

THE PROBABILITY IN THROWING DICES AND IN MEASUREMENT

F. Pavese

Torino, Italy, email: frpavese@gmail.com

Abstract – The paper summarises the main differences between the process of throwing dices and the measurement process, and draws some of the consequences on the meaning and use of the probability concept in the two cases.

Keywords: probability, dices throwing, measurement, uncertainty, aleatoric, epistemological, ontological, definitional

1. INTRODUCTION

“It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel. Doubtless, much of the disagreement is merely terminological and would disappear under sufficiently sharp analysis.” [1]

This paper, after a short introduction concerning the several meanings of the terms ‘uncertainty’ and ‘probability’, aims at limiting the illustration to a couple of interpretations of probability, discussing whether application of this concept to dice throwing has exactly the same meaning when it is used for measurement results, or there are differences.

The concept of probability is historically born from the speculation about the prediction in gambling problems, like that of the occurrence of a specific face of a fair dice in subsequent throws or of a fair coin in subsequent tossing.

In experimental science, when scientists realised that it is impossible to get ideally perfect and full information from measurement [2], the use of the concept of probability proposed itself as the most natural way to circumvent the difficulty and model chance.

The paper illustrates why and how, in general, the conditions of its application are basically different, by reviewing a range of positions. Some consequences are drawn.

2. UNCERTAINTY INTERPRETATIONS

The term ‘uncertainty’ has several meanings.¹ One can find a discussion on this issue from an epistemological point

¹ Often the terms "uncertainty" and "error" are used interchangeably, or one of the two is rejected (the second in [14]). "We find it convenient to distinguish them thus: 'Error' is the actual difference between a measurement and the value of the quantity it is intended to measure, and is generally unknown at the

of view in [6]: “Uncertainty is pervasive in most of the fields of science and technology (as it is also in real life and ordinary thinking), and in all cases it seems that what really matters and worries is the quantification of the ‘amount’ of uncertainty in some sense attributable to the considered statements, or events, be them possibly submitted to repetition (random events) or not (singular events). All that is done, in general, under a very unclear concept of what the word uncertainty could actually mean, but it is almost always done by aiming at the measurement of the variables affected, in very different contexts, by some kind of uncertainty.”

According to that approach, “the linguistic term uncertainty is not only imprecise, but has a very broad applicative spectrum in both ordinary life and Science. To scientifically approach what it could mean, it seems necessary to previously capture how their, also imprecise, mother-predicate $U = \textit{uncertain}$, the opposite of certain, is used in different contexts and with different purposes, and where can consequently be represented by a fuzzy set”.

Limiting to the scientific field, it actually did not only bring to the concept of ‘probability’, but also to the many other fields of chance, often called “imprecise probability”, first of all to the possibility and fuzzy type of reasoning [7, 8], but also to others, like previsions, lower and upper probabilities, or interval probabilities, belief functions, possibility and necessity measures, lower and upper previsions, comparative probability orderings, partial preference orderings, sets of desirable gambles, p-boxes, robust Bayes methods (see references, e.g., in [9]).

In particular, uncertainty measures are not necessarily additive [6, 10-12].

In measurement, that attribute is added to uncertainty to specify the intended meaning: “measurement uncertainty: non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used” (term 2.26 in [13]). It is useful to recall here also its Note 1: “Measurement uncertainty includes components arising from *systematic effects*, such as components associated with *corrections* and the *assigned* quantity values of measurement standards, as well as the *definitional* uncertainty. Sometimes estimated systematic effects are not corrected for but, instead, associated measurement uncertainty components are incorporated” (emphases added), useful in the subsequent discussion.

time of measurement. ‘Uncertainty’ is a scientist’s assessment of the probable magnitude of that error”. [22]

3. PROBABILITY INTERPRETATIONS

Probability is one of the concepts born to accommodate the concept of uncertainty, intrinsic in chance, as opposed to the term ‘certainty’ [2].

There are two broad categories of probability interpretations, which can be called "physical" and "evidential" probabilities (the following synthesis is taken from [3], for a full discussion see [4][5]).

Physical probabilities, which are also called objective or frequency probabilities, are associated with random physical systems such as roulette wheels, rolling dice and radioactive atoms. In such systems, a given type of event (such as the dice yielding a six) tends to occur at a persistent rate, or "relative frequency", in a long run of trials. Physical probabilities either explain, or are invoked to explain, these stable frequencies. Thus talking about physical probability makes sense only when dealing with well-defined random experiments. The two main kinds of theory of physical probability are frequentist accounts (such as those of Venn, Reichenbach and von Mises) and propensity accounts (such as those of Popper, Miller, Giere and Fetzer).

