

EXPERIMENTAL AND INTERDISCIPLINARY APPROACHES IN PHILOSOPHY: METHODOLOGICAL CAVEATS

*Marco Antonio Azevedo*¹

UNISINOS

*Jairo Othero*²

UNISINOS

RESUMO: Filósofos, especialmente aqueles que se autodenominam como naturalistas, parecem cada vez mais interessados em realizar pesquisas empíricas. Um caso típico é o da Filosofia Experimental, um campo emergente que faz uso de dados empíricos colhido por meio de inquéritos seguindo os mesmos métodos empregados nas ciências empíricas, nomeadamente na psicologia, a fim de apresentar provas em pesquisas ou argumentos filosóficos. Outro exemplo é oferecido pela participação ativa dos filósofos em grupos de pesquisa interdisciplinares nas neurociências. No entanto, os filósofos não são especialistas em metodologia de pesquisa empírica. Isso é transparente nos estudos mais representativos em filosofia experimental. Neste artigo, faremos algumas observações de advertência sobre as complexidades da metodologia das ciências empíricas. A recomendação final é de que é sábia a atitude de envolver-se em grupos de pesquisa interdisciplinares, buscando dominar as peculiaridades dos mais diversos delineamentos de pesquisa empírica nas diferentes áreas da ciência.

PALAVRAS-CHAVE: Filosofia Experimental, Ciências empíricas, Naturalismo, Metodologia Filosófica, Metodologia Científica, Metafilosofia.

ABSTRACT: Philosophers, especially those who classify themselves as naturalists, seem to be increasingly interested in conducting empirical researches. One case is Experimental Philosophy, an emerging field that makes use of empirical data gathered through surveys following the same methods employed in the empirical sciences, notably in psychology, in order to present evidence in philosophical arguments or inquiries. Another example is offered by the active participation of philosophers in interdisciplinary research groups in the neurosciences. Nevertheless, philosophers are not experts in the methodology of empirical research. This is transparent in the most representative studies in experimental philosophy. In this paper, we present some cautionary remarks on the complexities of the methodology of the empirical sciences. The final recommendation is that to engage in interdisciplinary research groups, seeking to master the peculiarities of the many empirical research designs in the most different areas of science, is the wisest attitude.

KEY-WORDS: Experimental Philosophy, Empirical Sciences, Naturalism, Philosophical Methodology, Scientific Methodology, Metaphilosophy.

¹ MD, PhD, Assistant Professor, Graduate Program in Philosophy, Unisinos.

² MD, PhD Student, Unisinos.

I. Introduction

The assumption that philosophy, especially in certain fields such as ethics, must be “empirically informed” is not in fact controversial advice (LOUGHLIN 2011). But this is different with the emergent field of *X-Phi*, or experimental philosophy (EP) (LACKMAN 2006; KNOBE 2004; KNOBE & NICHOLS 2008; APPIAH 2007, 2010). The same could be said about the interplay of philosophers with research groups in other fields of enquiry in the non-humanistic sciences (HAUG 2013; SMITH 2008, 2013; SPENCE, MICHAEL & SMITH 2014) – let’s call this trend “interdisciplinary philosophy” (IP). Both practices seem to represent methodological innovations in philosophical work. Both also display rigorous criticism of the methods of so-called “armchair” philosophy, even though EP is more radical in its criticism to traditional methods – “burning armchair” is said as the unofficial logo of EP (HAUG 2013).

Innovation? Maybe not, for, at least in the case of EP, their supporters say that the attitude is not actually new; it would be the same of traditional philosophy. The difference would be mainly *methodological*. According to Joshua Knobe, one of the main exponents of EP, traditional philosophers were not methodologically careful in the process of collecting evidence. He argues that EP is an original approach, unique in the history of philosophical research, since EP searches for empirical data employing the same methods as empirical sciences, more precisely, the same methods employed in contemporary researches in psychology, including quantitative statistical methods. This makes EP a relatively new area of research, even though, says Knobe, “much important work still remains to be done.” Nevertheless, “the field appears to be growing extremely rapidly”, so “[w]e will surely see a number of surprising results in the years to come” (KNOBE 2004). Well, this is Knobe’s prognosis at least. In fact, most philosophers agree that

the field has grown quickly since its inception, but it is disputable if their results are actually surprising.

Here is the main target of controversy: the study of subjects' "intuitions". In a paper published in the journal *Philosophical Psychology*, Knobe describes a couple of experiments designed to show that people manifest a peculiar asymmetry in their ascriptions of intentions: the so-called "Knobe effect" (KNOBE 2003a, 2003b). More explicitly, the research problem was whether evaluative considerations – considerations of good and bad, right and wrong, praise and blame – played any role in people's intentional ascriptions.³ The findings seem initially unsurprising: people attribute intentionality more often if they evaluate the actions negatively than positively. However, the interpretations of the data are subject to controversies (HOLTON 2010, CULLEN 2010) – we will see something about this below (section VI). It is claimed anyway that if Knobe is right, then philosophers should be more careful when they appeal to the intuitions of their readers when arguing philosophically. We cannot assume that readers' intuitions on any subject are true, at least not *simpliciter*, for human intuitions on many subjects are incoherent and biased.

