!) statistical tools – **statistical significance tests**, such as the chi-square test, the Student's t-test, the F-test (ANOVA), Wilks' lambda (MANOVA), Pillai's trace (MANOVA), Hotelling-Lawley's trace (MANOVA), or Roy's largest root (MANOVA).

At this point I would like to draw attention (referring also to Phase V) to an important distinction. According to Bruce M. King and Edward W. Minium (King and Minium, 2009, p. 26), one should distinguish the **statistical conclusion** from the **research conclusion**. The former refers to the content of the **statistical question**, which – as the authors cited above write – "[...] is a question about a numerical aspect of the observations" (King and Minium, 2009, p. 26). Applying, for example, the aptly chosen Student's t-test for testing a null hypothesis, $H_0: \mu_1 = \mu_2$ in opposition to an alternative hypothesis, $H_A: \mu_1 \neq \mu_2$, then one could reach the conclusion that the average result in group 1 (experimental) is significantly different to that occurring in group 2

(the control group). And the recorded difference is sufficiently large for it to be unlikely (the threshold usually adopted here is the probability value of $p = 0.05$) to have emerged accidentally (although this cannot be ruled out).

1.5. Phase V: interpretation of the research result ↦ research conclusion; generalising the research result

To return to the distinction given above, a statistical conclusion should not – although it sometimes does – be the end to the entire research procedure. This is because it constitutes a characteristic foundation for the research conclusion. And the **research conclusion** (King and Minium, 2009, p. 26):

> [...] is a conclusion about the subject matter. [...] Although the research conclusion derives partly from the statistical conclusion, we see that other factors must be considered. The investigator, therefore, must weigh both the statistical conclusion [but he may not confine himself to just this – J.B.] and the adequacy of the research design"

(which is distinctly emphasised in: Rosenthal, 1996). Let us note that the researcher, interpreting the results obtained, refers to empirical theory in the language in which he or she defined the variables under investigation.

Research studies in which people participate as the researched "objects" are far more complicated than those conducted by biologists, chemists or engineers. And the determinant of this complexity is not costly and technologically highly complex apparatus – characteristic of research in the exact sciences, in the life sciences, in biology or engineering – but what psychologists call the "individual differences" occurring between people. Not always can the results of research conducted on students (a population that psychologist-researchers choose willingly and too often) be simply generalised to embrace other populations. Sometimes the conditions in which laboratory experimental research is conducted are so far removed from non-laboratory conditions, that generalising them to embrace other populations and other conditions, without considering the impact on the dependant variable of variables beyond the researcher's control such as living conditions, preferences, tiredness, interests, occupational experience, and so on, distorts the conclusions. It alters their **external validity** (see Brzeziński, 2004). The researcher must also answer the question regarding the "durability" of the results obtained. This is about how long the results remain valid for after the completion of the research, before becoming outdated. Only thorough discussion of the possible impact on the depend-

ant variable of the confounding variables and concomitant variables not controlled by the researcher will allow the latter to draw up a cautious and reliable research conclusion – based on the statistical conclusion – and to generalise the research result.

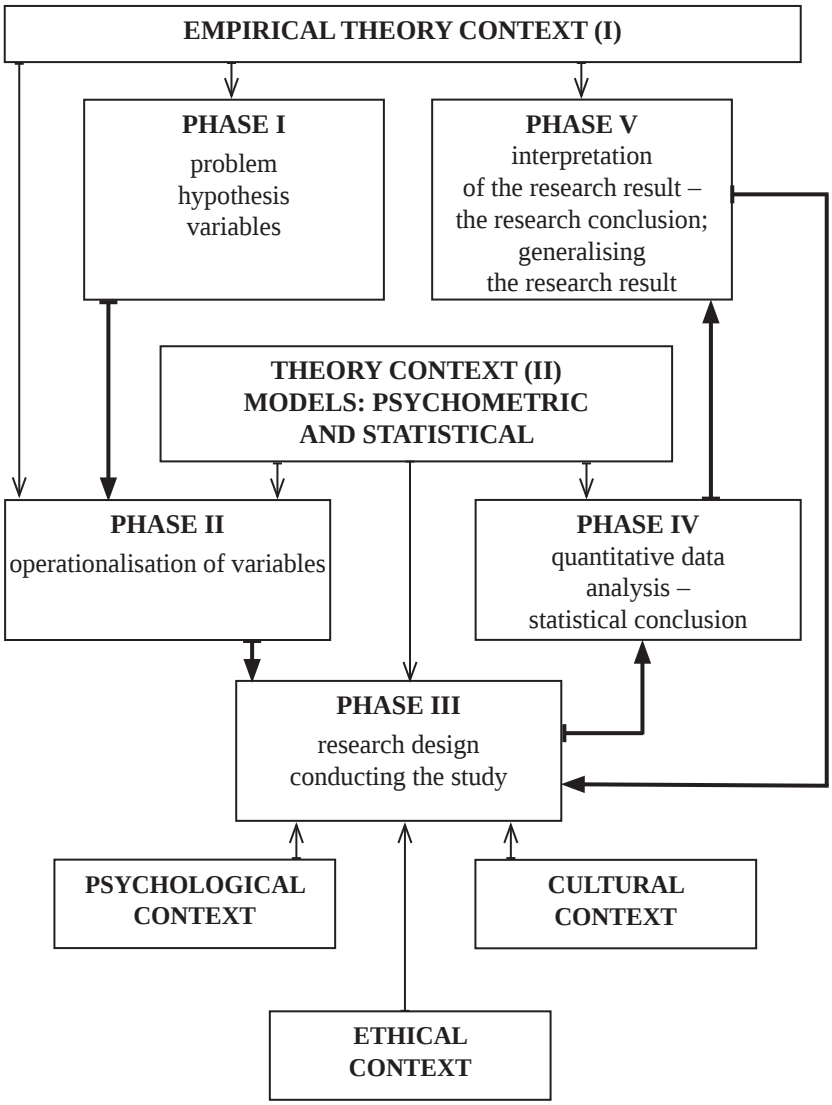

Fig. 1

2. Immersion of empirical research in four contexts: methodological, psychological, cultural and ethical

I shall now move on to discussing the significance possessed by the contexts listed above for understanding the determinants of the research process and the "fates" of the results obtained in the empirical social sciences. Fig. 2 shows their impact on research practice, and – indirectly – on social practice.

2.1. The context of methodological awareness – social and individual

When discussing the different phases of empirical research I indicated at the same time the determinants that it is subject to. These are illustrated in Fig. 1. As we can see, there are **two theoretical contexts – I and II** – in the foreground. They, in fact, determine the originality and methodological precision of a given empirical research study.

The first is the **context of the empirical theory**, because it is in this theory's language that the variables from which the researcher formulates the problems and hypotheses are defined. In psychology – though not only – which in the process of operationalisation of the variables refers to the specific metrological tools that psychological tests constitute, the role of the theory is accentuated. This theory, let's say for example psychological, constitutes scientific justification for the constructing of a psychological test applied in the phase of operationalisation of the variables. This connection between the test and theory was tackled in a work of breakthrough importance for psychological research practice (especially diagnostic) from 1955: *Construct validity in psychological tests* by Paul E. Meehl and Lee J. Cronbach (Cronbach and Meehl, 1955; Smith, 2005). The basic feature of a test is its **construct validity**. It is psychological empirical theory that makes any psychological test scientifically credible. The researcher formulates the research conclusion in its language. And ultimately the research conclusion is phrased in the theory's language.

The second is the context of two models: the **psychometric model** and the **statistical model**.

