Paper

Active – Passive: On Preconceptions of Testing

Krzysztof M. Brzeziński

Institute of Telecommunications, Warsaw University of Technology, Warsaw, Poland

Abstract—In telecommunications and software engineering, testing is normally understood to be essentially *active*: a tester is said to stimulate, control, and enforce. Passive testing does not fit this paradigm and thus remains the niche research subject, which bears on the scope and depth of the obtained results. It is argued that such limited understanding of testing is one of its many community-bound preconceptions. It may be acceptable in the current engineering approach to testing, but can and should be challenged in order to converge on the core concepts of the proposed science of testing ("testology"). This methodological work aims at establishing that there are no fundamental reasons for admitting the dominant role of the active element in testing. To show this, external (also extra-technical) areas are consulted for insight, direct observations, and metaphors. The troublesome distinction between (passive) testing and monitoring, as well as unclear relations between testing and measurements, are also addressed.

Keywords—behavior, development, metrology, monitoring, passive testing, reactive systems, Scientific Method.

1. Introduction

Testing is intertwined with the development (creation, construction, and further use) of artifacts – intentionally designed objects. Artifacts of a certain complexity are "mechanically" referred to as artificial *systems*. We understand testing as the umbrella term for a particular set of concepts, methods, and techniques of verification and validation (V&V), i.e., assessing whether a system is *correct* w.r.t. a given notion of correctness. This assessment leads, pragmatically, to deciding whether a system is *acceptable*.

Testing cannot be replaced by other, "non-testing" V&V techniques. Placed in a loop of development activities, testing is a vital element in achieving and maintaining correctness (and thus quality) of systems. The complexity of testing, however, is known to grow *exponentially* in the complexity of tested systems. Accordingly, despite the undisputable improvements in testing concepts and techniques, spectacular system failures (including those that entail loss of life), attributed to inadequate testing, still happen. In order to sustain the pace of development of complex systems (including telecommunications systems), it is necessary to seek improvements in testing *beyond* its current, relatively steady development. The aim of this work is to contribute towards this end. Its underlying assumptions and theses are briefly presented below (see [1]–[5] for discussion).

Testing is currently researched and practiced mostly by specialized groups, or schools, within separate communities concerned with particular classes of systems to be tested. The immediate context of this work are systems characteristic of information and communications technology (ICT) – a field defined by convergence of traditional *telecommunications* and informatics (*software engineering* and *computer science*). The convergence of concepts and approaches to the development (and thus – also to testing) of ICT systems is far from complete. It thus makes sense to refer, within ICT, to separate communities of *software* testing, *protocol* testing, *circuit* testing, etc. There are also groups concerned with testing outside ICT (*chemical* testing, *material* testing, etc.), with their own, important insight.

Testing communities speak particular languages (or, to quote Wittgenstein, they play different *language-games*). They are reluctant to borrow the concepts and terms from peer groups. Consequently, any *preconceptions* they may have on testing cannot be easily confronted with other patterns of understanding, and are very hard to uproot (even if they clearly form a crippling self-restriction). This phenomenon of conceptual and linguistic (terminological) "lock-in" has been noticed, e.g., by Lamport [private communication, 2010], who called it a "Whorfian syndrome".

Testing schools tend to follow the *engineering* approach, with its *apprentice* tradition of vocational study. The testing concepts and terms are defined stipulatively, to mean what a given community *wants* them to mean. This particular understanding, as well as skills for its practical use, are taught, and then checked during exams that lead to obtaining professional titles of a "certified tester" or the like. Accordingly, there are sources of community-bound "standardized knowledge" of testing [6]–[11]. There is, however, no common definition of testing that would be accepted *across* the testing communities. Lack of such definition indicates a serious problem with testing, as "*a definition influences future perceptions – a too narrow or misleading one may block future investigations for a long time*" [12].

In order to gain new perspectives, impetus, and funding, testing needs to properly address the issues identified above. To do so, it should transcend the limitations imposed by the apprentice model, and establish itself as a *science*, with academic recognition. This seems necessary not only for immediate professional application of testing, but also for its proper *teaching*, in a way that avoids seeding and perpetuating the existing preconceptions in the new generation of researchers – a concern that is not unique to testing [4]. This path has already been taken by *metrology* – the "*sci-*

ence of measurement and its application" [13, 2.2]. For the postulated new testing science, of the *design science* kind, we propose the name: "Testology". It would be the producer and bearer of *first principles* and *core concepts* of testing, regardless of its area and context of application, and would also allow forming the "applied testing" subsets and specializations, with meaningful relations to each other. Just as any other (design) science, testology is free to seek linguistic and conceptual *metaphors* [14] without any *a priori* restriction of the range of possible "donors", and to look into languages currently spoken by particular testing schools, in the hope that "*a core theory... can be synthesized from writings across a number of disparate fields*" [15]. Also the general understanding of "testing", encoded in everyday language and reflected in dictionary entries, should not be neglected. In this context, the existing sources of vocational knowledge on testing, including the definition(s) of testing contained there, are only one of the inputs available for consideration, and not *the authoritative* body of the concepts of testology.

To illustrate the postulated approach to testing, in the sequel we focus on a single idea that currently prevails in virtually all ICT-related testing schools, namely, that *testing is active*. We argue that it is a community-bound preconception – a conventional disciplinary modification of the concept of testing, which is not substantiated by any "deeper roots" of emerging testology. We further argue that sticking to this preconception is an unnecessary handicap for applied testology – valuable research on *passive testing* is currently conducted away from the mainstream, in a *niche* research area, which bears on the breath, depth, and consistency of obtained results. We claim that abandoning the "active" preconception should bring more consistency to the mainstream of testing research, by allowing the uniform treatment of both active and passive testing, which in turn may contribute to the postulated "nonlinear" improvement in testing.

2. Status of Active and Passive Testing

In telecommunications, software engineering, and most other technical disciplines, the prevailing intuition of testing is reflected in its operational characterization as an activity (stating what is being *done* while testing), in which a tester:

- generates and *applies* (sends) stimuli, or "test input data", in order to *control* a system under test (**Sut**) – to provoke and guide phenomena (in our case – mainly behavior) to be investigated by testing;
- *observes* phenomena as they appear under the influence of applied stimuli;
- *analyzes* the relation between observed phenomena and some reference (such as a pre-computed, intended behavior);
- *decides* on a suitable verdict, which expresses the assessment made.

This operational characterization is often taken as the *operational definition* of testing: all the enumerated operations are quite tangible, and their joint presence is said to constitute what shall (and, by complement, what should not) be regarded as testing. This characterization is then implicitly employed in the role of the definition of testing (as in [16, pp. 14–16], where, on 600+ pages, no other *explicit* definition of testing is given), or is suitably rephrased, as in [17]: "*The principle of testing is to apply inputs... and to compare the observed outputs to expected outputs*". Similar definitions[1], in various wording, prevail in "official", vocational compendia, dictionaries of terms, and meta-standards, and are also cited in research papers. Their common element is that they stipulate the *active* character of testing – a tester controls, solicits, and enforces. Active testing constitutes the mainstream of testing.

