Theory and Basic Research in Software Engineering ⋆

Alberto Sampaio and Isabel Braga Sampaio

Polytechnic Institute of Porto/Superior Institute of Engineering
{acs,ais}@isep.ipp.pt

Abstract. Empirical research is now common in the software engineering field. This paper aims to show the need for a reorientation of empirical research in software engineering through theory development and basic research. The importance of both for software engineering in the current development of empirical software engineering is explained and their relationship emphasized.

1 Introduction

The choice of software engineering (SE) techniques and tools should rest on facts and not just on opinions [1]. This leads to an empirical software engineering (ESE) or evidence-based software engineering and is an important step toward a SE discipline with scientific bases. Empirical studies in SE are being carried out for many years but they were not always so common [1, 2].

Meanwhile the situation changed and empirical studies are now frequent. Despite that change, rules of thumb are still frequent in software engineering, which is also replete with announcements of new methodologies, products, methods and techniques, these being frequently presented as some kind of panaceas to the respective areas. This situation can be seen as paradoxical. If in general there are many empirical studies, simultaneously we can not find a conducting wire for all that research, or a logical model, from where the inquiry might emanate and to where inquiry feedback can return.

In this paper it is argued that: it is theory that gives coherence to isolated SE empirical facts and that can help to explain the previous apparent paradox; and that this is also linked to nature of the accomplished research. Particular attention will be devoted to the second because the first is receiving more attention recently (e.g. [3]). In this paper we are chiefly concerned with the SE phenomena involving people, which require to be tested through empirical studies.

This paper is organized as follows. After this introduction, sections 2 and 3 discuss theory and kinds of research respectively. Section 4 describes the relevance of basic research and section 5 lists some issues about theory and basic research adoption in SE. Section 6 presents some conclusions.

2 Theory

Science is the pursuit of truths about the natural world usually expressed at an explanatory level [4]. With theory, researchers seek the explanation of phenomena and their prediction [5]. Thus, it can be said that the goal of science is the development of theory [5]. Since Galileo, theory has been a fundamental component of scientific research. With this relation between science and theory, then the word "scientific" would be better used to mean something that is related with theory. Recently, the importance of theory is been recognized in SE (e.g. [18]). Because of that and lack of space we are not going to reiterate and reinforce the importance of theory in SE.

Generically, a theory is a group of related propositions with the purpose of specifying the relationships among a group of concepts on a limited range of phenomena [5], from which the SE is part. In SE the propositions would be empirical statements because the variables would depend on empirical (real) world. A statement can be considered for theory when it tries to explain or predict [6].

Propositions relate concepts, so theory gives coherence to the set of concepts and, in last instance, to empirical research, that involves an operational form of those concepts. In this way, empirical results (facts) depend on a theory to contribute to knowledge growth.

The capacity of a theory in tie facts means that research based on theory would not produce isolated pieces of information, on the contrary, such information can be framed in a framework relating many studies. So, a theory contributes to a more global (systemic) view of the existing empirical information about a certain group of phenomena and, in this way, theory favors the systematization of existent scientific knowledge and the accumulation of new. This should be interpreted in an broad sense, including the appearance of alternative theories.

The researchers concerned with theory should have a specific attitude about empirical research. To these researcher is not enough to obtain the results. For them, to interpret results and find explanations for the experimental results using theory and possibly the inverse, see the theory in the light of empirical results, is a fundamental step in a empirical study. This can also take. In this way they can develop a more critical and clarified attitude about the SE field.

3 Kinds of Research

A question is sufficiently important to justify research when it seeks to satisfy a scientific need, or a social need [7] . A question satisfies a scientific need if its answers allow to increase the theoretical understanding of some phenomenon, making advance theory; and satisfies a social need, if it can help to solve some problem faced by the society or an organization [8]. Thus, two kinds of important questions can be recognized. Hypotheses for research can appear through theory, common sense, and previous research [9], where common sense should be interpreted in the context of the practice of software engineering, that is, not only from theory. Questions of the first kind then correspond to basic (or fundamental) research, while the ones of the second kind correspond to applied research [8]. Accordingly, research to be considered basic, should be based on theory and must provide feedback to the theory. This is not the case with applied research. The two kinds of research are now defined:

Definition 1 *Basic research - it is research done in order to understand the phenomena and making advance theory, with little or no concern with the application of the results, ie, without practical goals. It should satisfy the following two principles: 1) be based on theory; 2) possess a feedback loop feeding theory.*

Definition 2 *Applied research - it is research that seeks to supply information that can lead to the solution of practical problems, and which is expected to contribute to the improvement or progress of some process or activity, or the reach of practical goals. And the inclusion of theory is not mandatory.*

The criteria used to distinguish research were intention and theory. It is possible to subdivide research in other ways and detail. However, for the present discussion the splitting in two kinds of research is just enough. Although that separation seems clear, in practice this is not always so. Again, this is no impediment for the present discussion. It is logical that the level of theory in basic research can be used as an indicator of basic research quality.