Let us note that a psychological test is not a random patchwork of tasks, questions or statements, but is constructed in keeping with the requirements of a defined **psychometric model**. Such a model – accepted by the psychologist community for several decades now – is that presented in its fullest form in

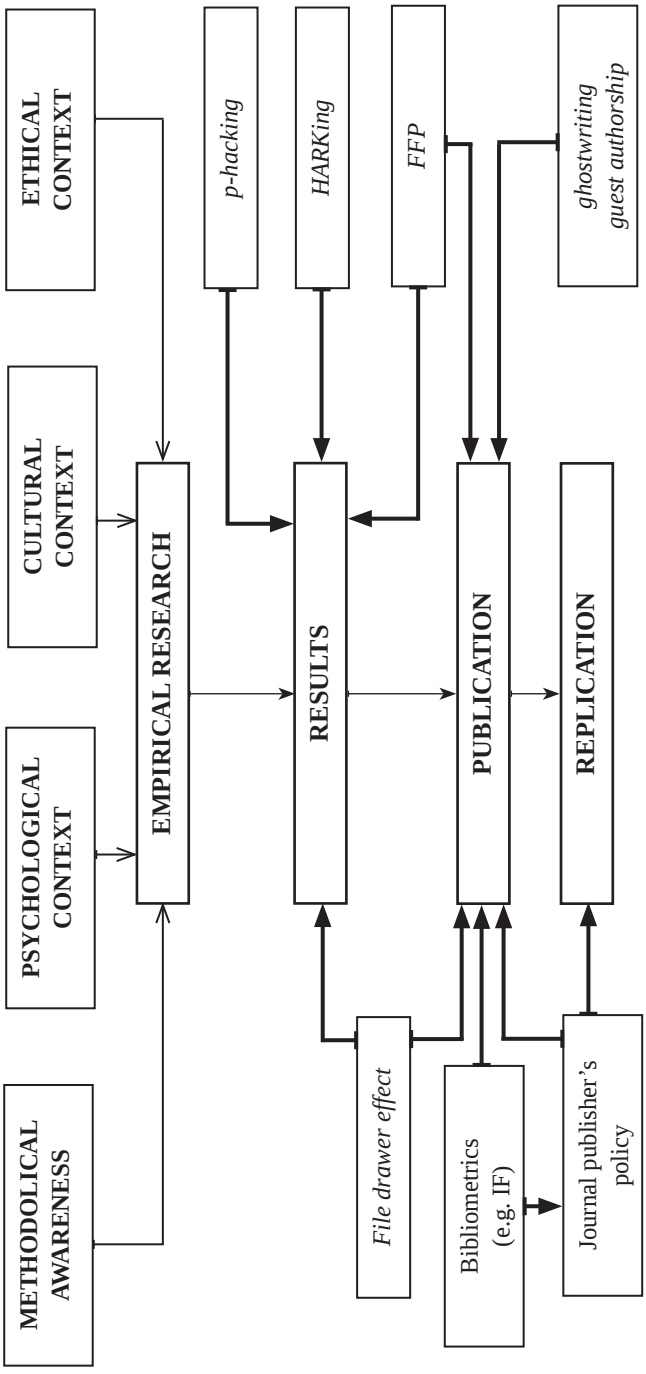
ETHICAL CONTEXT
CULTURAL CONTEXT
PSYCHOLOGICAL CONTEXT
METHODOLICAL AWARENESS
EMPIRICAL RESEARCH
p-hacking
HARKing
FFP
ghostwriting guest authorship
RESULTS
PUBLICATION
REPLICATION
File drawer effect
Bibliometrics (e.g. IF)
Journal publisher's policy

Fig. 2

the fundamental monography presenting *true score theory*, by Harold Gulliksen: *Theory of mental tests* (Gulliksen, 1950). This so-called "classical test theory" was later added to and developed by Frederic M. Lord and Melvin R. Novick in their monography: *Statistical theories of mental test scores* (Lord and Novick, 1968). The latest generation of test theories is being developed today: *Item Response Theory* IRT (see Embretson and Reise, 2000), also called "modern mental test theory." This subject-matter is so important for the educating of future psychologists that the program of master degree studies in psychology now has a separate subject entitled *Psychometry*.

Both the planning of empirical research (whether experimental or correlational) and the quantitative analysis of findings require familiarity with statistical models. The researcher not only describes the results obtained in the language of descriptive statistics, but also – for example within the framework of the NHST paradigm mentioned above – draws statistical conclusions regarding the empirically verified hypotheses. He or she also refers – when comparing the results of numerous studies conducted by different researchers on the same topic – to the more advanced method of analysis that is **meta-analysis** (see e.g. Kleka, 2011).

The two contexts, of the empirical theory and of the psychometric and statistical models, make up the more general **context of methodological awareness**. Methodological awareness is understood as a set of methodological rules and recommendations permitting – at a given stage of development of a field of science (here: the social sciences) – a defined shape of research practice that is realised in the form of research steps taken by researchers during the course of the research process.

The structure of the methodological awareness embraces empirical theories developed by generations of researchers and assigned to defined paradigms. It is in the language of a specific theory – as has been demonstrated above – that the researcher defines the variables and from which he or she draws up the problems and hypotheses. The empirical theory chosen by the researcher (or constructed by him or her from scratch) exerts a decisive impact on the entire research procedure. In particular – and in its language – the researcher interprets the research result and carries out its generalisation. In order to be able to conduct empirical research, the researcher must take a very important step: he or she must lend the theoretical variables (defined in the language of the theory concerned) empirical sense. This is achieved in the procedure of operationalisation of the variables. This (which is quite a common practice) refers to the psychometric models mentioned above. The results gathered after conducting empirical research planned according to

a defined standard, experimental or correlational, are subjected to statistical analysis. In order to verify the hypotheses, researchers refer to a defined statistical model.

One can distinguish a **social** and an **individual methodological awareness**. The latter may, sometimes significantly so, deviate from the former. And in particular – without guaranteeing that a given researcher will conduct the empirical research correctly.

I shall now move on to discussing the significance that the other three contexts have for understanding the determinants of the research process and the "fates" of the research results obtained in the empirical social sciences. Fig. 2 shows their impact on research practice.

2.2. Psychological context

The psychologist Saul Rosenzweig (Rosenzweig, 1933) already drew attention to the peculiarities of experimental psychological research (and also, in my opinion, of sociological or pedagogical research) back in the nineteen-thirties:

– the researcher becomes an element of the research situation,

– variables linked to and characterising the research subject, such as personality and motivation, and so on, influence the behaviour of this person in the research situation,

– interaction occurs during the research: "researcher – research subject."

A researcher conducting scientific research within the social sciences – and especially when, as with a clinical psychologist, sociologist or guidance counsellor at school, there is "face-to-face" contact with the research subjects – must take into account the fact that their subjects have some kind of perception (and whether correct or not is of no importance) regarding the purpose of the research in which they are participating, that they expect some kind of compensation for their involvement in the study, are concerned about the researcher being in possession of information disadvantageous for them, about the research conditions generating a certain level of discomfort, and are afraid of refusing to take part in the research, and so on. The physicist or biologist does not experience such "psychological" problems. Moreover, as numerous studies have gone to show, the researcher conveys his or her expectations regarding the result of the research to the research subject, thereby moulding their behaviour during the actual research. This gives rise to the so-called *interpersonal expectations effect*, also known as the *Rosenthal effect* (see Blanck, 1993; Trusz, 2013).

Martin T. Orne (1962/1991, 1973/1993) drew attention to the active role played in experimental research by the research subject. This person is capable of identifying the real purpose of a study despite it being concealed by the researcher or the usage of instructions masking that purpose (*deception*). The subject is guided by various indications provided by the researcher (who he or she is, what institution he or she represents), instructions given to the research subject, tools used by the researcher (e.g. psychological tests), or the research plan, and so on. M.T. Orne spoke of indications (variables) suggesting the research hypothesis (*demand characteristic of experimental situation*) enabling the research subject to decode the actual research goals, and not those misleading him or her. The masking measures applied in well-known experiments conducted by Stanley Millgram (see Millgram, 1963, 1965, 1974) or Solomon E. Asch (see Asch, 1951/2001) constitute a good illustration, although it must be pointed out that those particular masking measures used by the researchers cited above evoked serious ethical reservations.

Milton Rosenberg (1969/1991) in turn drew psychologist-experimenters' attention to the role of the variable that he called the variable of evaluation apprehension. In his opinion the research subject quite frequently treats participation in a scientific study as meaning the possibility of disclosure of information about him- or herself that he or she would prefer to remain unpublished. In addition, this person becomes alert and distrustful, and strives to identify the researcher's "real" intentions. Research subjects, depending on whether they consider spontaneous behaviour in the situation of a scientific study as threatening or not threatening their self-appraisal (sometimes inappropriately inflated) will either collaborate with the researcher or refuse to cooperate and modify their behaviour accordingly.