On the other hand, since the early 1980s there has been ongoing interest in *passive testing*, technically defined as a testing activity in which a tester does not influence (stimulate) a **Sut** in any way – it does not apply any test stimuli. Two typical approaches to such testing may be identified.

One party claims that the active character of testing reflects its *essence*, and thus cannot be surrendered. It is natural for this party to maintain that "passive testing" simply does not respond to the concept of testing – that it is a spurious interpretation, a mere *façon de parler*, or the case of confusion of tongues. Indeed, passive testing has not been identified as a dimension of the discourse space of testing, nor even alluded to, in the annotated bibliography of formal testing [22], the proceedings of the prestigious Dagstuhl Seminar on testing [23], taxonomies developed to get insight into the notion of testing [24], [25], standardized glossaries of terms pertaining to testing [6], [7] or broader software engineering activities [19]. It is also, apparently, not covered by the new, forthcoming international software testing standard ISO 29119. The telecommunications-oriented methodology of *conformance testing* [26] openly excludes passive testing from its scope. The standardized test language TTCN-3 [27] was meant to express active tests, and there have been very few proposals to re-use it also for passive testing [28].

The other party investigates passive testing basing on its arbitrarily adopted technical definition, without any deeper concern for methodological harmonization with active testing. This is how the majority of valuable results on passive testing have been achieved so far. In order to avoid the "politically incorrect" term, various euphemisms [3] are used: *observer*, *trace checker*, *the oracle*, *passive monitor*, *arbiter*, *supervisor*. Another indication of the present niche character of research on passive testing is its apparent discontinuity: frequent "restarts" and "re-inventions" of its key elements – a phenomenon not unknown in science, but in

[1] "*Implementation... is exercised with selected sequences of inputs*" [18]. "*...the process of operating a system or component under specified conditions* [as explained elsewhere – understood to be imposed by a tester]..." [19]. "*...testing always implies executing the program on (valued) inputs*" [20, ch. 5]. "*Software testing involves... systematically executing the software, while stimulating it with test inputs...*" [21].

this case it is particularly easy to be ignorant of previous work on passive testing (see [3] for examples).

Between these two approaches, there is an apparent gap: very little has been written on the fundamental methodological issue of whether passive testing "should" be admitted as *bona fide* testing. Serious attempts at starting a discussion at the meta-level, *about* passive testing, are known to have been vigorously rebuffed, as unnecessary, idle, and – allegedly – showing disrespect for "established and accepted truths". This stance is understandable in the vocational, *engineering* tradition, with its apprentice model, but questioning the present state of a conceptual framework is natural, healthy, and indispensable for any *science*, and should not be confused with the "know-better" attitude. This is yet another justification for testology.

Apart from the intellectual challenge of establishing a place for passive testing within testology, it can be shown that there is the growing *need* for it. Recently, a stream of reservations has been raised, by different authors, concerning the ability of testing, *as it is traditionally understood*, to respond to evolving needs, as briefly surveyed below.

Among the new tendencies and postulated further developments of formal model-based testing, [23] identifies:

- integration of test techniques, in order to be able to choose for every task their best combination,
- accepting that a product, however thoroughly tested, evolves and changes.

Both postulates may be re-cast in terms of passive testing, in the following way. Passive testing may be considered as a particular set of combinations of constituent elements, or "modules", of the general testing methodology; active testing would then be *another*, different set of such combinations. This conceptual and technological modularity was proposed by this author already in 1996 [29], and it has been researched since then under the name of protocol multimeter (PMM) [30]. One of the hypotheses tacitly adopted for active testing is that a system under test does not change *during* the tests, and that it is still meaningful to refer to test verdicts *after* the testing is over [31]. In fact, however, *all* real-world implementations do change, in unexpected ways and moments in time. This makes active testing, performed in finite sessions, inherently inconsistent with its hypotheses[2], while the "campaign-less" passive testing is not affected.

In his unpublished keynote speech at the recent software testing conference (ICST, Berlin, 2011), Ian Sommerville, the authority on the design and testing of ICT systems, put to doubt the universal validity of very foundations of testing, as it is currently researched and practiced within ICT. He identified these foundations as deriving from Hume's *reductionism* – reducing complex systems into manageable parts, simple enough to be understood, and interpreting the whole system in terms of interactions of these parts. This approach is conspicuous in the succession of activities in software testing: *unit*, *integration*, *system*, and *acceptance* testing. It is based on strong assumptions: that system boundaries, boundaries of its parts, and the detailed specifications and correctness notions for these parts can always be established, and that there is *control* over both the process of decomposition and putting together, and the operation of the parts (the latter being directly tied to active testing). In *systems of systems* (including global telecommunications and information technology systems), these assumptions simply do not hold: a system is multi-purpose (and these purposes are not established a priori), it exhibits emergent behavior ("*we put it together and strange things happen*"; ibid.), it is not built at once, it is unlimited in size and time scope, it is dynamically changing, it is not clear what constitutes its parts, the boundaries of its stipulated parts are constantly re-negotiated, and there is no *single* notion of its failure. The consequences and recommendations for testing include: basing the testing of such systems on "*actual system operation, not mythical specifications*" (ibid.), and accepting that there is no *single*, pragmatically meaningful result of testing (obtained by executing a particular test suite). Although this was not explicitly stated, passive testing clearly addresses both concerns.

In the sequel, in order to question the distinguished role of the "active" part of the concept of testing, we turn to external, arguably – more generic ideas, including those that had been established much earlier, before any current, disciplinary connotations had any chance to set in.

3. Towards the Generic Concept of Testing

It is possible to characterize testing very generally [3], in a way that avoids preconceptions, as:

- an activity with at least some empirical, *experimental* elements, the results of which can only be established *a posteriori*;
- where experiments are conducted on a particular object – *thing under test* (**Tut**);
- in order to evaluate a certain entity that partakes in testing – the *object of assessment* (**Ooa**);
- conducted with a certain *aim*. The quasi-equivalent formulations of this aim, adopted by different schools of thought, are:
  - to establish whether a given *relation*, normally – an equivalence or preorder holds between a **Tut** and a given *reference* (**Ref**) as often adopted within the formal testing community;
  - to establish whether a given *hypothesis*, which also means – all its necessary consequences, or "requirements" concerning a **Tut**, can be regarded as true (this is the essence of the logical approach to testing, as exemplified by the "industry-oriented" testing framework [26], and also that of the *Scientific Method*);

[2] Completeness of testing is usually defined as a *relative* notion, based on the assumption that hypotheses [32] *are* true.

– to obtain *knowledge* as to whether **Tut** corresponds to a **Ref** in a specific way (where the need for testing may be restated as the need to *know* if a system is correct – this is the language of *epistemology*).