Evidential probability, also called Bayesian probability (or subjectivist probability), can be assigned to any statement whatsoever, even when no random process is involved, as a way to represent its subjective plausibility, or the degree to which the statement is supported by the available evidence. On most accounts, evidential probabilities are considered to be degrees of belief, defined in terms of dispositions to gamble at certain odds. The four main evidential interpretations are the classical (e.g. Laplace's) interpretation, the subjective interpretation (de Finetti and Savage), the epistemic or inductive interpretation (e.g., Ramsey, Cox) and the logical interpretation (e.g., Keynes, Carnap).

Some interpretations of probability are associated with approaches to statistical inference, including theories of estimation and hypothesis testing. For example, the physical interpretation, is taken by followers of "frequentist" statistical methods, such as R.A. Fischer, J. Neyman and E. Pearson.

Statisticians of the opposing Bayesian school typically accept the existence and importance of physical probabilities, but also consider the calculation of evidential probabilities to be both valid and necessary in statistics. This article, however, focuses on the interpretations of probability rather than theories of statistical inference.

The terminology of this topic is rather confusing, in part because probabilities are studied within a variety of academic fields. The word "frequentist" is especially tricky. To philosophers it refers to a particular theory of physical probability, one that has more or less been abandoned. To scientists, on the other hand, "frequentist probability" is just another name for physical (or objective) probability. For those who promote Bayesian inference view, "frequentist statistics" is an approach to statistical inference that recognises only physical probabilities. Also the word "objective", as applied to probability, sometimes means exactly what "physical" means here, but is also used of

evidential probabilities that are fixed by rational constraints, such as logical and epistemic probabilities.

4. THROWING DICES OR TOSSING A COIN

The needed approach is a purely mathematical one, since the dice (or coin) is assumed to be ‘fair’ (perfect) and no interaction effect is assumed to occur from the way the throw is performed or with the environment of the dice and of its impact with a surface—nothing about the physics of the process. There are no influence factors affecting the outcome of a toss, which is *not even dependent on time*. Thus, the throws are assumed to be strictly a *repeatable* ideal process for an indefinitely long time, whence the certain, equal probability of getting each face. In addition, each face is mutually exclusive, and each throw is independent of any past or future throw. This type is also called ‘discrete probability’.

The meaning of uncertainty in this framework is that the prediction of the result of the subsequent throws is uncertain because the throwing process is strictly stochastic. No definitional uncertainty exists—for the ideal case. Speculations also exist about the effects of deviations on the long run from the strictly ideal assumption, but they enter the arena of experimental science.

Instead of dices or coins, many other frames are born that can be assimilated to the previous ones (cards, ...).

5. THE EXPERIMENTAL FRAME

The experimental frame of science is characterised by, at least, two facts:

- a. the concept of repeatability of the samples is subject to limitations, *namely in time*, so limiting also the justification to always consider the process as (only) a stochastic one;
- b. the concept of uncertainty applies not only to the unpredictability—within given (almost always finite) limits—of the next measurement result, but also to the possibility of occurrence of systematic effects, e.g. those born from epistemic reasons.

As to a., the VIM3 definition of “repeatability condition” is term 2.20 [13]: “condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the *same or similar objects* over a *short period of time*” (emphasis added). In fact, it is basically a tautology, because the expression “short period of time” means ‘so short that the conditions for repeatability hold’. Some kind of independent verification of the trueness of the condition is deemed necessary—when possible—concerning two features: (i) the stability in time of the measurand (*same or similar objects*); (ii) the repeatability of the operating conditions (all the remaining conditions).² On the contrary,

² Repeatability can be associated to the term “Type A uncertainty” of GUM, [14] not using the “error approach”. [13]

these conditions are both *assumed* as true for fair dices or coins.

As to b., systematic effects induce uncertainty components that are absent, by definition, in the case of fair dices or coins. They are also called ‘bias’,³ indicating the deviations of the measured values from a set of reference values forming what can be called a ‘reference condition’. For example, for a single additive bias B_i affecting a quantity X_i , is $X_i = \ominus X_i + B_i$ (the symbol of ‘standard state’ \ominus is borrowed here, for analogy, from physical chemistry to indicate the reference condition for which $E(B_i) = 0$), and $\ominus X_i + B_i = \ominus X_i - C_i$, where C_i is the so-called ‘correction’.⁴ [15]

A different category of systematic effects is what is causing the often-called epistemic uncertainty, [16] i.e., the one due to insufficient knowledge of known effects, which reflects into an *imperfect model* of the experimental conditions.