Not noticing or even outright ignoring the psychological context of the research, not controlling the variables described by M.T. Orne and M. Rosenberg – on the methodological status of the confounding variables – as well as ignoring the *interpersonal expectations effect* leads to the researcher describing not facts but *artefacts* (see Rosenthal and Rosnow, 2009).

2.3. The cultural context

In such social disciplines as psychology more attention has begun to be assigned to the cultural framework in which the research project is being carried out. As the authors of a report compiled in 2008 noted, the findings

of scientific research studies in psychology were obtained: "[...] upon Anglo Western middle-class, Eurocentric perspectives and assumptions" (American Psychological Association, 2008, p. 2).

The world, including that in which academics live and conduct their scientific research, is global in character. It does not comprise purely persons sharing European or American values. If, for example, psychology is to provide theories and research findings that can be relevantly transferred between groups differing in terms of culture, then it cannot turn a blind eye to the cultural context. And this dictate applies also to psychology understood as an occupation.

This requirement was noticed by, among others, the American Psychological Association, which compiled an important report on the topic: *Report of the task force on the implementation of the* multicultural guidelines (cf. American Psychological Association, 2008). Six multicultural guidelines form the lynchpin of the report (p. 6):

> Guideline 1: Psychologists are encouraged to recognize that, as cultural beings, they may hold attitudes and beliefs that can detrimentally influence their perceptions of and interactions with individuals who are ethnically and racially different from themselves.
>
> Guideline 2: Psychologists are encouraged to recognize the importance of multicultural sensitivity/responsiveness to, knowledge of, and understanding about ethnically and racially different individuals.
>
> Guideline 3: As educators, psychologists are encouraged to employ the constructs of multiculturalism and diversity in psychological education.
>
> Guideline 4: Culturally sensitive psychological researchers are encouraged to recognize the importance of conducting culture-centered and ethical psychological research among persons from ethnic, linguistic, and racial minority backgrounds.
> Guideline 5: Psychologists are encouraged to apply culturally appropriate skills in clinical and other applied psychological practices.
>
> Guideline 6: Psychologists are encouraged to use organizational change processes to support culturally informed organizational (policy) development and practices.

A new approach was proposed very recently (Hardin, Robitschek, Flores, Navarro and Ashton, 2014) for taking into consideration the cultural factor when analysing the relevance of the psychological theory. The situation is similar in regard to psychological tests. I believe that in addition to the traditional options for confirming **test relevance** (cf. American Educational Research Association, American Psychological Association, National

Council on Measurement in Education, 2014, pp. 13-22) one more may be added: the **models and demands of the culture** in which research subjects undergoing a specific psychological test live.

In bringing this necessarily brief description of the cultural context to a close, I shall cite one more fragment from a report by the APA (American Psychological Association, 2003, p. 390):

In analyzing and interpreting their data, culturally sensitive psychological researchers are encouraged to consider cultural hypotheses as possible explanations for their findings, to examine moderator effects, and to use statistical procedures to examine cultural variables (Quintana, Troyano and Taylor, 2001).

2.4. The ethical context (the researcher's conduct towards the research subjects)

Researchers who conduct scientific research with human participation must respect limitations (apart from legal, quite naturally) of an ethical nature. If we were compare to the 30s or 40s of the previous century, then we would have no difficulty noticing the increase in **ethical awareness** (via analogy to methodological awareness) not only among researchers, but also in universities offering, for example, studies in psychology. In the social sciences these limitations are most visible in the research practice of psychologists and in the training of future psychologists. Ignoring here the ethical context of the professional activity of psychologists (diagnostic, psychotherapeutic – see Brzeziński, 2015c, 2016d; Brzeziński, Chyrowicz, Poznaniak and Toeplitz-Winiewska, 2008) I would like to draw attention to the fact that the "codes of ethics" adopted by professional psychological organisations dedicate separate points to the topic of the ethicality of scientific research (e.g. the American Psychological Association, 2010b; Polskie Towarzystwo Psychologiczne, 1991; British Psychological Society, 1985; International Union of Psychological Science, 2008). Whereas abuse by professionals in their relations with clients (e.g. between psychotherapist or sexologist and patient) are universally condemned, and not only by other professionals but also by society, the combining of ethics and the methodological quality of scientific research is not so common. Here, however, I stand by the ethically sound position that there is no good (including in the ethical sense) psychological practice without it being supported with results from empirical research conducted properly in terms of methodology (see also Spendel, 2016). And such research must also respect the rights of the research sub-

jects. This is about the researcher (e.g. psychologist or sociologist) seeing in the participant of empirical research a person with naturally assigned rights, as embraced in codes of ethics. As an example one could point to the codes – which in my opinion are very good – drawn up by the American Psychological Association (2010) and the International Union of Psychological Science (2008). These oblige researchers to respect principles that I would consider fundamental for a researcher to function correctly in the world of contemporary science (see also Brzeziński, 2015c; Brzeziński and Toeplitz-Winiewska, 2015).

In the forefront we have respect for **human rights and dignity**. There is no justification for violating these rights in the name of a falsely understood superiority of achieving social goals over the rights assigned, naturally as it were, to the research subjects. Let us mention here the *Charter of Fundamental Rights of the European Union* (KPPUE, 2000); in Article 1, "Human dignity," it states: "Human dignity is inviolable. It must be respected and protected." And that requires no further comment.

The next principle, but more detailed than that indicated above, is that of **informed consent**, which happens to be embraced in every code. The recommendation here is unambiguous: participation in scientific research may take place only (!) with the consciously given consent of a person participating in it. This also means that this person understands the situation they are in after giving such consent, and accepts the consequences of participation in the scientific research in question! Of course incapacitated persons and children are excluded from this principle (where the consent of their legal guardians or parents is required).

And there are three other principles that should also be pointed out: **confidentiality**, **privacy** and **anonymity**. In keeping with these principles, the researcher does not "share" information obtained from a research subject (especially when this information constitutes so-called "sensitive data" about this person's private life, sexual preferences and beliefs, and so on). When gathering information from a research subject the researcher does not step outside of the area agree upon with this person beforehand, and in particular does not perform in relation to the subject as a psychotherapist or a "confessor." To word it briefly, the researcher respects the privacy of the research subject. When the researcher informs this person (e.g. via the Internet) that the research is anonymous in character, he or she does not attempt by deceit to acquire data enabling the subject's identification.

3. Institutionalised science

Today one rarely meets researchers conducting their activities outside of an institution of some kind, be it a higher place of learning, a research institute, university department or museum, etc. Researchers are squeezed into the corset of regulations that have not always been properly thought through, and which impose – while blocking authentic creativity – the way in which they proceed with results obtained from empirical research. By applying these regulations (e.g. regarding the parametric appraisal of research institutes) the persons responsible for the scientific condition of a research institute employing the researcher command him or her to publish in specific journals and not in others, and not to publish chapters in collective works as this is reflected in a low number of points, etc. In effect, researchers up against the wall or less resistant to such pressure, feeling desperate, take steps that qualify as "scientific misconduct" (Grabski, 2015). And the sub-points below deal with these unethical activities.

3.1. In pursuit of points

It is, of course, all about the points granted to researchers (but also, after they are added up, to the research units that employ the researchers laboriously gathering these points), which constitute the basis for assessment of their scientific activeness, and which may consequentially determine their employment by a specific research institute, academic and professional promotion, and whether or not they obtain a grant. The very elaborate system of periodic parametric evaluation of research units in Poland (more elaborate than is reasonable) is based precisely on granting points for researchers' various research products, and above all for their publications. Such reliance on points accumulated by a research institute (for publications, for patents, for promotional dissertations, etc.) is sometimes referred to ironically as "pointosis" or "pointomania."