A *test result*, or *test outcome*, always pertains to a **Tut**[3] – it records how a **Tut** behaved during a test. A *test verdict*, normally in {`Pass`, `Fail`, `Inconclusive`}, pertains to the object of assessment, which may, or may not be a **Tut**. The (non)identity of **Tut** and **Ooa** is subject to some debate. In engineering (and thus also in traditional ICT testing), a **Tut** is indeed taken to *be* the object of assessment (and thus, also the object of the ensuing corrective actions, if necessary), which is often reflected in various definitions of testing. In natural sciences, however, the object of assessment is normally a **Ref** – a hypothesis that explains and predicts facts about **Tut** – it is this hypothesis, not "the world", which may be found by testing to be defective. Redirecting the assessment is also possible, e.g., for *reverse engineering* [33].

The relations between **Tut**, **Ref**, and **Ooa** are just one dimension in a *matrix of choices* for sensible combinations of elements in the conceptual space of testing. Some of these combinations are actually claimed and occupied by different research schools, and other – are still to be "discovered" and tried out. In [1] it has been shown how much flexibility is to be gained by surrendering some *habitual* choices. Herein it is claimed that insisting on testing being *active* is one of such choices, the origin of which is no longer clear. Nowhere in the proposed "generic" exposition of testing "being active" appears as a necessary property of testing. In Aristotelian theory of predication, such property might be *essential* – included in a definition (as it is currently presented), or might be a *proprium* (*idion*) – still necessary, and derivable from a definition, but not explicitly present in this definition. In this author's opinion, "being active" appears rather to be of the third kind of predication – an *accident* of testing.

4. Testing and the Scientific Method

Testing, as a concept, did not emerge with technical systems. The important pre-technical, philosophical (epistemological) aspects of testing are present in the *Scientific Method* (SM) – a group of paradigms of sound scientific enquiry; in particular, we refer to one member of this group, attributed to W. Whewell, J. S. Mill, and K. Popper. It is primarily applicable to natural sciences, which does not preclude it from being a viable source of insight in a more technical context. As illustrated in Fig. 1, the application of SM consists in taking a series of steps:

– identifying a problem (i.e., a set of phenomena);

– stating a hypothesis that explains this problem – a statement p about "the world", preferably presented as a logical formula;

– deducing a set of the necessary logical consequences of the hypothesis: $\{p \rightarrow q_1, p \rightarrow q_2, \ldots\}$, where q_k *must* hold if p is indeed true;

– expressing the selected consequences in terms of their individual "empirical content" – predicted phenomena f, in principle amenable to empirical observation, such that q is true iff f "exists" (i.e., depending on its nature, *occurs*, *holds*, is *present* or *absent*);

– *testing the hypothesis* – performing *experiments* aimed specifically at confirming or denying the existence of predicted phenomena.

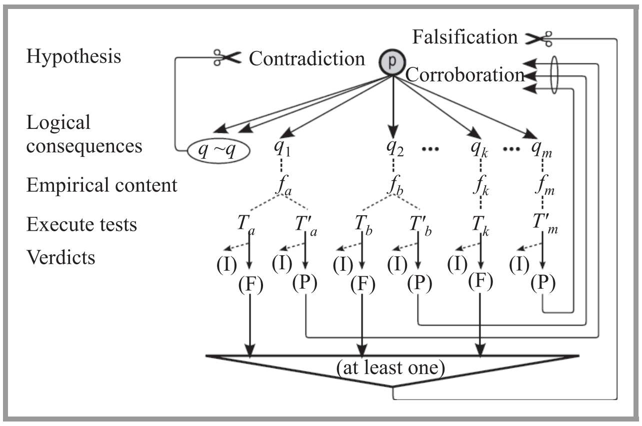

Fig. 1. Tests in the Scientific Method ([3]).

It is possible to check by purely formal means if the derived consequences of a hypothesis are non-contradictory. This non-empirical element appears in testing theories and practices as a "static phase" of testing; e.g., as *static conformance review* [26]. Being *a priori*, it does not count as testing proper, and is not presented as such in SM. In general, however, even the fundamental *a posteriori* character of testing is not universally admitted. Two quite opposite views on this matter, both voiced in "official" publications of the testing community, are: "*unlike dynamic testing... static testing techniques rely on...*" [8], and "*Different from testing, and complementary to it, are static techniques...*" [20]. This shows, again, how *arbitrary* the conceptual foundations of testing are, and further legitimizes raising and investigating doubts about these foundations.

Experiments entail empirical observation. The same method and means of such observation could be used in different *way*s, with different *aim*s, e.g., in the initial phase

[3] Admittedly, "**Tut**" is not an established term, but other, more conventional terms such as **Sut** (*System under Test*), **Iut** (*Implementation under Test*), **Eut** (*Equipment Under Test*) are too specific, being related to a particular test architecture or kind of tests.

of the application of the method – to "charge" one's intuition as to the phenomena about which one is to propose an explanation. This activity, although empirical, does not qualify as *testing* – it should rather be called *monitoring*.

Two extreme views on how to approach the testing of a hypothesis, or *directions of testing*, are (disregarding nuances): *verificationism*, according to which a hypothesis, to be accepted as true, must be convincingly confirmed or *corroborated*, (also – *verified*, in the sense: shown to be true), and *falsificationism* (attributed to Karl Popper), which holds that it is essentially not possible to empirically verify a hypothesis, and the only sensible (meaningful) direction is to try to *falsify* (refute) it. This very influential Popperian stance [34], taken to the ground of ICT, re-emerged in the well known observation by Edsgar Dijkstra that "*testing can only show the presence of bugs* [i.e., can falsify the claim of correctness] *but never their absence* [i.e., cannot verify that all the system's properties are as predicted]". Pure approaches, however, are extremely rare – practical applications of the scientific method almost always *combine* the elements of verification and falsification, in varying proportions (as was also explicitly postulated for technical validation activities in [35]). Popper admitted that corroboration *does* count scientifically, if obtained for genuinely risky predictions. In this sense, the Dijkstra's observation seems surprisingly shallow and misleading. It overlooks the very principles of *model-based testing* [24] with its accompanying assumptions (or *test hypotheses* [32]), under which it is perfectly possible to (conditionally) *prove* correctness.

The role of experiments is to confirm or deny the existence of phenomena, regardless of how "existence" and "experiment" are technically defined. The outcome of executed experiments is thus associated with verdicts: `P` (pass) if an experiment *confirms* the existence a predicted phenomenon, `F` (fail) if it *denies* this existence, and `I` (inconclusive) if neither holds. In general, `P` is not the converse of `F`, although in particular testing theories this may be the case. According to the idea of the "non-orthodox" Scientific Method, as shown in Fig. 1, tests-experiments for each phenomenon are divided into two groups: $\{T\}$ – tests aiming at falsification (so only able to issue a `F` or an `I`), and $\{T'\}$ – tests aiming at corroboration (so only able to issue `P` or `I`). For some predicted phenomena (like f_k and f_m), only one of these groups of tests may be present. It is also conceivable that a falsification and corroboration test be combined in a single experimentation unit, so that its verdict is in $\{\mathtt{P},\mathtt{F},\mathtt{I}\}$. This is the basic form of tests considered in [26]. It has direct representation in the linguistic devices of the TTCN-3 test language [27]. It is, however, by no means generic. Confirmation and refutation, not being the converse of each other, may need entirely different experimental approaches, and these are likely to translate into unrelated, orthogonal test programs, the composition of which may be unnecessarily complex and purely artificial.