Actually, this category is not exhaustive, and should better be spilt into two distinct categories of uncertainty:

–epistemic (imperfect knowledge, namely in science and technique); [17-19]

–ontological (ignorance about (some parts of) the phenomenon under study). [20]

The former occurs when an influence quantity is mis-evaluated or mis-modelled. The latter comprises, e.g., the case where an influence quantity it omitted from the model because the existence of its effect was missed. In both cases, it results into an imperfect modelling.

Still a distinct category of uncertainty comes from the “definitional uncertainty” [13, 14], which should not be confused with any of the previous ones. It concerns the *non-uniqueness* of (known) definitions of the measurand: different definitions. It cannot be considered epistemic when it concerns known issues, but it is up to the judgment of the experimenter to take some of the cases into account in the model, or not: this is a model non-uniqueness, not an imperfection of it.

Ontological, epistemic and definitional uncertainties are not of stochastic nature.

Figure 1 summarises the different components of uncertainty in experimental science (see, e.g., [20]).

The possible *time dependence* of the process under investigation is another feature of the experimental frame that is useful to consider separately.

It can be due to fluctuations in time series, or it may be due to non-repeatability of compounded data series subsequent in time.

In the first case, the terms used to indicate a dependence on time (drifting, dynamic, ... systems), or invariance from time (static, stationary, ...) may be different in different scientific and technical frames, or be semantically different. About the risk of consequent confusion, the reader is directed to the review performed in [34].

The time scale itself can be a reason for different random effects showing up, an example can be seen in the two-sample variance (Allen variance) studies of frequency standards. However, one cannot construe from it that always the fluctuations with time are of random nature: a mixed effect can build up, like in the very common case of a ‘drifting’ characteristics of an instruments from its initial calibrated state.

<p>(2) <u>Unknown-Knowns</u> <i>hidden knowledge</i> about an imperfectly/ incompletely-known subject matter</p> <p>(epistemic uncertainty)</p>	<p>(1) <u>Known-Knowns</u> <i>full knowledge</i> only possible in cases such: dices, coins, cards, ... probability type A</p> <p>(stochastic uncertainty)</p>
<p>(4) <u>Unknown-Unknowns</u> <i>ignorance</i> factor totally ignored</p> <p>(ontologic uncertainty)</p>	<p>(3) <u>Known-Unknowns</u> <i>gap in knowledge</i> discovery</p> <p>(epistemic uncertainty)</p>

Figure 1. Different components of uncertainty in experimental science, in the four possible combinations of known and unknown information.

5. DIFFERENCES BETWEEN THE TWO FRAMES AND SOME CONSEQUENCES

“Aleatory uncertainty represents an absolute limit. To use the coin toss example, having thrown the coin a thousand times we would be able to express with confidence the *probability* of a heads occurring, but that is all we can say about the next coin toss” (emphasis added) [20].

Probability cannot have the same meaning in experimental science, outside the aleatoric components of uncertainty.⁵

A useful classification of uncertainty components can be found in [21]:

- (i) Uncertainty in influence quantities,
- (ii) Uncertainty in model,
- (iii) Uncertainty in model parameters.

³ “Measurement bias: estimate of a systematic measurement error” (term 2.18 in [13]).

⁴ Notice that, in general, $E(C_i)$ is what is intended for ‘correction’.

⁵ One can even argue if probability is the only concept that can be used to deal with chance.

Components (i) can be either epistemic if referred to influence quantities properties, or random if referred to their measured values, or both. The influence quantities are normally subdivided into two groups: the “basic” ones (called “input quantities” in [14]) and the “derived” ones (i.e., the ones responsible for ‘bias’ [13] and needing ‘corrections’) [14]. The former are measured in indirect measurements (thus having at least one Type A uncertainty component), while the latter can be measured, or their values are obtained or inferred from previous knowledge (thus having in the first case at least one Type A uncertainty component, or only Type B components in the second).⁶

Component (ii) can be epistemic if referred to model imperfection, or ontological if referred to missed quantities. It also includes the definitional uncertainty.

Component (iii) can be stochastic if the values are obtained from measurement, or epistemic if the values are computed or inferred.