I believe (and I am not alone in this view – see Parnas, 2007; DORA, 2012; FNP, 2014; Nezlek, 2014; Towpik, 2015) that one cannot conduct a reliable evaluation, either of individual researchers or of research units, that is based solely or overwhelmingly on such bibliometric indicators as the Impact Factor (IF), Total Impact Factor (TIF), or – as happens in Poland – scoring for the scientific value of journals carried out using ministerial lists

of journals: A, B and C (see Brzeziński, 2015a, 2016a). As such, one has to agree with John B. Nezlek (see Nezlek, 2014, p. 595), that:

> Increasingly, raw bibliometric data are replacing judgment. One does not need to read a scholar's articles to evaluate what he or she has done. Simply calculating the *h*-index suffices. [...] **Bibliometric indices can and perhaps should be part of evaluation of scholarship, but they need to be part of evaluations; they cannot be the evaluations themselves** [highlighting – J.B.].

The evaluation procedure – ignoring *peer review* type assessments of one's scientific achievements – leads to serious anomalies. Appraisals expressed – and let's highlight this – using bibliometric scores by no means guarantee their accuracy and objectiveness. In fact, they become but a surrogate of a correctly conducted *peer review*. As if that were not enough, mediocre researchers and "ambitious" managers of research units yielding to the pressure of diverse factors (e.g. pressure from superiors to quickly complete a post-doctoral dissertation or the drive for a research institute to achieve a high category in its parametric assessment) resort to various types of abuse aimed at acquiring the largest possible number of points. More about this in the point that follows, 3.2.

3.2. In pursuit of "success" – plagiarism, ghostwriting, guest authorship, data manipulation via HARKing and p-hacking, data fabrication and falsification

I shall begin, once again, by citing the opinion of J.B. Nezlek (see Nezlek, 2014, p. 596):

> I think that the current fascination with bibliometrics and the types of pathologies Prof. Brzeziński discusses stem (in part) from the adoption of business models in science and education. Scholars are pressured to produce more and more under the false impression that numbers are magical indicators of quality, and careers become pursuits of numbers rather than pursuits of knowledge. This emphasis means is helping transform publication from a manifestation of the pursuit of knowledge to an indicator of success. What was an end in and of itself is becoming a means to an end – success and a reputation. Moreover, as such norms and practices become more salient and commonplace, I think we attract more individuals who are more focused on success rather than knowledge, and that such people, like Stapel, are more likely to violate ethical norms because the ends (success) justify the means (lying).

The practice of **plagiarising** others' texts (either in whole or in fragments, word-for-word or with individual words replaced and a different sentence structure than in the original) has been known for many years. Another worrying development is the buying of texts written (on the whole for a fee) by anonymous "specialists This is a practice known as **ghostwriting**. Occasionally a researcher (who may be a good specialist in his or her field) provides a hired author with a set of data (correctly gathered methodologically), and this ghostwriter works on this material, referring to the statistical procedures applied – which are sometimes very technically sophisticated – and writes an article meeting the editorial standards of a particular journal. Naturally, only the person commissioning such writing is actually credited as the author; the true author on the other hand is that unmaterialized "ghost In its primitive form, the commissioning party buys a dissertation on the basis of a press announcement.

Also alarming is the practice of the adding of superiors or thesis supervisors to articles written by their subordinates – by their assistants, doctoral students, master's students and so on. This is known as **guest authorship**.

It is worth focusing on the manipulation of empirical data by researchers (or one should rather say pseudo-researchers) with the aim of obtaining the desired situation, which of course is the publication of an article with the said data. This in turn brings one closer to the cherished goal of achieving academic promotion or employment in some research institution or other. The progress that has been achieved over recent decades in developing advanced methods of multivariable statistical analysis of data and in the significant simplification in their handling by computer has meant that researchers using modern, fast computers and highly sophisticated statistical software yield to the temptation of "juggling" the data in order to obtain the intended effect: confirmation of the research hypotheses. However, by so doing, they are betraying the ethics of scientific research.

In this context it is essential that attention be brought to two particularly worrying phenomena, descriptions and analyses of which have appeared in the columns of specialist scientific journals. This refers to **HARKing** and **p-hacking**.

Before approaching the editors of a journal with an article constituting a report on empirical research that has been conducted, researchers first of all – using facilitations, unimaginable just a few decades ago, provided by modern computers and the statistical software packages installed on them, for example SPSS, SAS or STATA – carry out various statistical analyses on sets of gathered empirical data. What are these analyses? They could,

for example, be the testing of the significance of differences between such parameters as the average results of the dependant variable in comparative groups, or the testing of correlation significance between variables. The researcher then tailors the content of the hypothesis to the results obtained from such statistical analyses that demonstrate statistical significance of the differences or correlations. Results that lack statistical significance are omitted. Then in the article the researcher formulates only those "original" hypotheses for which he or she knows that confirmation has been achieved, and then describes the statistical analyses carried out. If we take another look (see Fig. 1) at the logic of the research process, we see that the researcher first of all checked which of the analyses carried out (sometimes they are numerous) led to statistically significant results, and then "dressed" them in words, suggesting that such were his or her hypotheses put forward before conducting the empirical research. Hence the name of this procedure: **HARKing** (from the words *Hypothesizing After the Results are Known*). The researcher in turn comes across as a "seasoned" and "sagacious" researcher, who propounds research hypotheses whose content comprises highly accurate predictions.

The second procedure is also related to the manipulation of data in order to obtain as a consequence statistically significant results that have statistical significance at a level required in the empirical sciences and – importantly – respected by the editors of scientific journals, that is a p-value of 0.05. The researcher informs only of the results from research that led, for example, to the obtaining of differences between comparative groups, whose arithmetic averages differed significantly with *p* equal to or less than 0.05. Such procedure by a researcher resembles the behaviour of an angler who is not aiming to catch a fish of a particular species, but simply wants to catch a fish, any fish. Just as long as he does not have to go home empty-handed. You could also imagine a hunter who goes hunting and behaves in a similar way. Hence the humorous descriptions used for such a strategy: *fishing-expedition* or *hunting-expedition*. In source literature such conduct is defined as: **p-hacking**.

This search "whatever the cost" for statistically significant results, meaning ones that with a p-value of 0.05 allow for rejection of a null hypothesis (H_0) about there being no difference between population parameters (e.g. average results for the dependant variable in the compared populations) or no correlation between the investigated variables in favour of the alternative hypothesis (H_A) saying that the said differences or correlations have occurred, sometimes becomes a genuine obsession among researchers – and

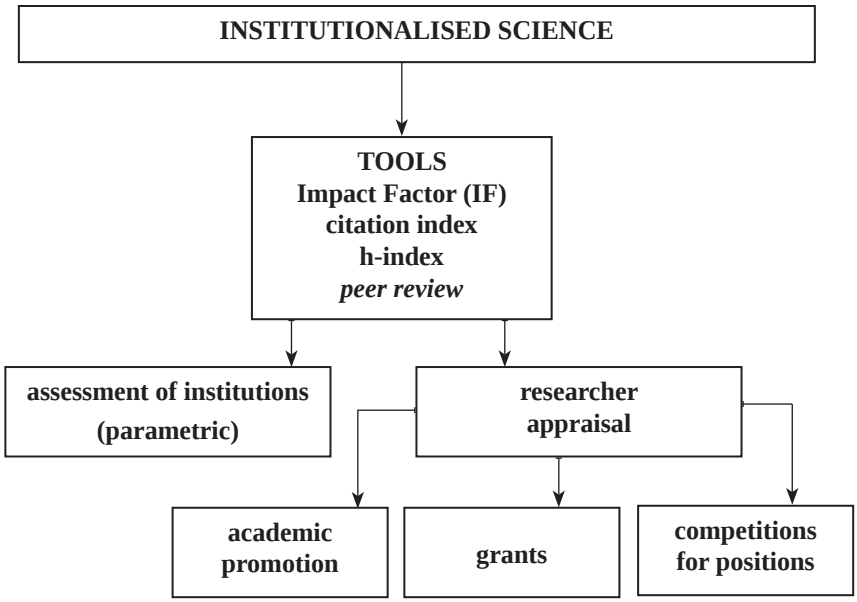

Fig. 3

especially among those harried by their superiors for lack of progress in their work and not providing the research institute employing them with parametrisation points, but also among those "ambitious" researchers dreaming of a rapid scientific career.

Let us conclude this recital of reprehensible procedures by drawing attention to primitive **falsification** through making changes to the values actually obtained from the measuring of variables or omitting data "deviating" from the assumed trend, and to the **fabrication** of data, meaning the making up of results from measurements that never were actually carried out.