Knobe's famous study is a paradigmatic example of EP research. Nevertheless, sometimes the study is described as "interdisciplinary". The study is said to be an interdisciplinary field that uses methods "normally associated with psychology to investigate questions normally associated with philosophy" (KNOBE et al 2012). So, it could be argued that this is not a paradigmatic example of EP but one of what I called above an IP research. Anyway, the question is: are the research methods employed the same as those used in other empirical fields? Knobe and his collaborators say: yes. This

³ Knobe revealed that Alfred Mele's "armchair" criticisms to his first experiments led him to modify his research questions.

seems to be true concerning Knobe's original study of 2003. In this study, the researcher made claims about people's standard moral intuitions, submitting subjects to "controlled experiments" and "the results are subjected to the usual statistical analyses" (KNOBE et al 2012, p. 82). Nevertheless, Knobe's study design is a target of several methodological criticisms (WOOLFOLK 2013) – we will see some of them in section VI.

IP philosophers by their turn are less criticized than the "experimentalists". Perhaps one reason is that they do not intend to represent any new philosophical movement. What they intend is to take seriously the tenets of methodological naturalism: if a given philosophical topic concerns substantive empirical issues, it should be informed about the science of that subject (SMITH 2014). They see themselves as contributing to a larger, interdisciplinary project of providing a fully comprehensive account of a given subject; so they do not pretend that philosophers should get data conducting experiments (even though they do not argue against it in principle), so they prioritize their insertion in interdisciplinary research groups. Philosophers are not experts in the methods of empirical sciences, so they do better if they work in company with good scientists. Nevertheless, as argued by Barry Smith, the scientific task of interpreting data can be better made with the help of philosophers, with their expertise in conceptual analysis. However, a critic could say that what those philosophers are doing is not actually philosophy but plain science. A friend of traditional philosophy could still say that the best philosophical work is still made only in the armchair.

Seen this way, EP philosophers seem to be more "radical" than the IP. EP philosophers do not give much value to the strategy of meddling in interdisciplinary groups. They want to conduct experiments for themselves. And they are more radical critics of their main opponents, the orthodox conceptual analysts (OCA) (KNOBE & NICHOLS 2008). In fact, complaints

are reciprocal. Criticisms of EP such as by Ernest Sosa (see LACKMAN 2004) show this emphatically. The experimentalists, says Sosa, can't prove that the human subjects of EP studies are actually responding to the philosophical principles at stake. OCA think that philosophical principles must be established only by means of conceptual analysis (or at least this is what their criticisms of EP seem to imply). But OCA's criticisms of EP end up showing, perhaps inadvertently, some internal methodological flaws in the studies of experimentalists. The lesson is that if philosophers want to conduct research themselves or if they want to participate in larger interdisciplinary teams, it is mandatory to them that they master the methodological subtleties of empirical studies. And they should master them not as philosophers of science, but as scientists.

In this paper we want to sustain that if philosophers want to take both enterprises (EP and IP) seriously, they should take the complex and diversified methodologies of the empirical sciences more seriously. As we will see, several criticisms of EP can be read as criticisms of EP's methodological commitments. Many of these criticisms come from philosophers or scientists well acquainted with the methodological diversity of empirical studies. Since EP and IP intend to be empirical and scientific, their supporters should be aware of the peculiar constraints of these different kinds of research designs in the sciences. They must clearly disclose the methods employed in their research, in the same way researchers of others fields of enquiry do. This will allow the reader a thorough evaluation of the inferential strength of their study.

But before dealing with this subject directly, let us firstly present a brief review of some critical topics in the philosophy of science, or more accurately, in the philosophy of the methodology of empirical studies.

II. Naturalism & Methodology

Both EP and IP are approaches committed to *methodological* naturalism (MN).⁴ Here we can distinguish two different MN trends, however. We will call them the enthusiastic and the moderate. *Enthusiastic* MN (EMN) claims that philosophical analysis does not have any substantial or productive role in philosophy. As Peter Railton formulates the position, it exists in the Quinean claim that philosophy does not possess a distinctive, a priori method able to yield substantive truths that, in principle, are not subject to any sort of empirical test. So, for EMN, good philosophy proceeds only *a posteriori*, “in tandem with perhaps as a particularly abstract and general part of the broadly empirical inquiry carried on in the natural and social sciences” (RAILTON 1989, 155–156). EMN also proposes that in ethics philosophical method is continuous with scientific method too, and thus metaethical questions should be answered within the framework of a broadly scientific epistemology (LEITER 2007, 8). In sum, EMN claims that philosophical analysis does not have any substantial or productive *scientific* role. In other words, philosophical analysis does not contribute to the progress of our knowledge in any substantial field. Maybe the contribution of philosophy is only negative. Maybe philosophy is for most of the EMN just a kind of literary critical enterprise.

Moderate MNs (MMN) accept conceptual analysis by their turn, but restrict it to the study of concepts.⁵ In substantive fields, MMN are on a par with the EMN: they say that philosophy should cohere with the results of and

⁴ It's argued that naturalism can be distinguished into two types, *methodological* and *substantive* (RAILTON 1990, GOLDMAN 1994, LEITER 1998). Methodological naturalists claim that philosophical theorizing should be continuous with empirical inquiry in the sciences, and they defend that “either in virtue of their dependence upon the actual results of scientific method in different domains or in virtue of their employment and emulation of distinctively scientific ways of looking at and explaining things” (LEITER 2014). Substantive naturalism is either the ontological view that only *natural* or *physical* things exist, or the semantic view that a suitable philosophical analysis of any concept must show it to be amenable to empirical inquiry.