Ontological uncertainty is generally not included in the budgets of experimental science,⁷ and definitional uncertainty is generally resolved by specifying the relevant type of definition—otherwise it becomes an ontological uncertainty component, excluded from treatment in the new GUM [32].

Epistemic uncertainty is often ‘randomised’, i.e. transformed into a stochastic component, by assuming ignorance about the position parameter, e.g. by assuming a null mean and estimating a range for the resulting uncertainty component. The latter can be set as an interval (e.g. “Maximum Permissible Error” (MPE, term 4.26 in [13]) or “Worst Case Uncertainty” (WCU) [33], or a non-probabilistic interval), or as the standard deviation, or a multiple of it, set by the chosen confidence interval (or degree of believe). [14]

6. AN EXAMPLE AND FINAL REMARKS

From the previous review it turns out that the case of dices and similar are, in experimental science, an over-simplification of a much more complex structure of uncertainty, such that the former concerns only the intrinsic stochastic component of it.

In the latter case, instead, the systematic effects are, in general, the major concern. Though they may have a stochastic component, they mainly involve, after all, the need of an assessment that involves basically a subjective judgment requiring a decision. A particularly useful discussion of this situation can be found in [22], where the seminal case (but a wider range of similar cases in measurement applies) of the treatment of uncertainty in the assessment of the values of fundamental constant, is treated from the viewpoint of psychology of measurement and decision theory—an interesting viewpoint basically ‘external’ to the metrology field.

Different reasons for ‘bias’ in the judgment are considered, bringing to what is called there “overconfidence”: “In several sets of analyzed measurements of physical constants, we have found consistent replication of a robust finding of laboratory studies of human judgment: reported uncertainties are too small. How could this apparent overconfidence arise? Experimental studies of human judgment have shown that such biases can arise quite unintentionally from cognitive strategies employed in processing uncertain information”.

However, not always this attitude is unintentional: one case “concerns the procedures chosen to assess the uncertainty. The recommended practice in physics is to consider all possible sources of systematic uncertainty when reporting results. However, without specific guidelines regarding what to consider and explicit recognition of the *subjective elements* in uncertainty assessment, one cannot be sure how comprehensively individual scientists have examined the uncertainty surrounding their own experiments. Conceivably, some of the apparent overconfidence reflects a deliberate decision to ignore the herder-to-assess sources of uncertainty” (emphasis added); “A second possible source of bias is that, unlike laboratory experiments on judgment, which can take great care to ensure that subjects are motivated to express their uncertainty candidly, real-world settings create other pressures”.

In the latter respect, precisely, “having a pre-existing recommended value may particularly encourage investigators to discard or adjust unexpected results, and so induce correlated errors in apparently independent experiments”. This has been observed in several circumstances, according to these authors [22], namely for the speed of light in vacuum c_0 , the basis for the unit of length,⁸ and the inverse of the fine structure constant, α^{-1} , the basis for the value of the electron charge e , so of the unit of electrical current.

The analysis in [22] is centred only on the *uncertainties* associated to the recommended values, while the aim of the actual CODATA Task Group [23] is also to *adjust the values* of the constants using the Least Squares Analysis (LSA) method, instead of the use of the mean—or another strictly statistical parameter—of the experimental values (based on probability). [24]

The accent on the subjective side of the uncertainty analysis, (see also [25]) may seem to rule out ‘frequentist’ methodology; however, this does not necessarily mean that the Bayesian one can always be better used instead. In Chapter 1 of [22], pag. 30-31, the authors say: “Bayes theorem is an uncontroversial part of probability theory. Bayesian inference is more controversial, because it treats probabilities as subjective, thereby allowing inferences that combine diverse kinds of evidence. Frequentistic probabilities requires evidence of a single kind (e.g. coin flips). Subjective judgements are only probabilities if they pass coherence tests. Thus, *probabilities are not just any assertion of belief*” (emphasis added).

⁶ Prior and posterior are not used in this paper necessarily in the Bayesian sense.

⁷ Several decades ago this was done sparingly, but systematically by the USSR school, as “undetected systematic errors”.

⁸ Notice that the history of the numerical values of c_0 stopped in 1983 with the ‘stipulation’, misinterpreted as a definitive value.

Actually, CODATA position had oscillations in the years, starting from a position where its major asset was not the use of the specific analytical treatment (the LSA) [26], but the preliminary critical review of the data and their screening. Later they shifted to a position where all the available data were used [27]. This was the 2010 position—presently the last one. Possibly, in addition to the criticism about the subjective nature of the screening, an abundance of data prompted the last decision, since a few outlying data would not anymore critically affect the resulting value and associated uncertainty.