The **fabrication** and **falsification** of data, plus **plagiarism**, known through the abbreviation **FFP**, are symptoms of the "scientific misconduct" mentioned earlier (see FPRM, 2002; Grabski, 2015, p.185). The main document on this topic defines this misconduct in research as follows (the highlighting is my own):

> Research misconduct is defined as fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results.
>
> – **Fabrication** is making up data or results and recording or reporting them.

– **Falsification** is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

– **Plagiarism** is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit.

– Research misconduct does not include honest error or differences of opinion.

Figure 3 (which should be "read" together with Fig. 2) shows what tools are used by contemporary institutionalised science for assessing the work of researchers and research institutes. The problem lies not in the fact that they are these specific tools and not others, but in the fact that they are used disproportionately while dismissing appraisals of the *peer review* type, and with a false conviction of the objectivization of appraisals made using these tools. This is especially the case in Poland.

4. Attempts at returning to normality

Following years of "rule" in institutionalised science – especially in Poland – of the bibliometric approach to assessing the research achievements of individual researchers and research institutes, referring to such indices as the IF and TIF in journals in which the researcher publishes, to the h-index or number of citations, an about-turn is taking place from this depersonalising evaluative practice in favour of a *peer review* approach (Brzeziński, 2015a, 2016a; Towpik, 2015; Fundacja na Rzecz Nauki Polskiej, 2014; DORA, 2012). Much more emphasis than ever before is being placed on the application of control procedures. I would like to devote more attention to three of these in point 4.2.

4.1. DORA

In my opinion we should acknowledge the importance of the *DORA Declaration* (DORA, 2012; Towpik, 2013), or in full the *San Francisco Declaration on Research Assessment. Putting science into the assessment of research* from 16 December 2012 (about which I wrote much more comprehensively in: Brzeziński, 2015a), which came into being in the community of cell biologists, the *American Society for Cell Biology* (the ASCB – a community I would call highly "parametricised") and a group of publishers of scientific journals. It was this *Declaration* that contained a distinct protest against the

mechanical application of bibliometric indices, including the impact factor (IF). Numerous global research institutions and researchers themselves, including a group of Nobel prize winners, put their signature to it. Among other things, the Declaration says the following:

> 1. Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions. [...]
> 18. Challenge research assessment practices that rely inappropriately on Journal Impact Factors and promote and teach best practice that focuses on the value and influence of specific research outputs.

I think that this is a step in the right direction.

4.2. Remedial actions

Provision of data. Reports of researchers applying unethical practices, ones that qualify as "questionable research practices" (QRPs) – something that seems to be snowballing, or at least much in recent years has been written of this on the pages of such esteemed journals as *Nature* and *Science* – are forcing the editing teams of self-respecting scientific journals to apply various techniques of self-defence. One of these, steadily gaining in popularity, is that obliging the authors of articles to provide the journal's editors with the raw data from the empirical research conducted. This provision of data (as writes Jachimczyk, 2015, p. 409), "[...] accumulated in research financed with public funds," is an acknowledgement of the *open access* movement in science. This idea is close to me, and I fully agree with Adam Jachimczyk's stance (Jachimczyk, 2015, pp. 410-411):

> For proponents of opening data the important things include lowering the costs of academic writing, since the availability of data online reduces the necessity for gathering the research material again [...]. The opening of data also contributes to an improvement in the quality of such writing, as it allows other scientists to conduct research based on the same material, which furthers the detection of misconduct in science and the discovery of cases of fabrication of research data [...].

However, not all researchers and not all publishers of the top international scientific journals are enthusiastic about this initiative. This is because they are afraid that somebody, at their "expense," will prepare a more inter-

esting paper, that they will have to share the contingent gains – including of a non-material nature – with others. In my opinion this problem could be solved by the persons who gathered the data having the exclusive right to use it for a certain length of time (one or two years), while researchers who use this data would have to cite the source in their publications.

Preregistration system. A large and immensely important part of the troubles related to HARKing and p-hacking could be solved through modification of editors' publishing policy – especially among editors running leading scientific journals in specific fields of science (for example those with high IF values). The thing is for magazines to open up to articles where "nothing" worked out for their authors, where the outcome of statistical significance of $p > 0.05$ was seen as a bad dream by the researcher. But why should we treat a p-value of 0.05 as "holy" (Skipper, Guenther and Nass, 1970: "the sacredness of .05")? The consequence of such a conservative stance is quite widely known as the **file drawer effect** and **publication bias**. In this context, worth noting is that *meta-analysis*, the method of modern review of source literature used to find empirical confirmation (and not falsification) for a specific hypothesis, particularly popular in the social empirical sciences (especially psychology), is seriously burdened by the *file drawer effect*. This is so because those scientific journals constituting the source material for such meta-analysis do not, on the whole, publish articles about a sought effect not occurring, or with borderline effects (insignificant values of the *effect size* scores). As a consequence of this *publication burden* (cf. Franco, Malhotra and Simonovits, 2014; Song, Hooper and Loke, 2013) **the meta-analysis results are overestimated**. How greatly? That we do not no. Publishers would have to modify their policy, and want also to publish "negative" articles.

In my opinion **"preregistration"** is a method heading in the right direction, although more time is needed for us to be sure. This is where the researcher announces a planned empirical research project to a given journal, one which following peer review would be provisionally accepted, while the author could then commence the research (cf. Chambers and Munafo, 2013; NeuroChambers, 2013). According to Chris Chambers and Marcus Munafo, as well as several dozen researchers who signed a special appeal in this matter (Chambers and Munafo, 2013):

> [...] unlike traditional scientific publishing, in which manuscripts are peer reviewed only after studies have been completed, registered reports are reviewed before scientists collect data. If the scientific question and methods

> are deemed sound, the authors are then offered "in-principle acceptance" of their article, which virtually guarantees publication regardless of how the results turn out.

It is precisely this method that is meant to be an effective method of fighting the practice of HARKing (cf. Kerr, 1998) or p-hacking (for example Lombrozo, 2014; Scot, 2013), as there must be concern regarding the results of a study into the incidence of QRP (*Questionable Research Practices*) in the psychologists' community (cf. Leslie, Loewenstein and Prelec, 2012) indicating a significantly large percentage of psychologists referring to unethical HARK research practices.

Fascination with this new "format" of conducting empirical studies will wear away in time. And I also believe that it will not succeed in supplanting current publishing practices. Knowing the publishers' concern regarding revenue from publications in "protected" journals, and we are not talking small-scale here (see Towpik, 2015), we can rest assured that the traditional style of publishing will continue. As for those who will want to achieve "success," if only sufficiently motivated to cheat, then their conduct will be dishonest. Such is the dark side of this promising publishing procedure.

Replications. Let us pose a question of fundamental significance for the social empirical sciences: What is it that really allows scientific knowledge (that generated as an effect of applying procedures and methods agreed upon in the academic community, in the studios of researchers – for example psychologists) to be distinguished from non-scientific knowledge? Knowledge that aspires to be called scientific is expected to be **intersubjective** (see Ajdukiewicz, 1983; Frankfort-Nachmias and Nachmias, 2001) – or more precisely, that it will be consistent with Kazimierz Ajdukiewicz's weaker principle of rationality: that it will be *intersubjectively communicable* and *intersubjectively verifiable*. This *intersubjectivity* is a quality of scientific knowledge that means it can be recreated by other researchers when they want to repeat an original scientific study.

Researchers and representatives of diverse empirical disciplines talk of the requirement for the **replication** of research studies. This problem of the replicability of research – in regard to studies conducted by psychologists – has, especially in recent years, become a problem raised not only in the columns of psychological journals. The replication of 100 studies conducted by a 270-person team under the supervision of Brian Nosek, a social psychologist at the University of Virginia, was a very extensive program of this kind. Interestingly, the results of these studies, conducted in the open

science system, were published in the prestigious weekly *Science* (Open Science Collaboration, 2015).