⁵ Several MMN do not see themselves as “naturalists”, since they use the label for those philosophers I here call “enthusiastics”. One example is Timothy Williamson (2014).

(at least when dealing with empirical questions) follow the methods of the empirical sciences, but they do not unanimously make the enthusiastic claim that only the hard sciences are our guide to the real constituents of the objective world (LEITER 2007). Some MMN even criticize physicalism, advocating for a more pluralist conception of the sciences and their methods (Mayr 1991, 104–5).

By their turn, OCA think that philosophical methodology is radically different from empirical methods. Following Quine and others, MNs claim just the opposite (RORTY 1992); if they were right, philosophical methodology and scientific methodology would be essentially similar. But in which way are they similar? What is the essential similarity between the methods of empirical investigations and methods of conceptual analytical philosophy?

III. Scientific Reasoning: Inductive or Deductive?

It is plausible that post-linguistic philosophers such as Rorty are wrong in saying that there is not any substantial difference in methodology between science and philosophy. But, independently of that, even if conceptual analysis has its proper methodological domain and is an enterprise substantially different from the empirical, this is not an argument that only philosophical works contain conceptual analysis, neither that only philosophers are apt to make it. Scientists also make conceptual analysis, at least in the preparatory part of their studies. Maybe philosophers are better in this enterprise. But even if surgeons are better in anatomy, this is not an argument that anatomy is an essential part of surgery and does not play any role in the art of clinicians. Maybe philosophers are better in conceptual analysis in the same sense that surgeons are better in anatomy. Philosophy can comprise conceptual analysis even if it is not what defines philosophy as an especial

branch of scholarly knowledge.

Let's accept that conceptual analysis is not an exclusive domain of philosophers, and let's assume that the same can be said about empirical methods with respect to scientists. This can be rightly assumed, for even if we recognize that scientists are better in empirical methods, this is not an argument for the claim that philosophers cease to behave as philosophers if they follow empirical methodological procedures in their researches. After all, as rightly claimed by the experimentalists, it is historically true that traditional philosophy was basically interested in empirical issues; and if the issue is empirical, and if the research question is also an empirical question, it is expected that the research method is suited to the nature of the issue; and since scientific method substantially changed after the first half of the twentieth century, philosophers persuaded to return to the traditional way of doing philosophy should learn how to adapt their research to the research methods already established in other fields of science. After all, much has changed in terms of methodology in the sciences since the 1950s.

With respect to questions of method, it is worth noting that simplification is not a virtue. Philosophers love reductions and simplifications. For example, when doing philosophy of science, philosophers usually talk of science as if it were a unique enterprise with a single basic methodology. So one traditional trend in philosophy of science was (and in fact continues to be) the disclosure of the logic behind the sciences (a paradigmatic OCA's enterprise). However, the result was a gross simplification, maybe only useful for undergraduate teaching. Elegant simplifications such as Hempel's deductive-nomological method, as well as Popper's falsifiability criteria, are not useful if we want to discriminate the central differences in the pluralist and diversified realm of science.

Take one famous “undergraduate” (and misleading) divide, the well-known separation between the *inductive* and the *deductive* sciences. This divide supposes a different “logic” behind those two modes of “reasoning” employed in the sciences (in a broad sense). Alfred Tarski was one who considered a difficult and problematic task to disclose the methodology of empirical sciences. In the preface of his *Introduction to Logic* (TARSKI [1941] 1995), Tarski says he is inclined to “doubt whether any special ‘logic of empirical sciences’”, as opposed to logic in general or the “logic of deductive sciences”, exists at all. So he advanced a view that, for the purpose of an adequate methodological treatment, an empirical science should be considered not merely as a scientific theory (that is, a system of asserted statements arranged according to certain rules), “but rather a complex consisting partly of such statements and partly of human activities” (TARSKI 1995, xii). This is a sensible caveat. If we want to understand the complexities and subtleties of the sciences we should understand them not only as bodies of statements and theories, but also as human *practices*. We unfortunately lack a good approach to scientific methodology in this pragmatic sense, that is, as a body of statements *and* practices.

Statistics is an example of a practice, maybe an inferential practice; and it is certain that if we want to understand the methodological subtleties in the sciences it should be important to study the methods of statistical inferences. It is a common belief that statistical inference is “inductive”. But, if statistical inference is inductive, why do the statisticians remark that the logic of statistical inference is actually *deductive*? See the following passage from a well-known handbook of statistics:

Both statistical and research hypotheses are tested by using deductive reasoning (...) to derive their consequences: what must also be true if the hypothesis is true. The mathematical, deductive reasoning for testing statistical hypotheses is part of

hypothesis-testing theory (...). The deductive reasoning for testing a research hypothesis is unfortunately not provided by a convenient theory; each investigator must develop the logical consequences of his or her own hypotheses. Thus, an obvious example of a deductive prediction is: *If cigarette smoking can cause lung cancer, then it is logical to expect more cases of lung cancer in a population that smokes cigarettes than in a population that does not* (BERNSTEIN & BERNSTEIN 1998, 55 – our italics).

We are taught that an argument is deductive if its conclusions cannot be false if the premises are true. So, if the conclusion of an empirical study is the conclusion of a deductive “reasoning”, then the “consequences” of the research follow deductively from the research hypotheses and “something else” (the auxiliary assumptions). But the example of Bernstein and Bernstein, however, does not express that: we surely can expect more cases of lung cancer in a population that smokes cigarettes than in a population that does not, even though “lung cancer” is not a consequence (in the logical sense) of “cigarette smoking”. It is, hence, reasonable to expect more cases of lung cancer in a population of smokers than in a population of abstainers, but the “logic” behind that is not simply a matter of logical consequence (as is well known to every undergraduate student in philosophy acquainted with the basics of classical logic).