In [22] a statistical analysis is performed on the reliability of the estimates of the constants values, namely of 40 recommended values for the period 1928-1973. A “surprise index” is computed (percent which fall outside the assessed 98% confidence interval, or outside 2.33s) and found to be an impressive 57%. In more recent times, the studies have dramatically lowered the experimental uncertainty, so the former statistics could have improved a lot. However, the already cited “bandwagon” effect of the recommended values could now have increased in importance, deserving an urgent update of analyses of this type.

In [22] the conclusions about that specific case are: “The underestimation of uncertainty in measurements of physics constants and compilations of recommended values seems to be pervasive. This evidence extends previous findings of overconfidence in laboratory studies of human judgment to a task domain of great practical importance. If reported uncertainties do not reflect the magnitude of actual errors, whether due to incomplete analysis or to judgment biases, the usefulness of those measurements is significantly diminished”. This is critical, e.g., for use in the “New SI” [24].

In general, it results that the use of examples concerning the case of dices, coins and similar, has very little value in studying the treatment of *experimental uncertainty*, and can even be deceiving, so, in general, it should be avoided.

7. APPENDIX: FALSIFICATION IN EXPERIMENTAL SCIENCE

The hereinbefore-asserted subjectivism in experimental sciences, with the need of judgment and decision, entrains the need of “verifications” (from Wittgenstein on) or of “falsification” (from Popper on) criteria.

The first method was later prevalently considered an impossible goal to reach in the lack of a general criterion about the sufficient number of verifications, and the difficulties arising from unavoidable epistemological limitations.

The second method was born *basically in an uncertainty-free context*: however, we learned that in measurement a single occurrence of falsification, even if reasonably proved, cannot be considered sufficient in the frame of uncertain knowledge. Not only repeated occurrences are needed, but “falsification is not possible without some threshold deviation which would be

considered sufficiently unlikely to reject the theory” [28]—also [29] is interesting, though one not necessarily always may share author’s opinions.

Early Popper (1936) said: “With the idol of certainty (*including that of degrees of imperfect certainty or probability*) there falls one of the defences of obscurantism which bar the way of scientific advance“, and “The relations between *probability* and experience are also still in need of clarification. In investigating this problem we shall discover what will at first seem an almost insuperable objection to my methodological views. For although *probability statements* play such a vitally important role in empirical science, they turn out to be in principle *impervious to strict falsification*” (emphases added). [30] Later he managed reconciling the probability concept by proposing first the theory of “propensity”, a further interpretation of probability. [31]

REFERENCES

- [1] L.J. Savage, The foundations of statistics. New York: John Wiley & Sons, Inc. (1954). ISBN 0-486-62349-1.
- [2] F. Pavese and P. De Bièvre, Fostering diversity of thought in measurement science, in Advanced Mathematical and Computational Tools in Metrology and Testing X (F. Pavese, W. Bremser, A. Chunovkina, N. Fischer, A.B. Forbes, eds.), Series on Advances in Mathematics for Applied Sciences vol. 86, World Scientific, Singapore, 2015, pp. 1–8.
- [3] Wikipedia, term “Probability interpretations”, as consulted on February 24, 2015, and references therein.
- [4] Hájek, Alan, Zalta, Edward N., ed., Interpretations of Probability, The Stanford Encyclopedia of Philosophy, online.
- [5] de Elía, Ramón; Laprise, René, Diversity in interpretations of probability: implications for weather forecasting, Monthly Weather Review 133 (2005) 1129–1143.
- [6] Enric Trillas, Some uncertain reflections on uncertainty, Archives for the Philosophy and History of Soft Computing, International Online Journal, Issue 1 (2013).
- [7] Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1 (1978) 3–28. doi:10.1016/0165-0114(78)90029-5.
- [8] Dubois, Didier; Henri Prade (1985). Théorie des possibilités. Masson, Paris.
- [9] Wikipedia, term “imprecise probability”, as consulted on June 30, 2015, and references therein.
- [10] G.L.S. Shackle, A Non-Additive Measure of Uncertainty, The Review of Economic Studies 17 (1949-1950) 70-74
- [11] I. Gilboa, D. Schmeidler, Additive representations of non additive measures and the Choquet integral, Annals of Operations Research 52 (1994) 43-65
- [12] D. Skulj, Non-additive probability, 2002. <https://www.stat.aau.at/Tagungen/Ossiach/Skulj.pdf>.
- [13] BIPM (2012) International vocabulary of basic and general terms in metrology (VIM), 3rd edn. BIPM/ISO, JCGM 200:2012. 2008 version with minor corrections. <http://www.bipm.org/en/publications/guides/vim.html>.
- [14] BIPM, Guide for the expression of uncertainty in measurement; JCGM 00:2008, ISO Geneva, at <http://www.bipm.org/en/publications/guides/gum.html>.
- [15] F. Pavese, Key comparisons: the chance for discrepant results and some consequences, Acta IMEKO, 2015, in press.
- [16] Wikipedia, term “uncertainty quantification”, as consulted on June 30, 2015, and references therein.