What did B. Nosek and his collaborators do that was so important and so disconcerting for psychologists? His team of 270 well-prepared researchers selected (using a thoroughly elaborated procedural algorithm) 100 articles out of 488 published in 2008, in three psychological journals: *Psychological Science* (PSCI), *Journal of Personality and Social Psychology* (JPSP), *Journal of Experimental Psychology: Learning, Memory and Cognition* (JEP: LMC), and repeated (replicated) the research described in them. 32 (of 55) articles were published in the JPSP, 28 (of 39) were selected from JEP: LMC, and 39 (of 64) from PSCI. 2 articles each covered 2 replications. As for the thematic scope of the research, it embraced 43 studies of a cognitive profile, and 57 of a social-personality profile. The findings of the replicated studies were subjected to comprehensive statistical analysis, with reference to tests of statistically significant differences, confidence intervals, effect size indices, correlation coefficients and meta-analysis. The studies were conducted between November 2011 and December 2014, and all findings are available online in the "open science" system (the article provides links to these comparisons, as well as to each of the 100 replications). To put things very briefly, this ambitious research project revealed that whereas statistically significant results (of $p < 0.05$) were obtained in 97% of the original studies, in the replication studies this level was much lower – at only 36%. Analysis of the *effect size* values showed that only 47% of such values obtained in the original studies fitted within the 95% confidence interval for the values of these indices in the replications.

So what do we learn of importance from these studies? If one were to use but a single sentence, then that novum could be boiled down to the observation that the level of "repeatability" of the results in the replicated studies was too low. How might one explain this worrying result? If we assume that the studies conducted by the international team supervised by B. Nosek were themselves conducted correctly in terms of methodology (and there are no methodological reasons to believe otherwise), then we have to deliberate over an answer to the following incredibly difficult question: Why was such a small percentage of the studies replicated successfully and with reasonable precision? Incidentally, problems with the repeatability of results also occur in other social disciplines.

Let us reiterate: a research study deserves to be called scientific if – when recreating the original conditions in which it was conducted (i.e. taking into account the profile of the human research subjects, the characteristics of the

research situation, the apparatus, measurement tools, and statistical procedures, etc.) – it can be repeated, providing similar results, independently of the author of the original research. One has to expect a certain level of error, since research studies are not conducted in ideal conditions by ideal researchers using ideal apparatus. One also has to accept the consequences resulting from the historical-cultural perspective and the specifics of the period of development in which the research subjects find themselves. And finally the passage in time from the original study to its replication is also not insignificant. This error will most probably be smaller in studies conducted in laboratories, and greater in social studies involving the participation of people answering questions in a questionnaire. Nevertheless, we do not expect it to be very large. Because only a result that can be reproduced sufficiently precisely in various places around the globe, and by competent researchers, is of interest from the point of view of the accumulation of scientific knowledge; hence in the title of the article by B. Nosek and his collaborators the term "reproducibility The social sciences (including the one of them most mature methodologically, psychology) are putting ever more emphasis on the requirement for replication (see e.g. Neuliep, 1991; Wojciszke, 2014). I would say that it's very good for the methodological condition of our scientific discipline, that such important works as this one are emerging, which will – I believe – long be the subject of discussion. As long as the discussion is constructive, and not destructive and boiling down to unentitled generalisations given in the catchy and sensationalist titles of comment in the press.

The highly-publicised scam committed a few years ago by a social psychologist at the University of Tilburg, Diederik Stapel, also contributed – and significantly so – to the tackling of issues related to protecting research practice from frauds of Stapel's ilk. This infamous social psychologist wrote several dozen "empirical" articles, supported by the results of entirely fabricated studies, and had them published in prestigious psychological journals (and, surprisingly, not in mediocre ones); other studies had glaring errors in their statistical analysis. The outcome of the work performed by three committees (see Levelt Committee, Noort Committee, Drenth Committee, 2012) appointed to analyse the effects of this gigantic scam was the removal from databases of dozens of articles from 15 years of the fraud's work. Yet one cannot help but ask: How was it possible for these fraudulent articles to be accepted for printing? Were the reviewers unreliable? Why were these prestigious journals' editors so naïve that they believed in the author only because he had a reputation as an outstanding researcher in the academic community of social psychology (and was also the dean of the psychology

faculty at his university)? How many more papers written by Stapel-like "researchers" are there in the databases of psychological literature? And the most important question: What can (must) be done for the Stapel affair not to happen again? The discovery of this fraud (incidentally achieved thanks to doctoral students being unable to replicate the studies) drew attention to the requirement (known for years to philosophers and science methodologists) of the replication of research as a criterion of its scientificity. The studies carried out by B. Nosek's group should constitute an enticement to refer to replication as an effective method for fighting artefacts.

Let us turn our attention now to the possible causes behind this appalling practice. I believe there are a few pathological factors contributing to the occurrence of undesirable and quite frankly shameful behaviour in the researcher community.

The first: conceit, rivalry, the desire to keep one's place at the top of the "best Working conditions and threats of a material nature are inconsequential here. Wanting to be among the best means living life in constant stress. All that guides persons such as Stapel (who, after all, belonged to the elite, and who was untroubled by material worries) is the unceasing preoccupation with not dropping out of the leading group, with always being present at prestigious conferences, and with being printed (and quoted!) in the best professional journals. And when ideas no longer suffice, one's resistance to temptation weakens and the subsequent slope is a slippery one.

The second: pressure from the employer – from the faculty head, the institute's director, the department's dean or the rector. Pressure exerted by research institutes' management on their staff, for them to bring in points from publications, to speed up their accumulation of scientific output essential for setting post-doctoral processes in motion, has been intensifying over recent years in Poland. Such excessive bibliometrics in the appraisal of research work, as discussed above, may in extreme cases lead to "shortcuts" being chosen: adding people to publications, artificially dividing and reproducing publications, plagiaries, buying entire works or only parts thereof (for example advanced statistical analyses), or attempts (sadly sometimes successful) at publishing works with data "improved" by their authors, and so on.

The third: social consent and barely perceptible consequences of such actions. Dysfunctional behaviours are also favoured by the absence of an explicitly assertive reaction by the academic community to the violation of academic standards, and particularly among university authorities (at all levels!) – especially when the person involved is "our" employee.

The fourth: an excess of mediocre higher places of learning (especially in the non-public sector). Despite staff requirements already being lowered when establishing new courses of study, and maintaining (although sometimes one would like to use the term "reanimating") those which, at least in the name of decency, should be closed down – the necessity of employing persons with a post-doctoral qualification (accepting second-job appointments of persons classed among the so-called staff minimum for courses at an ordinary degree level) means that applications for such academic advancement are coming from people who do not really feel such a need, who lack the capabilities, or who are incapable of writing a decent academic paper. Let us bear in mind that the law of normal distribution functions here as well. So what can they do? Either leave their place of learning (but where to?) or attempt a specific "short-cut"? Plagiarism, fabricating results, or amending statistical analyses of data are examples of the latter.

The next, fifth factor, is the publishing practice – also mentioned above – in psychological journals where the only articles that count for the publishers are those informing of research studies in which the researcher-author obtained a statistically significant result at the minimum required level of $p=0.05$! A desperate and unethical author then does everything (including data manipulation!) for the findings to be what the findings "should" be. After all, it is a matter of points!

But problems with the replication of research in psychology and other social sciences are not only generated by these five dysfunctional factors listed above. When we pass from the area of the empirical sciences, such as biology, physics or chemistry, into the area of the social sciences (psychology, sociology or pedagogy), then we also have to take into account the *psychological context* of the research, as described above. It is quite common for the variables of this context to be inadequately diagnosed (sometimes not diagnosed at all), or – all the more so – effectively controlled.

Summary

The determinants of the research process in the social sciences presented in this article are not the only ones. Due to the empirical character – as highlighted – of such social sciences as psychology, sociology or pedagogy, there is mention of four fundamental contexts in the practising of scientific research: methodological, psychological, cultural and ethical. The ultimate result of research will depend on the researcher discerning these determi-

nants and correctly referring to them. The article shows various ethical abuses – which must be constantly spoken of – that can be committed not only by the researcher, but also by those inciting the researcher to conduct his or her research and publish its findings in a specific manner. The article also indicates possible ways for counteracting the research and publication practices called into question.