However, what Bernstein & Bernstein intend to say becomes clearer in a passage from another of their books:

The logic of a hypothesis test is an argument based on deductive reasoning (...) that is [it is an argument] done in mathematical form. The logic begins with the null hypothesis as its starting premise and then examines the logical consequences of this premise—what must also be true if the null is true. In particular, it determines the conditional probability (...): Given that $H_0: \theta = \theta_0$ is true, what is the probability of getting a point estimate at least as different from the value hypothesized by the null (θ_0) as the one calculated from the sample (θ^*)? Or, in symbolic form:

P (point estimate at least as different from θ_0 as θ^* | H_0 is true).

(BERNSTEIN & BERNSTEIN 1999, 207.)

So it seems that Hempel was correct in saying that scientific justifications are deductive (HEMPEL 1945, 1966); but Hempel's deductive-nomological model is unfortunately overly simplified if we take it as a general description of statistical inferences of empirical comparative studies (even though some philosophers of science would say that this was not his intention). Since statistical tests are not (in a strict sense) deductive proofs, which is the deductive part of the statistical inferences? Anyway, present scientific tests or procedures (trials, scientific observations, etc.) are warranted by statistical inferences and not by *formal* deductions.

Popper seems to be more accurate when he describes empirical reasoning as a kind of *modus tollens* (POPPER [1959] 2002, 55–6). But his approach was also simplistic. Elliot Sober is more accurate: he calls the “probabilistic modus tollens” (PMT) the general form of the inference made by scientists when they interpret the conclusions of their empirical researches. The view is associated with the frequentist approach to statistics. Sober nevertheless dubbed this approach a “lazy” method (SOBEL 2008, 353). Maybe there is really something lazy, for methodology must be careful and not fast. Sober is a Bayesian, and he is a critic of the frequentist approach. Frequentists and Bayesians are taken to be opponents, but it is plausible that both approaches are in fact complementary. Unfortunately, we cannot deal with this subject here – to be honest, we are not sufficiently experts in statistics to deal with all the details of this famous controversy.

What is after all the best description of the “logic” behind statistical inferences? In fact, it is not easy to show which deduction is meant by statistical inference recommended by the frequentists. But a crucial, albeit neglected, point in the traditional approaches of the “undergraduate”

philosophy of science is just the formal description of the method of rejecting the *null* hypothesis.

Since we lack an adequate formal description of the “probabilistic modus tollens” that does justice to the inferential procedure of rejecting the null hypothesis, let us try to offer an (albeit provisional) suggestion (here are two versions):

I		
$(H_0 \vee H_1) \wedge (H_0 \rightarrow \neg E) \vdash (E \rightarrow H_1)$		
1. $(H_0 \vee H_1)$		
	P	
2. $(H_0 \rightarrow \neg E)$		
	P	
3.		E
	Sup	
4.	\neg	H_0
	2,3 MT	
5.		H_1
	1,4 DS	
6.	$E \rightarrow$	H_1
	3-5 DT	
p-value = $\text{pr}(E H_0)$		
If $\text{pr}(E H_0) \leq$ very small value		
(0,05 or 0,01), then $\text{pr}(\neg E H_0)$ is		
very high.		
If $\text{pr}(\neg E H_0)$ is very high, then		
$(H_0 \rightarrow \neg E)$ is a reasonable bet.		

II		
$(H_0 \vee H_1) \wedge (H_0 \rightarrow \neg E) \wedge E \vdash H_1$		
1. $(H_0 \vee H_1)$		
	P	
2. $(H_0 \rightarrow \neg E)$		
	P	
3. E		
	P	
4. H_0		
	Sup	
5. $\neg E$		
	2,4 MP	
6. $E \wedge \neg E$		
	3,5 C	
7. $\neg H_0$		
	4-6 RA	
8. H_1		
	1,7 DS	
Note: 4-6 can be omitted using		
<i>modus tollens</i> to infer 7		

As Bernstein & Bernstein state:

[the] conditional probability is determined by using a test statistic, which is a sample statistic (...) with these properties: (1) it allows comparison of the sample point estimate (θ^*) and the null's hypothesized value ((θ_0) , and (2) it is associated with a

probability distribution that is known under the assumption that H_0 is true (BERNSTEIN & BERNSTEIN 1999, 207).