Let us now combine the Scientific Method with the general view on tests presented in the preceding section. It can be seen that a hypothesis p is, at the same time, a **Ref** and an **Ooa**, while a **Tut** is a fragment of "the world", in which predicted phenomena occur. This setting can be brought closer to what is customary in ICT-related testing, by stating "**Tut** is correct" as p, and accepting, as the necessary logical consequences $\{q_1, q_2, \ldots\}$ of this hypothesis, the *requirements* (if a testing theory is cast in logic [36]) or particular features of a *behavioral model* (for a process-oriented approach). In this case, the consequences are obtained in a different way – they are not really *derived* from p, as very little can be derived, by pure logic, from "**Tut** is correct"[4]. Instead, some subset of consequences would contain the explicitly stated, *essential* requirements that define a correct **Tut**, and another, possibly *very large* set would contain its *propria*, derivable from the defining requirements, but not explicitly stated in a definition of **Tut** (and so, formally, not counted as essential). The place of *accidental* consequences in this picture is most dubious, as it is in philosophy in general. Accidents are not really instrumental in *distinguishing* things (correct implementations of A from correct implementations of A', or correct and incorrect implementations of A). Stating the "proper" set of q that would be subject to testing is one of the primary problems in testing theories. In natural sciences, new q are produced and subjected to testing continually. In technical testing, this problem is recast as *test generation and selection*. Associating particular q with different general *kinds* of properties (e.g., according to the Aristotelian concepts) is also tacitly practised, which transpires from, e.g., the telecommunications-oriented concept of "essential requirements" (as opposed to non-essential requirements, the testing of which might be skipped).

Within SM, "experiment" and "test" have, for all practical purposes, the same sense. What *is* this sense, has been investigated by philosophy of science [37], but nowhere within the context of the Scientific Method an experiment is described as necessarily *active*, i.e., that in which influence is purposefully exerted upon investigated phenomena. The distinction between *passive* (or "natural", or quasi-experiments) and *active* (or "controlled") tests/experiments has been, however, noticed and discussed. J. S. Mill calls them, respectively, *pure observation* and *artificial experiments* [38], and finds a place for both in scientific enquiry (and thus – in the Scientific Method). According to Mill, their essence is, respectively, to "*find an instance in nature suited to our purposes, or, by an artificial arrangement of circumstances, make one*". At a sufficiently high level of abstraction there seems to be "*no difference in kind, no real logical distinction, between the two processes of investigation... as the uses of money are the same whether it is inherited or acquired*". Mill does acknowledge the "great disadvantage" of pure observation, such as the apparent inability to ascertain causal relations and "*to produce a much greater number of variations in the circumstances than nature spontaneously offers*". He also identifies circumstances in which pure observation is advantageous, and his argu-

[4] If we ignore some quite fundamental, but still disputed philosophical consequences, such as "**Tut** exists" or "*something* is correct".

ments resemble the current expositions of the distinguishing features and applications of passive testing.

One Mill's remark, if taken literally, may provide the understanding of passive testing that *directly* maps onto active testing: "*Instead of being able to choose what the concomitant circumstances* shall be*, we now have to discover what they* are" (ibid.). The "concomitant circumstances" map to a particular *test purpose* as pursued by means of a particular *test preamble* (both being the elements of active tests, as stipulated in [26]). Instead of actively executing a test preamble, a passive tester recognizes it *if* and *when* it happens to occur. According to this interpretation, it is, in principle, possible to use the same test suite, however generated, for active *and* passive testing. The idea of recognizing a sequence of events embedded in a trace of behavior has been explored within the *pattern-matching* approach to passive testing [39], but, to this author's knowledge, it has not been developed so far as to suggest that these patterns may directly correspond to preambles taken from an active test suite.

5. Connotations of Passive Testing

As already stated, the prevailing operational characterization of testing derives from Mill's *controlled* experiments. Similarly, passive testing may be said to be based on *quasi-experiments* – the observation and assessment of phenomena that are not invoked (provoked, stimulated, influenced) by a tester. Pragmatically, this lack of influence may be intended or required for the following reasons.

The nature of a phenomenon may *not allow* for such influence (e.g., as in the investigation of the radiation spectrum of a distant star). In testing applied to technical systems, this translates to the absence of input port(s) – their genuine, physical absence, their administratively imposed inaccessibility for testing, or (as may be common for systems of systems) lack of information on whereabouts of these ports. Proposing that a **Tut** should provide the "testing ports" is a part of the *design for testability* framework [40]; one of its ideas postulates equipping a system with *additional* devices (interfaces and special functional properties), *specifically* for the purposes of its prospective, eventual testing. This approach has currency, e.g., in electronic circuit design, but is not advocated (or is even "prohibited") in most testing contexts in telecommunications.

A phenomenon may be "intensive enough". As an analogy, consider the dictionary meaning of "test" in chemistry: it is defined as a process of identifying the presence or the nature of a substance, *commonly by the addition of a reagent*. A reagent (and also a *catalyst*, which may be used for similar purposes) is analogous to a focused stimulus, which makes a phenomenon or substance *reveal itself*.

External stimulation (although feasible) might change or distort a phenomenon. There remains, however, a philosophical question as to whether passive observation really solves this problem. The *observer effect* (not to be conflated with the Heisenberg's *uncertainty principle*) is a posited principle, according to which a mere act of observation necessarily changes the phenomenon being observed. On the macro scale, in the setting of complex ICT systems, this concern may be safely dismissed[5].

A **Sut** may be a larger system whose *integrity, safety, and performance* critically depends on non-interference with its internal parts and processes. Normally, active tests would include incorrect, invalid, unexpected (inopportune [26]) stimuli that would have an *a priori* unknown effect upon a system[6], which may well be catastrophic to the system's mission – as in the U.S. network-wide failure of 1990 [42], and also in the Chernobyl's disaster, both caused by a stimulus that was not even incorrect or unexpected *per se*.

Finally, it may be *too cumbersome or costly* to build and operate the "sending" channel of a tester, through which stimuli would be administered.

One may claim that, regardless of any technical, local definitions, it is "common knowledge" that testing is active, as (supposedly) codified in the language and reflected in the common use patterns of the term, recorded in dictionary entries[7]. The passive nature of testing is stipulated, or at least not rejected, in the following dictionary entries for "test":

- *the means by which the presence… of anything is determined* (this closely resembles tests for the presence of a phenomenon in SM, which, as already indicated, do not have to be active);
- *trial – the examination before a judicial tribunal of the facts… in a case* (where the tribunal has no power to influence the course of the past events to re-enact alternative scenarios).

The active elements are emphasized in the following entries:

- *trial; to try out* (in order to be tested, something must be actually *used*, which connotes both-way interactions with this entity);
- *a set of standardized questions, problems, or tasks designed to elicit responses for use in measuring the traits, capacities, or achievements of an individual* (to elicit responses is the explicit role of stimuli in active tests).

Altogether, basing on [44] it may be concluded that the active and passive connotations of "test" are well balanced.