Bibliography

Ajdukiewicz, K. (1983). *Zagadnienia i kierunki filozofii. Teoria poznania. Metafizyka*. Warsaw: Czytelnik.

American Psychological Association (2010a). *Publication manual of the American Psychological Association* (6th edition). Washington, DC: Author.

American Psychological Association (2010b). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author; at: http://www.apa.org/ethics/code/principles.pdf.

American Psychological Association. (2003). Guidelines on multicultural education, training, research, practice, and organizational change for psychologists. *American Psychologist*, *58*, pp. 377-402.

American Psychological Association. (2008). *Report of the task force on the implementation of the* multicultural guidelines. Washington, DC: Author; at: https://www.apa.org/about/policy/multicultural-report.pdf.

APA presidential task force on evidence-based practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, pp. 271-285.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology. Why do we need them? What might they be? *American Psychologist*, *63*, pp. 839-851.

Asch, S.E. (1951/2001). Wpływ nacisku grupy na zmianę i zniekształcenie sądów. In J. Siuta (ed.), *Konteksty ludzkich zachowań* (pp. 17-30). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Blanck, P.D. (red.). (1993). *Interpersonal expectations. Theory, research, and applications*. New York: Cambridge University Press.

Bridgman, P.W. (1927). *The logic of modern physics*. New York: Macmillan.

British Psychological Society (1985). A code of conduct for psychologists. *Bulletin of the British Psychological Society*, *38*, pp. 41-43.

Brzeziński, J. (1976). *Struktura procesu badawczego w naukach behawioralnych*. Warsaw – Poznań: Państwowe Wydawnictwo Naukowe.

Brzeziński, J. (1978). *Metodologiczne i psychologiczne wyznaczniki procesu badawczego w psychologii*. Poznań: Wydawnictwo Naukowe UAM.

Brzeziński, J. (2004). Trafność zewnętrzna terenowych badań eksperymentalnych. *Zagadnienia Naukoznawstwa*, *161*(3), pp. 455-469.

Brzeziński, J. (2012a). Jakich kompetencji badawczych oczekujemy od psychologa? In H.J. Grzegołowska-Klarkowska (ed.), *Agresja, socjalizacja, edukacja. Refleksje i inspiracje* (pp. 383-409). Warsaw: Wydawnictwo Akademii Pedagogiki Specjalnej.

Brzeziński, J. (2012b). Co to znaczy, że wyniki przeprowadzonych przez psychologów badań naukowych poddawane są analizie statystycznej? *Roczniki Psychologiczne*, *15*(3), pp. 7-39.

Brzeziński, J. (2015a). Jeżeli oceniać (jednostki naukowe i badaczy), to JAK oceniać. Przeciwko *IF*, a za *peer review*. *Nowotwory Journal of Oncology*, *65*, pp. 476-480; also at: http://www.staff.amu.edu.pl/~brzezuam/.

Brzeziński, J. (2015b). Czy (i jakie) potrzebne są testy inteligencji? In *Seminarium Polskiej Akademii Umiejętności, T. XI, 2013–2015* (pp. 49-76). Kraków: Wydawnictwo Polskiej Akademii Umiejętności.

Brzeziński, J. (2015c). O powinnościach psychologów wobec społeczeństwa. Zawód psychologa jako zawód zaufania publicznego. W: E. Jeliński, Z. Stachowski, S. Sztajer (eds), *Ratio, religio, humanitas. Miscellanea dedykowane Profesorowi Zbigniewowi Drozdowiczowi* (pp. 45-57). Poznań: Wydawnictwo Naukowe Wydziału Nauk Społecznych UAM.

Brzeziński, J. (2016a). Przeciwko depersonalizacji i nadmiernej standaryzacji procesu ewaluacji w nauce. *Zagadnienia Naukoznawstwa*, *52*(1), pp. 127-137.

Brzeziński, J. (2016b). Towards a comprehensive model of scientific research and professional practice in psychology. *Current Issues in Personality Psychology*, *4*(1), pp. 2-10; also at: http://www.staff.amu.edu.pl/~brzezuam/.

Brzeziński, J. (2016c). *Metodologia badań psychologicznych* (5th edition). Warsaw: Wydawnictwo Naukowe PWN.

Brzeziński, J. (2016d). Chapter 5. Etyka postępowania psychologa klinicznego. In L. Cierpiałkowska, H. Sęk (eds), *Psychologia kliniczna* (vol. 1, pp. 81-98). Warsaw: Wydawnictwo Naukowe PWN.

Brzeziński, J., Chyrowicz, B., Poznaniak, Toeplitz-Winiewska M. (2008). *Etyka zawodu psychologa*. Warsaw: Wydawnictwo Naukowe PWN.

Brzeziński, J., Toeplitz-Winiewska, M. (2015). Psycholog jako zawód. Sześć powodów dla których powinniśmy być razem. *Przegląd Psychologiczny*, *58*(4), pp. 405-418.

Bunge, M. (1967). *Scientific research I. The search for system*. Berlin: Springer-Verlag.

Carnap, R. (1959). The methodological character of theoretical concepts. In H. Feigel and M. Scriven (eds), *Minnesota Studies in the Philosophy of Sciences* (vol. 1, pp. 38-76). Minneapolis,MN: University of Minnesota Press.

Chambers, Ch.D., Feredoes, E., Muthukumaraswamy, S.D., Etchells. P.J. (2014). Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, *1*, pp. 4-17. At: http://orca.cf.ac.uk/59475/1/AN2.pdf.

Chambers, Ch., Munafo, M. (2013). *Trust in science would be improved by study pre-registration*; at: http://www.theguardian.com/science/blog/2013/jun/05/trust-in-science-study-pre-registration.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, NJ: Erlbaum.

Cronbach, L., Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

DORA (2012). *San Francisco Declaration on Research Assessment. Putting science into the assessment of research*, at: http://www.ascb.org/files/SFDeclarationFINAL.pdf?157685.

Embretson, S.E., Reise, S.P. (2000). *Item response theory for psychologists*. Malwah, NJ: L. Erlbaum.

FPRM (2002). *US Federal Policy on Research Misconduct*; at: www.federalregister.gov/articles/2000/12/06/00-30852/executive-office-of-the-president-federal-policy-on-research-misconduct-preamble-for-research#h-16 .

Fundacja na Rzecz Nauki Polskiej (2014). *Oświadczenie władz Fundacji na rzecz Nauki Polskiej*, at: www.fnp.org.pl/assets/O%C5%9Bwiadczenie-Rady-i-Zarz%C4%85du-FNP.pdf.

Franco, A., Malhotra, N., Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. Science, *345*(6203), pp. 1502-1505.

Frankfort-Nachmias, Ch., Nachmias, D. (2001). *Metody badawcze w naukach społecznych*. Poznań: Zysk i S-ka Wydawnictwo.

Grabski, M.W. (2015). Uczciwość i wiarygodność nauki. In M.W. Grabski. *O nauce w Polsce – zamyślenie* (pp. 171-197). Kraków: Polska Akademia Umiejętności; reprint from: *Nauka*, 2009, *2*, pp. 37-59.

Gulliksen, H. (1950). *Theory of mental tests.* New York: J. Wiley.

Harasimczuk, J., Cieciuch J. (2012). Podstawowe standardy edytorskie naukowych tekstów psychologicznych w języku polskim na podstawie reguł APA. Warsaw: Wydawnictwo Liberi Libri; at: http://www.liberilibri.pl/harasimczuk.

Hardin, E.E., Robitschek, Ch., Flores, L.Y., Navarro, R.L., Ashton, M.W. (2014). The cultural lens approach to evaluating cultural validity of psychological theory. *American Psychologist*, *69*, pp. 656-668.

Hornowska, E. (1989). Operacjonalizacja wielkości psychologicznych. Założenia – struktura – konsekwencje. Wrocław: Zakład Narodowy im. Ossolińskich.

International Union of Psychological Science (2008). *Universal Declaration of Ethical Principles for Psychologists*; at: www.iupsys.net/about/governance/universal-declaration-of-ethical-principles-for-psychologists.html.

Jachimczyk, A. (2015). Otwarte dane badawcze. Casus polskich instytutów. *Zagadnienia Naukoznawstwa*, *51*(4), pp. 409-424.