- [17] L.P. Swiler, T.L. Paez, R. L. Mayes, Epistemic Uncertainty Quantification Tutorial, Proceedings of the IMAC-XXVII, 2009 Orlando, Florida USA, 2009 Society for Experimental Mechanics Inc.
- [18] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? Does it matter?, Special Workshop on Risk Acceptance and Risk Communication, March 26, 2007, Stanford University, pp. 1–13.
- [19] T. O’Hagan, Dicing with the unknown, *Significance* 2005, 132–133.
- [20] M. Squair, Epistemic, ontological and aleatory risk, 2009, blog <http://criticaluncertainties.com/2009/10/11/epistemic-and-aleatory-risk/>.
- [21] F. Pavese: “Mathematical and statistical tools in metrological measurement”, 2013, Chapter in Physical Methods, Instruments and Measurements, [Ed. UNESCO-EOLSS Joint Committee], in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK, <http://www.eolss.net>.
- [22] M. Henrion, B. Fischhoff, Assessing Uncertainty in Physical Constants, Ch. 9 in “Judgement and Decision Making“ (B. Fischhoff, Ed.) Routhledge (2013) pp. 172-187, 113649734X, 9781136497346, taken from American Journal of Physics 54, 791 (1986); doi: 10.1119/1.14447.
- [23] CODATA <http://physics.nist.gov/cuu/Constants/index.html> , <http://www.bipm.org/extra/codata/> .
- [24] F. Pavese, Some problems concerning the use of the CODATA fundamental constants in the definition of measurement units, *Metrologia* 51 (2014) L1–L4, with online Supplementary Information.
- [25] F. Pavese, Subjectively vs. objectively-based uncertainty evaluation in metrology and testing, IMEKO-TC1-TC7-2008-030, available on <http://www.imeko.org/index.php/proceedings#> under TC7 (2008); F. Pavese, On the degree of objectivity of uncertainty evaluation in metrology and testing” *Measurement* 2009 42 1297–1303.
- [26] B.N. Taylor, NBSIR 81-2426, 1982, National Bureau of Standards (today NIST), Gaithersburg USA.
- [27] P.J. Mohr, B.N. Taylor and D.B. Newell, CODATA Recommended Values of the Fundamental Physical Constants: 2010, *Rev. Modern Phys.* 84 (2012) 1–94.
- [28] J.R. Grozier, Can Theories be Falsified by Experiment?, MSc essay (UCL, London, 2012).
- [29] J.R. Grozier, Falsificationism, Science and Uncertainty, MSc Dissertation, London Centre for the History of Sciences, Technology and Medicine, 2013, pp. 1–29.
- [30] K. Popper, *The Logic of Scientific Discovery* (Taylor & Francis e-Library ed.). London and New York: Routledge / Taylor & Francis e-Library (2005).
- [31] K. Popper, The Propensity Interpretation of the Calculus of Probability and of the Quantum Theory. In *Observation and Interpretation*, Korner & Price (eds.), Butterworth Scientific Publications, 1957, pp. 65–70.
- [32] JGCM, Guide for the expression of uncertainty in measurement, Draft of 14 December 2014 delivered for comments to ISO TC69.
- [33] da Silva R.J., Santos J.R., Camões M.F., Worst case uncertainty estimates for routine instrumental analysis, *Observations on the worst case uncertainty*, *Analyst* 127 (2002) 957–963; L. Fabbiano, N. Giaquinto, M. Savino, G. Vacca, Observations on the worst case uncertainty, *Journal of Physics: Conference Series* 459 (2013): 2043, doi:10.1088/1742-6596/459/1/012038.
- [34] K.H. Ruhm, Are fluctuating measurement results instable, drifting or non-stationary? - A survey on related terms in

Metrology, Journal of Physics Conference Series 459 (2013): 2043, doi: 10.1088/1742-6596/459/1/012043.