4 Relevance of Basic Research

The lack of basic research is a matter of concern for at least three reasons: firstly because basic research has as its purpose the understanding of a phenomenon; secondly because fundamental research is considered of the largest importance in the innovation of a knowledge area; finally because it allows significant advances in the long-term by opening research avenues for applied research. The second reason can be easily justified by the natural curiosity underlying basic research.

There is the view that pure research, contrary to applied research, should not be practiced in software engineering because it does not allow the study of the important aspects (practical application) of SE. Knowing that SE researchers have expended so many efforts to study accidental properties of software [10] mainly using applied research, becomes clear that such view is questionable. If it is basic research that satisfies scientific needs, then to find the essential properties and processes of SE more basic research would be needed.

As much as it was possible to search, few studies exist about the influence of the two kinds of research in the progress of a discipline. A classic reference is the study by Comroe and Dripps [11] where it is showed the largest importance of basic research for the advance of the medicine practice. There are, however, interactions between the two kinds of research, making to think that both are indispensable and not mutually exclusive. Basic research also involves risks and costs, but it is a positive investment with returns in the long-term (examples e.g. in [12, 13]). There are several studies showing the economic benefit of basic research. For example, the review in [13] confirms that basic research is a good investment, even if difficult to measure.We are not aware of such studies in SE.

When all software engineers perform the same kind of research there is no distinction between the activity of a SE researcher, possibly from a research lab, and the activity of SE practitioners that research their solutions just to see if they work. Moreover, there is no distinction between the kinds of research developed among the labs themselves. This is not done to classify labs, but the kind of research they perform can distinguish labs. If a research project/lab is claimed to be of fundamental kind, then it must present a strong theoretical basis in its research. This clarification could be relevant when applying for financial grants. The public funding of basic research should be done carefully to prevent adverse effects to society and science [14].

The proliferation of publications in line with the large science can limit great science and good ideas can pass unnoticed [15]. This also happens in computer science according David Parnas [19], although he presents the problem from the "need to publish" perspective. Because basic research is not forced to immediate applicability, a reorientation towards it can reduce those problems.

5 Issues Adopting Theory and Basic Research in SE

For more than two decades authors are calling attention to the "people problems" in SE (p.557) [16], which includes management and organizational issues. Also, SE is not a science and in software the level of control is very distinct from the level found in the physical sciences. Furthermore, it is accepted that software engineering differs from more traditional engineering. Software engineering is, by definition, an applied discipline, and seen to a certain extent as applied computer science. Theory development in an applied field is a "demanding and difficult" task [17]. To conclude, the adoption of theory formation approaches based on physical laws cannot be done automatically.

The covering-law model for explanation could not overcome certain issues pointed out by their critics and several researchers proposed alternative explanation models [21] for example based on counterfactuals analysis. Suppose it is known that when a team used the SEMAT essence framework, the team succeed - for the sake of simplicity succeed will mean whatever the reader wants. A possible counterfactual question would be, what would happen if the team did not use the framework, would it succeed or not?

Traditionally, generalizations in explanation are either laws or purely accidental. But the distinction between laws and accidental generalizations has proven to be very difficult. Traditionally, it was done by verifying how much a proposed law complies with law definition. Outside the natural sciences the explanatory generalizations do not fit any of those two possibilities. According Woodward [23] there is a middle ground between these two possibilities. That is, there is an intermediate region where some generalizations cannot be seen as laws because they are only accidentally true but not purely accidental.

Counterfactuals play a fundamental and valuable role in some causal models (see e.g. [20, 23]). It is also possible to extract predictions from a counterfactual. For the predictions of a counterfactual statement to be valid the laws and the boundary conditions must remain invariant for a certain period. However, even when prediction is not assured, counterfactuals can have explanatory value [20].

Experiments (followed by quasi-experiments) are the most adequate research method among all scientific methods for the study of causal relationships [22]. They are also superior even when the effect is not "directly" manipulable, by allowing higher counterfactual inference than non-experiments. Unfortunately, they are limited because [22, p.9] "knowledge of the effects of manipulable causes tells nothing about how and why those effects occur." In other words, experimentation is suitable to obtain "causal descriptions", but less suitable to "causal explanations" requiring more effort [22].