Jaworska, A. (2006). Główne nurty w metodologii badań nad skutecznością psychoterapii – w poszukiwaniu „trzeciej drogi." In J. Brzeziński (ed.), *Metodologia badań psychologicznych. Wybór tekstów* (pp. 116-147). Warsaw: Wydawnictwo Naukowe PWN.

Kerr, N.L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, pp. 196-217.

King, B.M., Minium, E.W. (2009). *Statystyka w psychologii i pedagogice*. Warsaw: Wydawnictwo Naukowe PWN.

Kirk, R.E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd edition). Belmont, CA: Brooks.

Kleka, P. (2011). Statystyczne kryteria przydatności raportu z badań do metaanalizy. In J.M. Brzeziński (ed.), *Metodologia badań społecznych. Wybór tekstów* (pp. 99-114). Poznań: Zysk i S-ka Wydawnictwo.

Kmita, J. (1973). *Wykłady z logiki i metodologii nauk*. Warsaw: Państwowe Wydawnictwo Naukowe.

KPPUE (2000). Karta Praw Podstawowych Unii Europejskiej [Charter of Fundamental Rights of the European Union]. *Dziennik Urzędowy Unii Europejskiej* (C83, vol. 53 from 30 March 2010) (pp. 389-403); at: http://www.lgeringer.pl/pliki/traktat_lizbonski_pl.pdf.

Krajewski, W. (1998). *Prawa nauki. Przegląd zagadnień metodologicznych i filozoficznych*. Warsaw: Książka i Wiedza.

Kuhn, T.S. (1996/2001). Struktura rewolucji naukowych [The Structure of Scientific Revolutions] (Polish translation of 3rd edition, 1996). Warsaw: Fundacja Aletheia.

Leslie, K. J., Loewenstein, G., Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, pp. 524-532.

Levelt Committee, Noort Committee, Drenth Committee (2012). Flawed Science: The fraudulent research practices of social psychologist Diederik Stapel. (Unpublished report.) At: https://www.tilburguniversity.edu/upload/3ff904d7-547b-40ae-85fe-bea38e05a34a_Final%20report%20Flawed%20Science.pdf.

Lombrozo, T. (2014). Science, trust and psychology in crisis; at: http://www.npr.org/blogs/13.7/2014/06/02/318212713/science-trust-and-psychology-in-crisis.

Lord, F.M., Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, pp. 371-378.

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, *18*, pp. 57-76.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper & Row.

NeuroChambers (2013). *Changing the culture of scientific publishing from within*; at: http://neurochambers.blogspot.co.uk/2012/10/changing-culture-of-scientific.html.

Neuliep, J. W. (red.). (1991). *Replication research in behavioral sciences*. Newbury Park, CA: Sage Publications.

Nezlek, J.B. (2014). Znaczenie norm i wartości. Uwagi do artykułu Jerzego Brzezińskiego o stanie polskiej psychologii. *Roczniki Psychologiczne*, *17*(3), pp. 593-597.

Nisbett, R.E. (2016). *MINDWARE: Narzędzia skutecznego myślenia*. Sopot: Smak Słowa.

Nowak, L. (1974). *Zasady marksistowskiej metodologii nauk. Próba systematycznej rekonstrukcji*. Warsaw: Państwowe Wydawnictwo Naukowe.

Orne M.T. (1962/1991). Eksperyment psychologiczny z punktu widzenia psychologii społecznej ze szczególnym uwzględnieniem wpływu zmiennych sugerujących hipotezę i ich implikacji. In J. Brzeziński, J. Siuta (eds). *Społeczny kontekst badań psychlogicznych i pedagogicznych. Wybór tekstów* (pp. 15-32). Poznań: Wyd. Nauk. UAM.

Orne M.T. (1973/1993). Komunikowanie się w sytuacji eksperymentalnej: dlaczego jest istotne, jak jest oceniane i jakie ma znaczenie dla trafności ekologicznej. In J. Brzeziński (ed.), *Psychologiczne i psychometryczne problemy diagnostyki psychologicznej* (pp. 31-68). Poznań: Wyd. Nauk. UAM.

Parnas, D.L. (2007). Stop the numbers game. *Communications of the ACM*, *50*(11), pp. 19-21; also at: www.ce.sharif.edu/~ghodsi/PaP/stop_the_number_game.pdf.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research. Explanation and research* (3rd edition). Fort Worth, TX: Harcourt Brace College Publishers.

Polskie Towarzystwo Psychologiczne (1992). *Kodeks etyczno-zawodowy psychologa*. Warsaw: Polskie Towarzystwo Psychologiczne; at: http://www.ptp.org.pl/modules.php?name=News&file=article&sid=29.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science, *Science*, no. 349(6251) DOI: 10.1126/science.aac4716; also at: http://www.

psykologforbundet.se/Documents/Psykologtidningen/Aktuellt%20Pdf/Science%20aug%202015.pdf.

Quintana, S. M., Troyano, N., Taylor, G. (2001). Cultural validity and inherent challenges in quantitative methods for multicultural research. In J.G. Ponterotto, J.M. Casas, L.A. Suzuki, C. M. Alexander (red.), *Handbook of multicultural counseling* (2nd edition, pp. 604-630).Thousand Oaks, CA: Sage.

Reichenbach, H. (1938). Experience and prediction. *An analysis of the foundations and the structure of knowledge*. Chicago, IL: University of Chicago Press.

Rosenberg M.J. (1969/1991). Warunki powstawania oraz konsekwencje lęku przed oceną. In J. Brzeziński, J. Siuta (eds), *Społeczny kontekst badań psychologicznych i pedagogicznych. Wybór tekstów* (pp. 61-135). Poznań: Wyd. Nauk. UAM.

Rosenthal, R. (1996). Nauka a etyka w przeprowadzaniu badań psychologicznych oraz analizowaniu i przedstawianiu ich wyników. *Czasopismo Psychologiczne*, *2*, pp. 37-46.

Rosenthal, R., Rosnow, R.L. (2009). *Artifacts in behavioral research*. New York: Oxford University Press.

Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review*, *40*, pp. 337-354.

Scott, S. (2013). Pre-registration would put science in chains; at: http://www.timeshighereducation.co.uk/comment/opinion/science-in-chains/2005954.article.

Skipper, J.K., Guenther, A.L., Nass, G. (1970). The sacredness of .05: A note concerning levels of significance in social science. W: R.E. Henkel, D.E. Morrison (ed.), The significance test controversy. A reader (pp. 155-160). London: Butterworths.

Smith, G.T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, *17*, pp. 396-408.

Song, F., Hooper, L., Loke, Y.K. (2013). Publication bias: what is it? How do we measure it? How do we avoid it? *Open Access Journal of Clinical Trials*, *5*, pp. 71-81; also at: www.dovepress.com/getfile.php?fileID=16644.

Spendel, Z. (2016). O dylematach „metodologia – etyka" we współczesnej psychologii. *Czasopismo Psychologiczne*, *22*(1), pp. 15-26.

Towpik, E. (2013). Deklaracja z San Francisco nt. Oceny Badań Naukowych (DORA): Czy to początek końca dotychczasowej metody oceny parametrycznej? *Nowotwory Journal of Oncology*, 63, pp. 501–504.

Towpik, E. (2015). IF-mania: Journal Impact Factor nie jest właściwym wskaźnikiem oceniania wyników badań naukowych, indywidualnych uczonych ani ośrodków badawczych. *Nowotwory Journal of Oncology*, *65*, pp. 465-475.

Trusz, S. (ed.). (2013). *Efekty oczekiwań interpersonalnych. Wybór tekstów*. Warsaw: Wydawnictwo Naukowe Scholar.

Wilkinson, L. & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, pp. 594-604.

Winer, B.J., Brown, D.R., Michels, K.M. (1991). *Statistical principles in experimental design* (3rd edition). New York: McGraw-Hill.

Wojciszke, B. (2011). Systematycznie Modyfikowane Autoreplikacje: logika programu badań empirycznych w psychologii. In J. Brzeziński (ed.), *Metodologia badań społecznych. Wybór tekstów* (pp. 19-54). Poznań: Zysk i S-ka Wydawnictwo.

Figures