Distinctions made in the purely technical context are also much less clear-cut than it is usually admitted. One

[5] Although there have been insightful discussions on the behaviour of automata, in which both principles have been used (at least metaphorically) under the name of "complementarity" [41].

[6] If this effect were known *a priori*, testing would not be needed at all.

[7] This is a genuine and acknowledged problem. For example, in [43] it is noted that *"the field of IS* [Information Systems] *development is severely hampered by the limitation of meaning derived from the everyday use of some representative words…"*.

of the early attempts at harmonizing active and passive testing [45] has been to employ two testers operating in parallel: an *active* tester for confirmation tests leading to the `P` verdicts, and a *passive* tester, originally called a *trace analyser*, for refutation tests leading to the `F` verdicts. This solution was based on the earlier idea of separating the problem of choosing and applying *test inputs* (stimuli – the "pure" active part of testing) from the problem of assessing the behavior of a **Tut** for *these* or *any other* stimuli that may have been provided, by any means [46]. In both approaches, the passive testing functionality may form a part of a compound (effectively – active) tester, or may, in the limit, form the *whole* of a tester – a self-contained *passive tester*. The conditions for such transformations were discussed in [3]. In these early stages, a passive tester was apparently treated as a *bona fide* entity, and not as a metaphor. Interestingly, one of the first direct uses of the term "passive testing" (instead of various euphemisms) was made in [47], when the initial acceptance for passive testing as (a kind of) testing seemed to evaporate.

6. Testing versus Monitoring

It is surprisingly difficult to precisely state the difference between *monitoring* and *testing* a system. The common tendency so far has been to conflate monitoring with passive testing[8]. The prevailing intuition is that monitoring is *not* testing, so "passive testing is not testing" as well. We argue that it is both necessary and possible to keep the two notions apart – they have different *sense*, even if, in the limit, they may refer to technically "the same" activity.

Let us assume that at least some level of technical instrumentation is necessary, and that the suitable technical instruments: a *monitor* and a *passive tester*, respectively, are present. In the following, we treat monitoring as *using* a *monitor*, resp. testing as *using* a *tester*, and we look at each of the constituent parts of the decomposed concepts separately. Using a thing (an apparatus) presupposes the existence of a *user* – some external entity that is *not* a monitor (resp. a passive tester). Clearly, not *every* use of technical instruments lies within the scope of the respective notions – using a monitor to hammer down nails would certainly not count as monitoring. The pertinent question is *what* use of a monitor (tester) makes for monitoring (testing), and how this "proper" use is related to the functionality of the instrument.

We first consult the basic dictionary meanings of "monitoring" [44], noting the recurring use of two key terms: *looking* (or *watching*) and *seeing*:

1. *Listening to transmitted signals in order to check the quality of the transmission*. Monitoring is thus performed *in order to* check (some properties), but checks themselves are left to the user. Consider a medical monitor (e.g., an electrocardiograph). The output of a monitoring system is a stream of data, suitably (e.g., graphically) *presented* so that it can be conveniently *interpreted*. The interpretation itself rests with the doctor, who on different occasions can *look* at (the same) data from different perspectives, in order to *see* if there is any activity of the heart, or if the heart-beat is regular, etc.
2. *Observing, recording, or detecting (an operation or condition) with instruments that have no effect upon the operation or condition*. This definition stresses the *passive* and *technical* character of the operation. "Detecting" suggests the higher-level functionality that will be later assigned to an *extended* monitor.
3. *Keeping track of, checking continually*. This stipulates a "campaign-less", possibly infinite process.

[8] "*Monitoring is … called passive testing…*" [48].

Points (2) and (3) correspond to the joint characteristics of monitoring and passive testing, while point (1) seems useful for differentiating the two. It relies on differences between *looking* (watching, listening) and *seeing* (hearing). According to [44], to *look* means: to direct one's glance, attention, consideration (to *watch* – to keep under attentive view or observation, as in order to see something); to *see* means: to perceive (things) mentally, to discern, to understand, to recognize.

The first approach to differentiating between monitoring and passively testing a *thing under investigation* – **Tui** (it is not known yet whether it is *thing under monitoring* – **Tum**, or **Tut**) is based of different *levels* of interpretations (see Fig. 2). *Monitoring* is a technical counterpart of watching a particular aspect of a system. Monitoring provides, in a *passive* way, a *continued* stream of processed data on the behavior of a **Tum**. These data are intended to be interpreted by an external process (in particular, but not exclusively, by a human operator), where the underlying phenomena or properties of a **Tum** can be *seen*. Unlike monitoring, *passive testing* involves both, the (syntactic) process of *watching*, and the (semantic) interpretative process of *seeing*. The latter may still be performed by a human test operator (as is often stipulated in approaches to testing characteristic of software engineering). The output

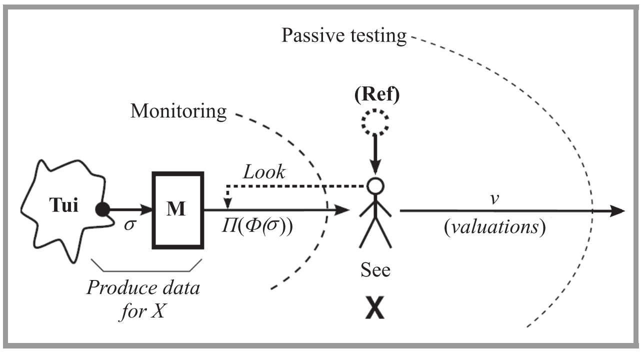

Fig. 2. Monitoring and passive testing.

of the passive testing process is a stream of *interpretations*, or *valuations* (ν) of monitored data.

A basic technical apparatus for monitoring, a *simple* (or *plain*) *monitor* (**M** in Fig. 2) is thus assigned functions for: syntactically transforming (filtering, projecting) a stream of "raw data" about the behavior of **Tum** into a stream of processed data: $\sigma \mapsto \sigma' = \Phi(\sigma)$; and suitably presenting (formatting) the processed data: $\sigma' \mapsto \Pi(\sigma')$, so that the stream of formatted data can be *conveniently* interpreted. The presentation function is not mere aesthetic decoration; it is an important part of the notion of monitoring. Both syntactic processing and pragmatic presentation mode depend on the intended external interpretation process – the understanding is that of "monitoring *for*..." or "monitoring *in order to*...". Monitoring is thus *not* purpose-agnostic.

Experience shows that it is quite common to shift some semantic interpretation functions (e.g., raising an alarm if a threshold is exceeded) from an external process to the monitoring process itself, which will now be referred to as *extended monitoring*. A monitor enhanced with the interpretation function **I** is an *extended* monitor (**Me** in Fig. 3a). Such a device would be able to, e.g., raise an alarm when certain threshold values are exceeded (as in a class of medical monitors), while still being referred to as a monitor, and not a tester. A single layer of inbuilt interpretations yields what might be called *tier 1* of *extended monitoring of* **S**. There may be many consecutive tiers of extended monitoring.