In sum, researchers begin stating their own hypothesis that will represent the hypothesis of the study, H_1 . The problem is that the researchers cannot test evidences directly in order to confirm this hypothesis (we could call this “the Popperian problem”); but if they can find evidence that makes some hypothesis H_0 very unlikely, assuming that this is roughly the same as assuming that this hypothesis is false (a pragmatic supposition), since the falsity of this hypothesis (the null) entails the hypothesis of the study (H_1), then they can prove (deductively) what they want. So the method (or the *maneuver*) consists in establishing a very small probability for the study findings (the evidences) conditional to the null hypothesis, and then calculating what the study actually shows. If the calculated probability of the observed data conditional to the null hypothesis is very low (for example, 0.05 or 0.01), the researcher is authorized (in fact, not “deductively”, but maybe “inductively”, that is, for practical aims) to conclude that the data confirm the study hypothesis, that is, H_1 (since either H_1 or H_0 is true). This is a very ingenious procedure, and in fact it was a very ingenious *invention*. This mathematical invention was responsible for the remarkable progress in the statistical methods employed by the empirical sciences after the second half of the twentieth century; it is hence responsible for the most relevant part of the scientific knowledge of our present era.⁶

⁶ The method of hypothesis testing is, in fact, a creation, invented in the period 1915–1933 by three great men (not usually mentioned in philosophy undergraduate textbooks): Ronald A. Fisher (1890–1962), Jerzy Neyman (1894–1981), and E. S. Pearson (1895–1980) (NEYMAN & PEARSON 1933; LEHMANN 1993). Some say that the modern theory of testing hypotheses began even earlier, with a student's discovery of the *t* test in 1908. There are certainly more people involved than those four great men. Fisher was a genius, with several contributions to science, especially in biology. In statistical methods, one of his main contributions is the idea of **design** of experiments (FISHER 1965). The present variety of study designs with different “levels of evidence” began with Fisher's proposals. Fisher is also the main exponent of the frequentist approach; he is a historical opponent of the Bayesian approach to statistics.

IV. Experiments trump observations?

We present hypotheses as best explanations for what we know or observe (LIPTON 1994; SOBER 2000, 2008). It seems that in science, as in philosophy as well as in ordinary reasoning (HARMAN 1965), we depart from observations and then try to explain them. So Bayesians are right; but this is not the whole story.

Some scientific reasoning seems to depart from a very special kind of observations: the experimental. Pragmatic philosophers think that this is what makes science qualitatively different from any other kind of reasoning (HACKING 1983). The reason is that scientific data are not *there* in the world; scientists *create* them. As remarked by the philosopher Ian Hacking, “the popular image of scientists is someone in a white coat in a laboratory.” Experiment was declared the royal road to knowledge. Scientists must “twist the lion’s tail” (HACKING 1983), so they *manipulate* the world in order to learn its secrets.

Hacking’s approach to experimental science seems persuasive. Experimental science is the icon of all sciences. In it, scientists are sovereign: they cannot be deluded by nature for, in the experimental field, in the laboratory, scientists are in control. In it they can command nature, at least partially. It is usual to say that scientists explain the phenomena that they discover in nature. But in fact what happens in labs is different. Hacking thinks that this fact is so neglected that “we lack a name for it.” Hacking calls it the *creation of phenomena*: “[s]cientists create which then become the centerpieces of theory” (HACKING 1983, 120). So it seems that environmental manipulation made in order to get appropriate data is the essence of all scientific method. But this is not completely right.

Hacking is not fully correct in saying that traditional scientists always “create” the phenomena they intend to explain theoretically. This is true in the case of experimental physics and maybe in other fields of experimental sciences, but this is not true about other very traditional fields of enquiry, such as astronomy, for example. In astronomy, instruments serve to observe objects obviously not created in the lab. Biologists also depart from observations not created by them. Psychologists seem to do both. Experimental psychologists create what they try to explain theoretically, but probably the most part of psychology is still observational, not experimental.

Hacking’s pragmatist approach is right in highlighting that a central characteristic of scientific research is that it is a *technical* procedure. But it is not true that the essential aspect of this technic is that scientists always create the phenomena they intend to explain; in observational studies (cohort studies in clinical medicine, for example), scientists do not create the phenomena, even though they do not simply observe or find them in the world: they systematically discriminate the phenomena they want to explain.

There is something important here. Experiments are not *the* essential component of scientific enterprise, even though it is true that science is not only a “reasoning” procedure, that is, it is not pure *noetic* activity (as it seems to be the case for traditional philosophy, at least for conceptual analysts). Science is also technic, procedure, that is, *poiesis*. Besides experiments, there are also observational procedures that we accept as scientific; so the essence of scientific method must be something common not only to the experimental but also to the observational variant.

V. Inductive Behavior and modern Evidence-Based Practices

There is, however, something right in the pragmatic approach. Perhaps Neyman’s approach to science as an “inductive behavior” is a clue

(NEYMAN 1957). “Induction” for Neyman is not a different kind of reasoning, neither does it involve any special kind of logic; what marks induction is a special kind of attitude to the process of getting knowledge. Which is the common characteristic of the Neumanian “inductive” attitude?

In order to understand this, let’s separate four different varieties of research methods. It is commonly accepted that all empirical researches can be divided into four classes: the descriptive, the analytical, the observational and the experimental.

Take the example of modern epidemiology. Clinical epidemiologists usually discriminate between two different kinds of studies of medical interventions they call “experimental” and “observational”, which differ in scientific strength and feasibility (see the figure below). Epidemiologists say that analytical (that is, comparative) studies can be either experimental or observational. In observational studies researchers only observe what happens with patients who, for various reasons, do or do not get exposed to an intervention. Observational studies can be “analytical”, that is, comparative; or “descriptive”, that is, non-comparative. Statistical tests were developed in order to permit reliable comparisons. The main example of an analytical observational study is a cohort study. Observational studies are more feasible than experimental studies, but they are more prone to systematic errors. Descriptive studies are not analytical, that is, they do not have hypotheses that can be statistically tested. They comprise only descriptions of observations, data, or their frequencies:

	EXPERIMENTAL	OBSERVATIONAL
ANALYTICAL	RCT QUASI-EXPERIMENTS	<ul style="list-style-type: none"> •COHORTS •CROSS-SECTIONAL COMPARATIVE (TRANSVERSAL) STUDIES •ECOLOGICAL STUDIES •SERIES OF CASE-CONTROLLED STUDIES
DESCRIPTIVE	NONE	FREQUENCY STUDIES (INCIDENCE, PREVALENCE OR CROSS-SECTIONAL STUDIES) CASE REPORTS ETNOLOGICAL STUDIES

Fletcher & Fletcher explain that experimental clinic-epidemiological studies are in fact a special kind of cohort studies in which the conditions of study (that is, selection of study groups, nature of interventions, management during follow-up and outcomes) are specified by the investigator “for the purpose of making unbiased comparisons” (FLETCHER & FLETCHER 2005, 128). So, in experimental trials, “the investigators are conducting an experiment, analogous to those done in the laboratory” (also WALLMAN 2001, 92–3).

Experimental trials are population studies. Population studies are studies of multiple observations (note that ‘population’ here does not apply only to living individuals). But note that not all experiments are population studies. Lab experiments in physics are not necessarily population trials. Let’s return now to the view advanced by Hacking that “experiments are the creation of phenomena” (HACKING 1983, 229), that is, “to experiment is to create, produce, refine and stabilize phenomena” (HACKING 1983, 230). As said above, this is persuasive. Note, however, that this applies to paradigmatic experiments in physics and chemistry, but not to other kinds of good and

legitimate empirical researches. For example, this is not true in the case of observational studies, and this is only partially true in the case of population trials (for the phenomena under study occur in natural individuals). So, unless we think that only physics is a legitimate representative of science, Hacking's approach is partial and defective.

We are in effect lacking a good approach to the unity of scientific method. Maybe the idea of unity in science is a myth, for in fact we have in science a variety of study designs. Each design seems to be adequate to certain fields and intents. As done by epidemiologists (in so-called evidence-based medicine), we can draw a pyramid of the different study designs and classify them by their evidentiary strength and reliability. Maybe experiments are at the top of this pyramid; they are arguably the most reliable kind of research method. But in several domains the best "inductive practice" is not the "experimental". In clinical medicine, for example, it is sometimes better to opt for a cohort study, a non-experimental study, and in some circumstances cross-sectional studies are good enough; in psychology experiments are sometimes simply impossible to be made, so observations are the best choice for reasons of feasibility. Science must develop even with imperfections, but this is not a reason to consider those studies with lower inferential power defective. However, it is important to calculate those imperfections in order to relativize the conclusions of each research.

VI. *Experimental philosophy?*

"Experimental Philosophy" is a brand name; anyway, X-philosophers claim to be conducting literal experiments – or at least this is what experimentalists pretend to be doing. However, why not "observations"? Why not "qualitative research" (WOOLFOLK 2013, 82)? For experiments are difficult to perform; observations are more feasible and sometimes they are

even more reliable and informative, especially in the human sciences. As for qualitative research, it is patent that presently they evolved methodologically to the point of being considered by many as the preferred method in the humanities. In fact, choosing the experimental method, experimentalists have become a target of criticisms from several parties.

Let us consider Knobe's study on intentions (KNOBE 2003b). Is it really a typical example of an experiment? In fact, it is what clinical-epidemiologists would call a "quasi-experiment". Quasi-experiments share similarities with standard experimental designs (that is, randomized controlled trials), but they differ because they lack the requirements of randomization. This is not explicit in the text of the study, but it seems clear that in Knobe's study the sampling was chosen by convenience. *Convenience sampling* is a sampling technique where subjects are selected because of their convenient accessibility and proximity to the researcher.⁷ This is a method widely used in pilot studies; but Knobe's study was not intended to be a pilot study. In fact, Knobe did not clarify all the study methods. He only writes that, in order to assess the issue (viz. whether people's judgments about intentionality depend in part on the moral opinion they have about the behavior of the observed subjects – we can assume this is the study hypothesis), he conducted "a simple experiment." So the human subjects were "80 people spending time in a Manhattan public park." We take as obvious that the method of sampling was by convenience.

Anyway, how were the subjects selected? Knobe did not mention anything about that. So another important question is: of which population is this sample intended to be representative? Can it be inferred that the population is comprised of the totality of adult human beings (non-cognitively deficient) or, as some could also conclude rather ironically, is its population the

⁷ Woolfolk calls convenience sampling "opportunity sampling" (WOOLFOLK 2013, 82).

sum of all WEIRD people (that is, western and educated, from industrialized, rich and democratic countries)? We could ask those questions, for the author does not present any information about sex or age, as we also do not have information about the criteria for subject inclusion or exclusion in the experiment. Moreover, the study is certainly not blinded. Hence, there are good reasons for being cautious in interpreting the study results.⁸

These methodological inquiries launch doubts about the internal validity and reliability of EP studies (WOOLFOLK 2013, 80-1). Philosophers sympathetic to EP and IP cannot simply turn a blind eye to the criticisms of X-phi researches; and they seem to be right in stressing the limitations of X-phi experiments in the field of normative ethics. Take, for example, Ernst Sosa. In an interview, he argued that “[t]he experimentalists can’t prove that their subjects are actually responding to the philosophical principles at stake.” And one reason is that “people may be distracted by certain language choices or irrelevant details in the scenarios being posed” (see LACKMAN 2006). One could interpret this criticism as saying that subjects’ responses in studies such as those conducted by Knobe can be biased, and hence cannot be reliable (WOOLFOLK 2013, 82). But note that is not in fact a criticism from an OCA’s point of view. In other words, this is not a criticism from an orthodox armchair stance; this is a general methodological criticism. The criticism is this: X-phi studies have internal methodological flaws inherent to their study designs. This criticism does not imply that empirical approaches are equivocal “in principle”, even less that empirical findings are irrelevant to philosophical inquiry. The criticism also does not imply that philosophers with strict