If causal explanation is so important for science and so for basic research, and knowing that experimentation is well suitable to show causal relations but not so much to show simultaneously the mechanisms of that causal relation, the question is how can we proceed to uncover these mechanisms. That task is called "detective work" by Scriven [22] and requires intuition. According [22] it requires decomposition of causes.

It is well-known that performing experimental studies in SE engineering is a difficult task and consequently it is also difficult to obtain causal descriptions and their generalization. Being so, then all the experimental effort should be properly rationalized and, as explained before, theory can be of great value by organizing the results of empirical studies.

6 Conclusions

This work is mainly concerned with SE research when involving people. The thesis (hypothesis) here expressed were: (1) the need for theory development that could guide research and contribute to the integration and systematization of empirical evidence; (2) and, concomitantly, the need of more fundamental research. That is, the success of theory development, from (1) depends on (2), a reorientation of a part of empirical research towards fundamental research and vice-versa. Can also be concluded that: *To endorse theory development and simultaneously negate basic research can be seen as a contradiction.*

The paper also presented issues that should be considered when developing theory and basic research in SE.

This paper did not propose the rejection of research performed without an explicit theoretical source because hypotheses for research can appear from outside of theory. Additionally, a concern with fundamental research does not imply that all research should be fundamental. However, we are not able to specify any percentage as a guide and a deeper analysis should/will be done.

References

[1] Harrison, W., Basili, V.R., Editorial, Empirical Software Engineering, Vol.1, N.1, pp.5-10, Kluwer Academic Publishers, 1996.

[2] Zelkowitz, M.V., Wallace, D.R., Experimental Models for Validating Technology, IEEE Computer, pp.23-31, May, 1998.

[3] Johnson, P., Ekstedt, M., & Jacobson, I., Where's the Theory for Software Engineering?, Software, IEEE, 29(5), 96-96, 2012.

[4] Sachs, Mendel, Concepts of modern physics: the Haifa lectures. Imperial College Press, Distributed by World Scientific, 2007.

[5] Anderson, D.C., Borkowski, J. G., Experimental Psychology: Research Tatics and Their Applications, Scott, Foresman and Company, Illinois, USA, 1978.

[6] Bailey, K.D., Methods of social research; 4th ed., Free Press, 1994.

[7] Manheim, J.B., Rich, R.C., Empirical Political Analysis, 4th ed., Longman, 1994.

[8] Miller, D.C., Handbook of Research Design and Social Measurement, 5th ed., Sage Publications, 1991. (1st ed. in 1964.)

[9] Mitchell, M., Jolley, J., Research Design Explained, 3rd ed., Harcourt Brace College Publishers, Holt Rinehart and Winston, Inc., 1996.

[10] Brooks, F.P., The Mythical Man-Month, 20th anniversary edition, Addison-Wesley, 1995.

[11] Comroe J.H., Dripps, R.D., Scientific Basis for the Support of Biomedical Science, Science, April, pp. 105-111, 1976.

[12] Remedios, C., The Value Of Fundamental Research: A discussion paper prepared for the IUPAB Council, IUPAB, 2006. (URL: http://iupab.org/publications/value-of-fundamental-research; last accessed January, 2013)

[13] Salter, A.J., Martin, B.R., The economic benefits of publicly funded basic research: a critical review, Research Policy, Volume 30, Issue 3, 1, Pages 509-532, March 2001.

[14] Nelson, R.R. Reflections on The simple economics of basic scientific research: Looking back and looking forward. Industrial and Corporate Change, OUP, 15.6: 903-917, 2006.

[15] Popper, K.R., The Myth of the Framework: In Defence of Science and Rationality, (Edited by MA Notturno), 1994.

[16] Seaman, C.B., Qualitative Methods in Empirical Studies on Software Engineering, IEEE Transactions on Software Engineering, pp.557-572, Vol.25, N.4, July/August, 1999.

[17] Lynham, Susan A., Quantitative Research and Theory Building: Dubins Method, Advances in Developing Human Resources, 4: 242-276, August 2002.

[18] GTSE 2014: Proceedings of the 3rd SEMAT Workshop on General Theories of Software Engineering. ACM, New York, NY, USA (2014)

[19] Parnas, D.L.: Stop the numbers game. Commun. ACM 50(11), 19–21 (Nov 2007)

[20] Pearl, J.: Causality: models, reasoning and inference, vol. 29. Cambridge Univ Press (2000)

[21] Pfeifer, J., Sarkar, S.: The Philosophy of Science: An Encyclopedia. Psychology Press (2006)

[22] Shadish, W.R., Cook, T.D., Campbell, D.T.: Experimental and quasi-experimental designs for generalized causal inference. Wadsworth Cengage learning (2002)

[23] Woodward, J.: Making things happen: A theory of causal explanation. Oxford University Press (2003)