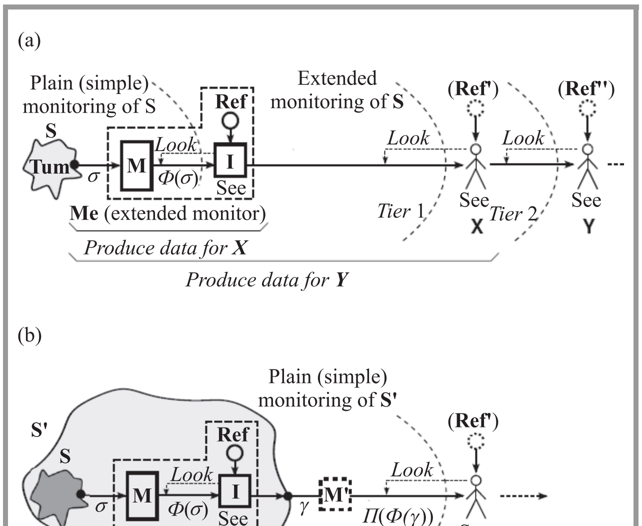

Fig. 3. Plain and extended monitoring: (a) tiers of extended monitoring; (b) shifting the boundary of the monitored system.

One way to conceptually dispose of the notion of extended monitoring (and to stay with "just" monitoring), is to re-position the boundary of a system under monitoring – the stream of interpretations is now regarded as raw data γ on the behavior of *another* system **S'** (Fig. 3b). The original system **S** is now *embedded* in a context, which makes it clear that its behavior can only be investigated indirectly[9].

When interpretations become a part of the monitoring process itself, the distinction between monitoring and passive testing, as proposed in Fig. 2, seems to collapse – the output of *both* processes is now a stream of interpretations. Additionally, with the growing number of tiers of extended monitoring (Fig. 3a), there is no clear point at which monitoring would "magically" change into (passive) testing. It thus becomes apparent that another, additional criterion for distinguishing monitoring and passive testing is necessary. We take this additional criterion to be the *kind* of interpretations that are carried out within the respective processes, as already hinted in [49]. Let us consider a pair: $< B, C >$ consisting of a particular *behavior*, and circumstances (*conditions*) in which this behavior is exhibited. We claim that monitoring and (passive) testing differ in the pragmatically meaningful, logical ordering of the elements of this pair. In (extended) monitoring, interpretation (valuation) serves to infer, from the observed behavior, the conditions (circumstances), or the general *mode of operation* of a **Tum**, such as "being overloaded (congested)", "being down", "being stable", "being under attack" (in the context of intrusion detection [1]), or "being dead" (in the medical context). This is consistent with the view that "*monitoring consists of measuring properties of the network, and of inferring an aggregate predicate from these measurements*" [50]. In testing, interpretation is related to the defined circumstances (conditions), called in this context *test purposes*. In active testing, a test system, steered by a test program, *establishes* (forces) these conditions, while in passive testing a test system *recognizes* them. Note that the same understanding was also arrived at earlier, although in a different way.

It follows from the foregoing discussion that monitoring can be located as a lower-layer functionality with its results interpreted by testing, or as a higher-layer functionality acting on a stream of lower-layer test verdicts. According to this view, both monitoring and passive testing are "full", but *different* functionalities, with no fixed subordination relation between them. We conclude that passive testing *can* be distinguished from monitoring, and thus can be freed from one of its strongest "non-testing" connotations.

7. Testing versus Measurements

Intuitively, testing and measuring are closely related (as in: test *and* measurement), but distinct concepts. This intuition makes metrology an interesting source of concepts and mechanisms to be directly imported, and also general insight and analogies. Some links between the two domains have already been briefly identified in [3]. To be a measurement, determining/assigning a value must be based on empirical observation of a real, existing object – similarly

[9]This setting can be re-cast as *verification-in-context* [35]. This is also the classical telecommunications setting, where **S**, called **Iut** (*Implementation under test*), is embedded in, and only indirectly accessible through, other parts of a **Sut**.

for testing. According to [51], the necessary conditions for calling an evaluation a measurement are: a well-defined, external *reference*, and a well-defined measurement *operation* which can be carried out independently of any specific measurer. These two postulates have always been the cornerstones of formal testing: the former is at the very core of *model-based* testing, and the latter was given due consideration, e.g., in [26] as "Conformance Assessment Process".

Surprisingly, the in-depth, direct, non-metaphorical discussion of the relations between measurements and testing is lacking. The measurement community at least declares interest in suitably extending and adjusting their methodology so as to accommodate testing, but any reciprocal effort from the testing community has not manifested itself. The metrological interest in testing is mainly due to the conceptual troubles with applying traditional concepts of measurement (implicitly focused on *physical* quantities[10]) to information technology artifacts, with their *logical* properties [53]. Despite this declared interest, the current effects of harmonization are modest. Conspicuously, "measurement" and "testing" are defined not side-by-side, but in different metrological documents: "Measurement – *process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity*" [13, 2.1] vs. "Testing: *determination of one or more characteristics of a given object of assessment according to a procedure*" [54]. One of the rare explicit explanations of the distinction is "*sometimes made by considering testing to be a measurement or measurements together with a comparison to a specification*" (ibid.). This explanation is, however, flawed in that "comparison to a specification", in a broad sense, is also the *internal* element of most methods of measurement.

The main obstacle to directly applying metrological concepts to testing seems to be the core concept of metrology – the *measurand*, which is controversial in itself, and subject to internal, metrological debate [55]. A measurand is a *quantity*, i.e., a property that has a magnitude that can be expressed using numbers [13, 1.1] (more generally – expressed symbolically). The minimum requirement seems to be that the objects of measurement can be *ordered* w.r.t. the magnitude of a quantity in question. The mainstream concepts of metrology pertain to such quantities that meaningful algebraic operations on their values (expressions of magnitude) can be defined, so that the results of these operations reflect the empirical relations between quantities of respective objects. Metrology also explicitly admits *ordinal* quantities, which can be (numerically) expressed and which enter into (empirical) ordering relations, but with no corresponding algebraic operations on the expressions of their values (e.g., garment sizes: $\{\mathtt{XS}, \mathtt{S}, \mathtt{M}, \mathtt{L}, \mathtt{XL}\}$). The values of these quantities can be obtained by a conventional measurement procedure. There are also properties that have been specifically excluded from the scope of the concept of "quantity", and thus also from the scope of "measurement" – *nominal* properties that have no magnitude [13, 1.30], although can be assigned a (symbolic) value. Sex and colour have been given as an example (ibid.), although this is debatable – ordering (the "value" of) humans by sex or race has often, sadly, been practiced[11], and colour has an obvious "objective" value (wavelength). Despite this somewhat arbitrary exclusion, there have been attempts at the metrological treatment of taxonomic, nominal relations (such as postal codes [56]).