⁸ Woolfolk mentions Guglielmo & Malle paper in order to support his criticism of Knobe’s methodological choices. In a series of six experiments, Guglielmo & Malle claimed that there is a confound in Knobe’s original comparison given that, in Knobe’s study, indifference to negative outcomes is viewed as more indicative of desire than is indifference to a pro-social outcome. Moreover, “when given opportunities to select among manifold descriptions of the actor’s attitude toward the actions, participants rarely referred to the harming of the environment as “intentional,” preferring instead the term “willingly” or “knowingly” (GUGLIELMO & MALLE 2010).

philosophical intent could never perform well-designed experimental approaches. The criticism points in another direction: if philosophers want to conduct empirical studies, they must pay rigorous attention to the methodological principles commanding researches in the empirical sciences.

Consider Woolfolk's main criticism of "experimentalism" in philosophy. According to Woolfolk, "experiments conducted by experimental philosophers frequently fail to meet the methodological standards that are articulated by the experts on research design in those fields they would emulate" (WOOLFOLK 2013, 80). This is exactly the same opinion we have and want to emphasize:

The biobehavioral sciences have developed, via dialectic and trial and error, techniques for avoiding threats to sound inductive reasoning. Strict adherence to those methods can minimize the fallacies that so often plague not only the everyday judgments of ordinary people but also the inferences that trained scientists draw from systematically collected data. Indeed, philosophers examining a textbook on experimental design can rather quickly recognize various devices that are practical applications of Mill's Methods (WOOLFOLK 2013, 80).

The same can be said not only of Mill's Methods but also of the general descriptions of the "logic" of scientific method made by philosophers of science. Undergraduate philosophy of science is not a guide to understand the complexities of present methods in the different fields of scientific research.

VII. Why not interdisciplinary research?

One obvious conclusion we can reach now is that philosophers cannot overlook the methodological questions in empirical science, and this is especially important for those that consider themselves naturalists. If philosophers want to make empirical researches, it is mandatory to master the

subtleties of the methods of the many different “inductive practices”. And if Harman is right that philosophical inferences are also IBE (as we agree), then conceptual “analysis” certainly is, and should be, a part of any good empirical inquiry. This in fact opens the case for a preference for interdisciplinary empirical studies philosophically oriented, over philosophical “experimentalism”.

We hope it becomes clear that we are not against “experiments” in philosophy; far from it. But we should be more methodologically cautious if we want to make philosophical researches scientifically responsible. We should take scientific methodology very seriously, and this involves avoiding methodological simplifications and a careful respect of the appropriate methodological guidelines for each different type of empirical research. In effect, it is wiser not to travel alone in this new crossing, but to do so in the company of the more experienced. And this is a sensible advantage of IP over EP.

REFERENCES

APPIAH, Kwame Anthony. *Experiments in Ethics*. Cambridge, Massachusetts: Harvard University Press, 2010.

_____. The New New Philosophy. *New York Times*, December 9, 2007. At <http://www.nytimes.com/2007/12/09/magazine/09wwln-idealabt.html?_r=2&>. Accessed 03/08/2015.

BERNSTEIN, Stephen & BERNSTEIN, Ruth. *Schaum's Outline of Theory and Problems of Elements of Statistics I: Descriptive Statistics and Probability*. Schaum's Outline Series. New York: McGraw-Hill, 1998.

BERNSTEIN, Stephen & BERNSTEIN, Ruth. *Schaum's Outline of Theory and Problems of Elements of Statistics II: Inferential Statistics*. Schaum's Outline Series.

New York: McGraw-Hill, 1999.

CULLEN, Simon. "Survey-Driven Romanticism." *Review of Philosophy and Psychology* 1, 2010: 275–296.

FISHER, Ronald Aylmer. "The Place of the Design of Experiments in the Logic of Scientific Inference." *Sankhyā: The Indian Journal of Statistics, Series A*, 27 (1), 1965: 33–38.

FLETCHER, Robert & FLETCHER, Suzanne. *Clinical Epidemiology, The Essentials*, 4th edition. Philadelphia, Pennsylvania: Lippincott Williams & Wilkins, 2005.

GLYMOUR, Clark. "Explanation, Tests, Unity and Necessity." *Nous* 14 (1), 1980: 31–50.

_____. *Theory and Evidence*. Princeton, NJ: University Press, 1980.

GOLDMAN, Alvin I. "Naturalistic Epistemology and Reliabilism," *Midwest Studies in Philosophy*, 19, 1994: 301–320.

GUGLIELMO, Steve & MALLE, Bertram F. "Can Unintended Side Effects Be Intentional? Resolving a Controversy over Intentionality and Morality." *Personality and Social Psychology Bulletin*, 36, 2010: 1635–1647.

HACKING, Ian. *Representing and Intervening. Introductory Topics in the Philosophy of Natural Science*. Cambridge, UK: Cambridge University Press, 1983.