For testing, the "measurand" would be correctness, and the conventional expression of its value is a verdict, in $\{\mathtt{Fail}, \mathtt{Inconc}, \mathtt{Pass}\}$, or, possibly, in $\{0, \frac{1}{2}, 1\}$ (if this should bring more metrological connotations). It may seem to be the nominal, taxonomic property, officially – beyond the scope of metrology. In testing, however, the verdicts (reflecting the "magnitude" of correctness) do introduce ordering on systems – incorrect systems are "less than" correct ones, and there may be different implementations of a standard that are "equally correct". There is also the explicit ordering on verdicts: $\mathtt{P} \rightarrow \mathtt{I} \rightarrow \mathtt{F}$, built into the semantics of the test language TTCN [27]. On the other hand, composing a correct and incorrect system may yield a system that is correct or incorrect, which *a priori* cannot be established by applying any operations to their individual correctness values. This is why, after conformance testing, combined systems are subjected to interoperability tests.

It may be concluded that the *direct* application of metrological concepts and language to testing seems no more controversial than the ongoing debates within metrology itself (including the notion of a measurand). If, however, mutual harmonization is for any reason unacceptable, then metrology may always be used as a source domain for *scientific metaphors* [5] aimed at explicating testology. As in any similar case, the metrological community has no "right" to stop testology from applying such metaphors (or to enforce the observance of all the metrological definitions and agreements to the last detail). The only criterion of the validity of metaphors is their effectiveness. Obviously, the canonical metaphor to try out is: "Testing is measurement".

Having established the applicability, either direct or metaphorical, of metrological concepts to testing, we now briefly return to the "active-passive" dimension. Measurement is popularly believed to be essentially *passive*, as it is intended to assess the object of measurement "as it is". Contrary to this impression, the techniques of measurement, which clearly constitute some part of its essence, are explicitly divided into passive (as in measuring the radiation spectrum of a body) and active. Measuring resistance may be performed actively, by applying a certain voltage (a stimulus) and observing the resulting current. The same quantity may also be measured passively, by observing the

[10]"*. . . the measurement of a well-defined physical quantity — the measurand*" [52, 1.2]

[11]Such pragmatic "ordering" should not be *a priori* rejected, in view of the general shift from regarding measurement as *determination* (of some elusive "true value") towards treating it as *assignment* [51].

relation between the current and the voltage in a circuit, while refraining from actually applying either. Both would be readily called "measurement", with no methodological and linguistic reservations. It is acknowledged that under certain circumstances one technique is preferred over the other, or is exclusively applicable, but no paradigmatic preference is given to either. This symmetry is a feature of metrology that should, by analogy, at least be given due consideration in testology. Should metrology be consulted for insights, testology would not find there any justification for its current asymmetric views on the active nature of testing.

8. Concluding Remarks

Any scientific community, including the testing community, is free do define the scope of its interest, conceptual horizon, and terminological (linguistic) devices. Such choices are, however, not beyond the scope of external scrutiny. They are also often the object of *internal*, intra-disciplinary debate. For example, Gaudel [32] felt that it was necessary to examine the general dictionary entries to recharge the failing intuition of testing. Similarly, within the broad context of information systems there are schools of thought that, dissatisfied with a certain methodological lock-in, try to re-define their discipline in terms of *semiotics*. Also investigations into how people use words have a long tradition in social sciences and philosophy of science.

The presented high-level methodological discussion of testing is not the first of its kind. It is similar in vein, and complementary (but more focused in scope) to [57]. It also builds on [1] and [3], where an attempt is made at identifying and dismissing spurious incompatibilities between the "testing-like" concepts developed by different research communities, and on [5], which surveys the methodological aspects of looking for insight and borrowing concepts.

The aim of this work has been not to arbitrarily fix a terminological "misunderstanding", but to show how testology could be freed from a particular family of preconceptions that seem to impede one direction of its development.

Acknowledgements

This work presents the motivation and results of the fundamental research track of Research Task 7: "Verification and validation of network protocols by passive testing", within the framework of Research Project PBZ-MNiSW-02-II/2007 contracted by the Polish Ministry for Science and Higher Education and financed with the 2007-2010 reseach funds. Fragments of the author's work [3] were re-used in Section 4, in accordance with the publishing agreement.

References

[1] K. M. Brzeziński, "On common meta-linguistic aspects of intrusion detection and testing", *Int. J. Inform Assurance and Secur. (JIAS)*, vol. 2, no. 3, pp. 167–178, 2007.

[2] K. M. Brzeziński, "A joint meta-linguistic taxonomy of intrusion detection and testing/verification", in *Proc. 2nd Int. Worksh. Secure Informa. Syst. SIS'07*, Wisła, Poland, 2007.

[3] K. M. Brzeziński, "Towards the methodological harmonization of passive testing across ICT communities", in *Engineering the Computer Science and IT*, S. Soomro, Ed. In-Tech, 2009, pp. 143–168.

[4] K. M. Brzeziński, "On conceptual struggles over 'testing'", in *Proc. XIV Poznańskie Warsztaty Telekomunikacyjne PWT 2010*, Poznań, Poland, 2010.

[5] K. M. Brzeziński, "Standards are signs", in *Proc. 15th EURAS Ann. Standardization Conf.*, Lausanne, Switzerland, 2010, pp. 43–60.

[6] *Software testing. Vocabulary.* BS 7925-1. British Standards Institution, 1998.

[7] *Standard Glossary of Terms Used in Software Testing, version 2.0*. ISTQB (Glossary Working Party), 2007.

[8] *Certified Tester. Foundation Level Syllabus*. ISTQB, 2007.

[9] *Certified Tester. Advanced Level Syllabus*. ISTQB, 2007.

[10] A. Spillner, T. Linz, and H. Schaefer, *Software Testing Foundations*. Rocky Nook, 2007.

[11] G. Bath and J. McKay, *The Software Test Engineer's Handbook*. Rocky Nook, 2008.

[12] C. F. Tschudin, "On the structuring of computer communications", Ph.D. dissertation, University of Geneve, 1993.

[13] VIM, *International vocabulary of metrology – Basic and general concepts and associated terms*. Joint Committee for Guides in Metrology (JCGM), 2008, vol. JCGM 200.

[14] R. R. Hoffman, *Metaphor in Science*. Lawrence Erlbaum Associates, Inc., 1980, pp. 393–423.

[15] S. Purao, C. Y. Baldwin, A. Hevner, V. C. Storey, J. Pries-Heje, B. Smith, and Y. Zhu, "The sciences of design: observations on an emerging field", Harvard Business School, Working Paper 09-056, 2008.

[16] K. Naik and P. Tripathy, *Software Testing and Quality Assurance: Theory and Practice*. Wiley, 2008.

[17] L. Cacciari and O. Rafiq, "Controllability and observability in distributed testing", *Inform. Software Technol.*, vol. 41, no. 11-12, pp. 767–780, 1999.

[18] C. Sunshine, "Formal techniques for protocol specification and verification", *Computer*, vol. 12, no. 9, pp. 20–27, 1979.

[19] *IEEE Standard Glossary of Software Engineering Terminology*. IEEE Std 610-12. IEEE, 1990.

[20] *Guide to the Software Engineering Body of Knowledge*. SWEBOK, IEEE, 2004.