HARMAN, Gilbert. "The Inference to the Best Explanation." *The Philosophical Review*, 74 (1), 1965: 88–95.

HAUG, Matthew C. (Org). *Philosophical Methodology: The Armchair or the Laboratory?* London, UK: Routledge, 2013.

HEMPEL, Carl G. "Studies in the Logic of Confirmation." *Mind*, New Series, 54 (213), 1945: 1–26.

_____. *Philosophy of Natural Science*. New Jersey: Prentice-Hall, 1966.

HOLTON, Richard. "Norms and the Knobe Effect." *Analysis*, 70 (3), 2010: 1–8.

KNOBE, Josha & NICHOLS, Shaun. *Experimental Philosophy*. Oxford, UK: Oxford University Press, 2008a.

_____. “An Experimental Philosophy Manifesto,” in Knobe, Joshua & Nichols, Shaun (eds.), *Experimental Philosophy*. Oxford, UK: Oxford University Press, 2008b, p. 3–14.

KNOBE, Joshua; BUCKWALTER, Wesley; ROBBINS, Philip; SARKISSIAN, Hagop; SOMMERS, Tamler & NICHOLS, Shaun. “Experimental Philosophy.” *Annual Review of Psychology*, 63 (50), 2012: 72–73.

KNOBE, Joshua. “Intentional Action and Side Effects in Ordinary Language.” *Analysis*, 63, 2003a: 190–193.

_____. “Intentional Action in Folk Psychology. An Experimental Investigation.” *Philosophical Psychology*, 16 (2), 2003b: 309–324.

_____. “What is Experimental Philosophy?” *The Philosophers’ Magazine*, 28, 2004. At <<http://pantheon.yale.edu/~jk762/ExperimentalPhilosophy.pdf>>. Accessed 03/08/2015.

LACKMAN, Jon. The X-Philes: Philosophy Meets the Real World, *Slate*, March 2, 2006. At <http://www.slate.com/articles/health_and_science/science/2006/03/the_x_philes.html>. Accessed 03/08/2015.

LEHMANN, Erich Leo. “The Fisher, Neyman-Pearson Theories of Testing Hypotheses. One Theory or Two?” *Journal of the American Statistical Association*, 88 (424), 1993: 1242–1249.

LEITER, Brian. “Naturalism and Naturalized Jurisprudence,” in Brian Bix (ed.), *Analyzing Law: New Essays in Legal Theory*, Oxford: Clarendon Press, 1998.

_____. *Objectivity in Law and Morals*. Cambridge Studies in Philosophy and Law. Cambridge, UK: Cambridge University Pre, 2007.

_____. “Naturalism in Legal Philosophy,” in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, 2014. At

<<http://plato.stanford.edu/archives/fall2014/entries/lawphil-naturalism/>>.

Accessed 04/12/15.

LIPTON, Peter. *Inference to the Best Explanation*. London, UK: Routledge, 1994.

LOUGHLIN, Michael. "Criticizing the data. Some Concerns about Empirical Approaches to Ethics." *Journal of Evaluation in Clinical Practice*, 17, 2011: 970–975.

MAYR, Ernst. *One Long Argument*. Cambridge, MA: Harvard University Press, 1991.

NEYMAN, Jerzy S. & PEARSON, Egon S. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231, 1993: 289–337.

NEYMAN, Jerzy S. "'Inductive Behavior' as a Basic Concept of Philosophy of Science." *Revue de l'Institut International de Statistique/Review of the International Statistical Institute*, 25 (1/3), 1957: 7–22.

POPPER, Karl. *The Logic of Scientific Discovery*. Routledge Classics, [1959] 2002.

RAILTON, Peter. "Naturalism and Prescriptivity." *Social Philosophy and Policy*, 7, 1989: 151–174.

_____. "Naturalism and Prescriptivity," in E.F. Paul, *et al.* (eds.), *Foundations of Moral and Political Philosophy*, Oxford: Blackwell, 1990.

SMITH, Barry C. "Same Compounds: Different Flavours?" *Proceedings of Wine Active Compounds*, 2008: 98–102.

_____. "Philosophical and Empirical Approaches to Language." In Haug, Matthew C. (Org). *Philosophical Methodology: The Armchair or the Laboratory?* London, UK: Routledge, 2014. Chapter 16.

SOBER, Elliott. *Philosophy of Biology*. Boulder, Colorado: Westview Press, 2000.

_____. *Evidence and Evolution: The Logic behind Science*. Cambridge, UK: Cambridge University Pre, 2008.

SPENCE, Charles; CHARLES, Michael & SMITH, Barry C. “Airplane, Noise and the Taste of Umami.” *Flavour*, 3, 2014: 2–4. At <<http://www.flavourjournal.com/content/3/1/2>>. Accessed 03/08/2015.

TARSKI, Alfred. *Introduction to Logic and to the Methodology of Deductive Sciences* [1941]. Mineola, New York: Dover Books, 1995.

Van FRAASSEN, Bas C. “Constructive Empiricism Now.” *Philosophical Studies*, 106, 2001: 151–170.

_____. *The Scientific Image*. Oxford, UK: Oxford University Press, 1980.

_____. *The Empirical Stance*. New Haven, CT: Yale University Press, 2002.

WOOLFOLK, Robert L. “Experimental Philosophy: A Methodological Critique.” *Metaphilosophy*, 44 (1–2), 2013: 79–87.