[21] L. Frantzen and J. Tretmans, "Model-based testing of environmental conformance of components", in *Formal Methods of Components and Objects (FMCO'06)*, LNCS 4709. Springer, 2007, pp. 1–25.

[22] E. Brinksma and J. Tretmans, "Testing transition systems: an annotated bibliography", in *Proc. MOVEP 2000*, Nantes, France, 2000, pp. 187–195.

[23] E. Brinksma, W. Grieskamp, and J. Tretmans, "Summary", in *Perspectives of Model-Based Testing* of *Dagstuhl Seminar Proceedings*, E. Brinksma, W. Grieskamp, and J. Tretmans, Eds., no. 04371. IBFI, 2005.

[24] M. Utting, A. Pretschner, and B. Legeard, "A taxonomy of model-based testing", Univ. of Waikato, Hamilton, New Zealand, Working Paper 04/2006, 2006.

[25] J. Ryser, S. Berner, and M. Glinz, "On the state of the art in requirements-based validation and test of software", Tech. Rep. IFI-98.12, Univ. of Zurich, May 1998.

[26] *Conformance Testing Methodology and Framework*. ISO/IEC 9646. ISO/IEC, n.d., vol. 1–7.

[27] *MTS; The Testing and Test Control Notation version 3*. ETSI ES 201 873, ETSI, n.d.

[28] K. M. Brzeziński, "Intrusion detection as passive testing: linguistic support with TTCN-3", in *DIMVA*, LNCS 4579. Lucerne: Springer, 2007, pp. 79–88.

[29] K. M. Brzeziński, D. Mastalerz, and R. Artych, "Practical support of testing activities: the PMM Family". COST 247 WG3 Internal Report, IT P.W., LTiV Tech. Rep. 965, 1996.

[30] K. M. Brzeziński, "Weryfikacja i testowanie", *Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne*, no. 4, pp. 139–140, 2010 (in Polish).

[31] K. M. Brzeziński, "Testowanie w cyklu życia systemu: nieregularności meta-standaryzacji", in *Krajowe Sympozjum Telekomunikacji i Teleinformatyki (KSTiT)*, Warszawa, 2009 (in Polish).

[32] M.-C. Gaudel, "Formal methods and testing: hypotheses, and correctness approximation", *Formal Methods*, pp. 2–8, 2005.

[33] K. M. Brzeziński, A. Gumieniak, and P. Jankowski, "Passive testing for reverse engineering: specification recovery", in *Proc. IASTED Int. Conf. Paral. Distrib. Comput. Netw. PDCN 2008*, Innsbruck, Austria, 2008, pp. 27–32.

[34] K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge, 1963.

[35] L. Heerink and E. Brinksma, "Validation in context", in *Proc. 15th IFIP Int. Symp. Protocol Specification, Testing and Verification PSTV 1995*, Warsaw, Poland, 1995, pp. 221–236.

[36] *Framework on Formal Methods in Conformance Testing.* ITU-T Z500. ITU-T, May 1997.

[37] M. Heidelberger, "Experimentation and instrumentation", in *Encyclopedia of Philosophy*, D. M. Borchert, Ed. Thomson Gale, 2006, vol. 10, pp. 12–20.

[38] J. S. Mill, *Of Observation and Experiment*. Routledge and Kegan Paul, 1974.

[39] J. A. Arnedo, A. R. Cavalli, and M. Núñez, "Fast testing of critical properties through passive testing", in *Proc. IFIP Int. Conf. Testing Commun. Syst. TestCom 2003*, Sophia Antipolis, France, 2003, pp. 295–310.

[40] R. Dssouli and R. Fournier, "Communication software testability", in *Protocol Test Systems III*. North-Holland, 1991, pp. 45–55.

[41] K. Svozil, "Extrinsic-Intrinsic Concept and Complementarity", in *Inside Versus Outside*, H. Atmanspacher and G. J. Dalenoort, Eds. Berlin: Springer, 1994, pp. 273–288.

[42] P. G. Neumann, "Cause of AT&T network failure", *Risks Dig.*, vol. 9, no. 62, 1990.

[43] M. Boahene, "Information systems development methodologies: are you being served?" in *Proc. 10th Australasian Conf. Information Syst.*, Wellington, New Zealand, 1999, pp. 88–99.

[44] *Webster's Encyclopedic Unabridged Dictionary of the English Language*. Gramercy Books, 1996.

[45] R. Wvong, "A new methodology for OSI conformance testing based on trace analysis", Master's thesis, University of British Columbia, 1990.

[46] G. von Bochmann and O. B. Bellal, "Test result analysis with respect to formal specifications", in *Proc. 2nd. Int. Worksh. Protocol Test Syst.*, Berlin, Germany, 1989, pp. 272–294.

[47] D. Lee, A. N. Netravali, K. K. Sabnani, B. Sugla, and A. John, "Passive testing and applications to network management", in *Int. Conf. Netw. Protoc. ICNP'97*, Atlanta, USA, 1997, pp. 113–122.

[48] J. Tretmans, "Testing concurrent systems: a formal approach", in *10th Int. Conf. CONCUR'99*, Eindhoven, The Netherlands, 1999, pp. 46–65.

[49] K. M. Brzeziński, "Towards Practical Passive Testing", in *Proc. IASTED Int. Conf. Paral. Distrib. Comput. Netw. PDCN 2005*, Innsbruck, Austria, 2005, pp. 177–183.

[50] M. Dilman and D. Raz, "Efficient reactive monitoring", *IEEE J. Selec. Areas Commun.*, vol. 20, no. 4, pp. 668–676, 2002.

[51] L. Mari, "The role of determination and assignment in measurement", *Measurement*, vol. 21, no. 3, pp. 79–90, 1997.

[52] *Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement*. GUM. Joint Committee for Guides in Metrology (JCGM), 2008, vol. JCGM 100.

[53] "Metrology for Information Technology (IT)". NISTIR 6025. NIST, White paper, 1997.

[54] *Conformity Assessment – Vocabulary and General Principles*. ISO/IEC 17000, ISO/IEC, 2004.

[55] A. C. Baratto, "Measurand: a cornerstone concept in metrology", *Metrologia*, vol. 45, pp. 299–307, 2008.

[56] R. M. Olejnik, "Kod pocztowy jako przykład metrologicznej skali nominalnej", *Pomiary Automatyka Robotyka*, no. 7–8, pp. 186–188, 2004 (in Polish).

[57] T. S. E. Maibaum, "The epistemology of validation and verification testing", in *Proc. 17th IFIP Int. Conf. TestCom 2005*, Montreal, Canada, 2005, pp. 1–8.

Krzysztof M. Brzeziński is Assistant Professor at the Institute of Telecommunications, Warsaw University of Technology, Poland. He obtained his Ph.D. in Telecommunications from the same University in 1995. His research interests concentrate on rigorous, formalized design and testing of distributed (esp. telecommunications) systems, and theory of standardization. He is the author of a book on ISDN technology (also translated into Russian), four book chapters, over 30 conference papers, and over 60 research reports. He is a certified TTCN-3 specialist.

E-mail: kb@tele.pw.edu.pl
Institute of Telecommunications
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland