CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Political Studies

Department of Security Studies

Master's Thesis

2019 **Matěj Štěpánek**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Political Studies

Department of Security Studies

The Art of (In)Accuracy: A Meta-Analysis of the European Values Think-Tank´s Forecasts

Master's thesis

Author: Bc. Matěj Štěpánek

Study programme: Security Studies

Supervisor: RNDr. Jan Kofroň, Ph.D.

Year of the defence: 2019

Declaration

1. I hereby declare that I have compiled this thesis using the listed literature and resources only.
2. I hereby declare that my thesis has not been used to gain any other academic title.
3. I fully agree to my work being used for study and scientific purposes.

In Prague, July 30, 2019 Matěj Štěpánek

References

ŠTĚPÁNEK, Matěj. *The Art of (In)Accuracy:A Meta-Analysis of the European Values Think-Tank´s Forecasts.* Praha, 2019. 105 pages. Master's thesis (Mgr.). Charles University, Faculty of Social Sciences, Institute of Political Studies. Department of Security Studies. Supervisor RNDr. Jan Kofroň, Ph.D.

Length of the thesis: 195 519 characters (with spaces; without footnotes, list of references and list of appendices)

Abstract

Probabilistic forecasts represent a potentially indispensable tool for policy advising, strategic planning, or provision of possible scenarios of future development. It is clear, however, that inaccurate forecasts can entail serious consequences. At best, unsuccessful forecasting attempts may discredit such potentially valuable method in the eyes of decision-making elites. At worst, wrong predictions may lead to the misallocation of scarce resources or to the unnecessary securitization. Nonetheless, probabilistic forecasts have seldom been used in the realm of the Czech security analyses, studies, or debates. Thus, the European Values Think-Tank´s research project is a pioneering attempt to utilize the probabilistic forecasting in the Czech politico-security sphere. Due to the fact that the think-tank developed its probabilistic forecasts to help the Czech security elite with strategic planning, the thesis aims to verify the accuracy and predictive capabilities of the European Values. The broader goal is to bring, by the accuracy assessment, the rigor into the Czech probabilistic-forecasting debate. Additionally, the thesis also compares the predictive capabilities of the European Values with the alternative – foreign – forecasts, as well as with other means of accuracy verification. The results of the accuracy analysis, as well as the comparison, show that the predictive capabilities of the European Values can be seen – for now – as poor ones. The thesis, therefore, closes by offering key recommendations for the improvement of the accuracy of any future enterprises.

Abstrakt

Pravděpodobnostní předpovědi představují potenciálně nezbytný nástroj pro politická doporučení, strategické plánování a možné scénáře budoucího vývoje. Na druhou stranu je zřejmé, že nepřesné předpovědi mohou vést k vážným důsledkům. V lepším případě mohou neúspěšné předpovědi vést ke zpochybňování takto potenciálně užitečné metody v očích politických činitelů, v tom horším k chybné alokaci již tak omezených zdrojů či zbytečné sekuritizaci. V českých bezpečnostních analýzách, studiích či debatách byly pravděpodobnostní předpovědi zatím využívány zřídka. Z tohoto hlediska představuje výzkumný projekt think-tanku Evropské hodnoty průkopnický pokus o využití pravděpodobnostních předpovědí v české politicko-bezpečnostní oblasti. Vzhledem k tomu, že think-tank vytvořil pravděpodobnostní předpovědi za účelem pomoci českým bezpečnostním elitám při tvorbě strategického plánování, je cílem práce prověřit přesnost a prediktivní schopnosti Evropských hodnot. Širším záměrem je pomocí vyhodnocení přesnosti přinést do české debaty téma preciznosti pravděpodobnostního předpovídání. Práce také porovnává prediktivní schopnosti Evropských hodnot s alternativními – zahraničními – předpověďmi a dalšími nástroji pro ověřování přesnosti. Výsledky analýzy přesnosti i porovnání ukazují, že prediktivní schopnosti Evropských hodnot lze považovat – prozatím – za slabé. Na závěr proto práce poskytuje klíčová doporučení pro zlepšení přesnosti všech budoucích iniciativ.

Keywords

Probabilistic forecasts, Future, Security environment, Europe, Czech Republic

Klíčová slova

Pravděpodobnostní předpovědi, budoucnost, bezpečnostní prostředí, Evropa, Česká republika

Název práce

Umění (ne)přesnosti: Meta-analýza předpovědí Think-tanku Evropské hodnoty

Acknowledgement

I would like to express my gratitude to RNDr. Jan Kofroň, Ph.D. for his constructive comments, expert advice and helpful guidance during the writing process of my thesis.

I would also like to thank Lori, mom, and my family for their endless support and words of encouragement.

Table of Contents

Introduction: In Pursuit of the Forecast

Anticipating the future represents one of the major human desires. Throughout every historical era of humanity, our fundamental cognitive architecture propels us to tackle the long-standing enemy: the uncertainty about the future. To borrow an expression from famous probabilistic statistician Bruno de Finetti: "In almost all circumstances...we all find ourselves in a state of uncertainty".[1] But this perennial condition is equally a very unpleasant one to our biological settings. We – as human beings – naturally chase both the certainty and its less "resonant" types (e.g. high probabilities, etc.), as much as we crave the patterns of conditionality and predictability in our world. We have therefore been historically fascinated by oracles, prophets, clairvoyants, whole religious and philosophical systems; even by Laplacian conception of knowledge or Newtonian scientific determinism.[2] In this regard, we utilize predictions and forecasts as a more accessible form of an antidote to the uncertainty.

Hence, we are all forecasters in our daily lives. When thinking about the fastest way home, launching a business (a matter of potential benefits), leaving school or about the solution in a dangerous situation (a matter of survival), we try to base "appropriate" decisions on our expectations about yet unknown future dynamics.[3] This relates to another remarkable – practical rather than purely cognitive and biological - property of predictions and forecasts that steers us to use them. Predictions and forecasts serve as our guidance in decision-making by adumbrating us a range of probable future outcomes; and also by providing us with an early-warning tool and time to prepare counteraction to the processes before they really unfold.[4] Not surprisingly, and despite a resonant critique emanating from ontological and epistemological shortcomings (as well as from notorious cases of forecasting failure, such as was the Financial Crisis of 2007-2008) of the "foresight", many scientific disciplines and relevant institutions across various fields have therefore accepted forecasting as a pivotal method of their inquiry. Meteorologists are traditionally called upon to predict the weather; economists routinely give forecasts of economic rebounds, national banks forecast inflation levels or output growth, doctors of medicine assess the probability of patient´s survival.[5] Even while writing this diploma thesis, the Bank of England has forecasted a 22% probability of

[1] DE FINETTI, Bruno. Theory of probability: A critical introductory treatment. John Wiley & Sons, 2017. p. 21.
[2] KRISTÓF, Tamás. Is it possible to make scientific forecasts in social sciences?. Futures, 2006, 38.5: p. 564.
[3] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016. p. 1.
[4] SLOVIC, Paul; LICHTENSTEIN, Sarah. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational behavior and human performance, 1971, 6.6: p. 652.
[5] LAI, Tze Leung, et al. Evaluating probability forecasts. The Annals of Statistics, 2011, 39.5: p. 2356.

shrinking of the British economy in 2019 (as a consequence of "Brexit" dynamics).[6] It should also be noted that some of these disciplines have gradually improved their accuracy in forecasting the events or trends by using statistical models or "expert" knowledge, but above all, by using the probabilities, (which lie at the semantic heart of forecasting; see in the next chapter), accuracy checking and continuous revision of forecasts (so-called Bayesian thinking).[7]

The ability to make predictions and forecasts as much accurate as possible, and to revise them if needed, is sometimes even more crucial in the political-security area. For decision-making stakeholders, forecasting plays an indispensable role in policy advising, provision of possible scenarios, and strategic planning (e.g. for next steps in foreign policy or development of the Grand Strategies). The same holds for the readiness in the military sphere (e.g. the military simulations). In terms of enhancing and facilitating anticipation of the threats to national security, forecasts are essential.[8]

Nonetheless, forecasts – and especially the probabilistic ones – have seldom been used in the realm of the Czech security analyses, studies and debates, both practically (e.g. consultations for the government) and scientifically. We prefer explanation rather than prediction within our political-security discourse. Thus, the European Values Think-Tank´s research project - *Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016-2019*[9] – is both a unique attempt and a significant breakthrough in utilizing probabilistic forecasting. The "foresight" has finally been brought into the Czech political-security debate. But the European Values forecasts have also received attention due to its flagrant alarmism (see Kalous, 2018).[10] It is clear, however, that flawed predictions and forecast can entail serious consequences. Whereas the wrong allocation of resources for political action, panic, unneeded securitization or missed opportunities represent the less serious consequences, the unnecessary loss of human lives is the worse one (e.g. false forecast before the 2003 Invasion of Iraq).[11] This should not give the

[6] The road not taken: Brexit has not caused much economic damage. Until now. In: The Economist [online]. [cit. 2019-03-03]. Available from: https://www.economist.com/britain/2019/02/16/brexit-has-not-caused-much-economic-damage-until-now

[7] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 14.

[8] MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: p. 418-419.

[9] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

[10] KALOUS, Miroslav. HOW (NOT) TO PREDICT THE FUTURE?: Analysis of several pioneering studies in the field of Czech political and security scenario-building. In: Obrana a strategie [online]. [cit. 2019-03-03]. Available from: https://www.obranaastrategie.cz/en/archive/volume-2018/1-2018/articles/how-not-to-predict-the-future.html

[11] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016. p. 3.

impression that Czech political-security decision-makers make choices based on the European Values forecasts, but on the other hand, the European Values Think-Tank is not an insignificant institution without any power to shape our security discourse. Conversely, the Think-Tank provides Czech decision-makers with expert recommendations and works as a platform for dialogue amongst – not only – politicians and experts.[12]

The main purpose of this diploma thesis is, therefore, to assess the accuracy and predictive capabilities of the European Values forecasts. However, the aim is not "just" to draw attention to the potential flaws in forecast results capable of – at least to some extent – leading to false conclusions of the Czech decision-makers. Since the European Values´ research project represents the first serious attempt to employ the probabilistic forecasting, the broader goal is to feed into the debate about, and bring the rigor into, the forecast-making. Following first research question "*How accurate are the European Values forecasts using classical methods of validation?*" emanates from the above mentioned main "accuracy-checking" purpose. Yet it subsequently encompasses the broader goal – to bring, by checking the accuracy, another level of rigor into the incipient forecasting efforts in the realm of the Czech security; to point out possible methodological errors and thus help with the future improvement of the forecast-making.

The second research question of "*How does the European Values´ research project perform compared to the results of alternative forecasts?"* pursues – too – the main as well as the broader goal, this time by the comparison of the think-tank´s results. The fact is that the think-tank´ accuracy alone tells us too little about whether its result is relatively good or rather a poor one. Good results are needed in order to avail unquestionable benefits of forecasting and calm the critical voices of forecasting pessimists. Thus, the true evaluation lies in comparing the European Values´ accuracy with the accuracy of alternative foreign institutions, as well as with other means of forecast (accuracy) comparison.

Moreover, both an analysis and identification of potential shortcomings of the European Values´ research project naturally pave the way for eventual recommendations. Nevertheless, it is reasonable to reaffirm that the main motivation of this work is not to criticise the European Values´ research project. It deserves our recognition, hopefully, for kick-starting the new and potentially very useful method for political-security inquiry in the Czech Republic.

[12] Ideologové, nebo odvážní bojovníci s propagandou? Think tank Evropské hodnoty dráždí politiky. In: iROZHLAS [online]. [cit. 2019-07-20]. Available from: https://www.irozhlas.cz/zpravy-domov/ideologove-nebo-odvazni-bojovnici-s-propagandou-think-tank-evropske-hodnoty-drazdi-politiky_201701131300_dbernardy

The thesis is divided into seven chapters. The first chapter considers the great philosophical debate regarding the possibility of knowledge, which is furthermore reflected in views of forecast optimists and sceptics. Chapter two explains the methodology of bringing the verification process into the think-tank´s research project. The European Values´ research project, along with the presentation of the key data-sets, is finally introduced in the third chapter. The analysis of the data and the assessment of the think-tank´s accuracy are addressed in the fourth and the fifth chapter. These two chapters should also highlight accuracy errors and methodological flaws of the research project. Subsequent accuracy-comparison between the European Values´ and alternative foreign forecasts, or other means of forecast verification, is addressed in the sixth chapter. Based on the six-chapter findings, the seventh chapter offers some recommendations for future improvement of forecast making in the Czech Republic. The last part of the thesis summarizes the key findings.

1. THEORETICAL FRAMEWORK

Does it make sense to try to build scientific probabilistic forecasts (i.e. based on the collection and an assessment of data) in social sciences? This question emanates from a more fundamental and traditional debate within the philosophy of social science, concerning the possibility of anticipating the future, thus, the possibility of knowledge as such. And both forecasting optimists and sceptics more or less base their arguments on such a dichotomy. Whereas extreme optimists perceive forecasts almost as a panacea for uncertainty, sceptics liken it to a Chimera.

1.1 Possibility of Knowledge and the Philosophy of Science

From the perspective of knowledge, forecast ultra-optimists sympathize, to some extent, with thoughts of several philosophical streams. Two of them – logical empiricism and scientific determinism – are in agreement with positivism. Logical empiricism, which revises the positivism of the Vienna Circle, is committed to the idea that the orderliness, linearity, and causality of the natural world also exist in social processes. Since these "universal laws" are objectively observable and hence open to our knowledge, we should be able to explain and predict social phenomena. The future is predictable and observation is all that is required.[13] This belief leads to another assumption, often criticized by forecast sceptics for being perilous (see below): the growth of knowledge along with an increase in predictions of future events will enable us – but mainly the researchers and scientists – to both guide and change social dynamics (of course, under the political-leadership control). Such social engineering bears a striking resemblance to ideas of French sociologist Henri de Saint-Simon (scientists should guide society).[14]

Scientific determinism, too, links the possibility of objective – meaning also scientific - knowledge with the possibility of making predictions and forecasts about the future. This stream of thought similarly believes in the existence of complete and certain knowledge. Therefore, a gradual gathering of information will get us closer to the universal theories about the past and present. From this point forward, the future of social processes can be rationally calculated and our social world is as deterministic as, for example, classical mechanics or

[13] BALDUS, Bernd. Positivism's twilight?. Canadian Journal of Sociology/Cahiers canadiens de sociologie, 1990, p. 151.
[14] VAN VUGHT, Frans A. Pitfalls of forecasting: fundamental problems for the methodology of forecasting from the philosophy of science. Futures, 1987, 19.2: p. 186.

modern physics.[15] Moreover, this past-present-future logic can be found in the Harold Lasswell´s thinking ("it is possible to be oriented in the continuum of past, present, and future dynamics") or Herman Kahn´s Projective approach stressing that present and immediate past will make future anticipation much simpler.[16] The optimistic view about the increasing collection of data itself (about past and present, hence about the future) arises out of Laplacian omniscient conception of knowledge. Pierre-Simon Laplace asserted that growing knowledge must entail greater predictability. In the end, having all the data about the present means certainty of everything in the future.[17]

Another philosophical stream providing forecast optimists, and especially the proponents of Futures studies with some theoretical foundation, is Critical realism. But this time not so deterministically. Critical realism again suggests the possibility of knowing the future based on what knowledge we have about the past and the present. However, the knowledge is what we are justified to believe, and even though some predetermined elements (causal structures) about the world and social processes can be uncovered by data and logical deduction, something is fundamentally unpredictable. Thus, the future as such is uncertain, but there is at least something in the future that can be predicted. In comparison with the previous two other streams, critical realism is not perceived as purely utopian by knowledge-pessimists and forecast sceptics.[18]

Unsurprisingly, forecast sceptics´ thoughts emanate from a different philosophical position on the possibility of knowledge, and consequently, on predicting. In this case, it does not really make sense to compare processes of society to the motion of physical objects. Newtonian laws of nature do not apply to the complex, non-linear nature of society (see Lorenz´s Chaos theory bellow). Karl Popper, therefore, argued that since there are only trends and tendencies in social dynamics, not predictable "scientific" laws, we simply cannot predict the future because every trend may immediately change at any moment.[19] Moreover, stringent version of such indeterminism challenges even the existence of natural laws, hence, the possibility of achieving the objective, scientific knowledge. One can clearly see that by looking at the Hume´s "Problem of tomorrow" (originally referring to the question of inductive-deductive reasoning). Although the sun has risen every day so far, it may not, indeed, rise tomorrow. To say the obvious, it is not that Hume´s problem of tomorrow seeks

[15] HOSNI, Hykel; VULPIANI, Angelo. Forecasting in light of big data. Philosophy & Technology, 2018, 31.4: p. 561.
[16] STRAUSS, Harlan J.; ZEIGLER, L. Harmon. Delphi, political philosophy and the future. Futures, 1975, 7.3: p. 185-186.
[17] TETLOCK, Philip E.; GARDNER, Dan. Superforecasting: The art and science of prediction. Random House, 2016, p. 9.
[18] VAN DER HEIJDEN, Kees. Scenarios and forecasting: two perspectives. Technological forecasting and social change, 2000, 65.1: p. 31-34.
[19] KRISTÓF, Tamás. Is it possible to make scientific forecasts in social sciences?. Futures, 2006, 38.5: p. 563.

to criticise natural laws, but in our context, it can be used to highlight an important issue of knowledge and forecast sceptics: we can never know the future will be like past, so the uncertainty is omnipresent.[20]

According to famous philosopher Nicholas Rescher, terminal limits and confines of the possibility of reaching an all-encompassing objective scientific knowledge do exist in our world. For Rescher, who accepts Kantian limits of science (limits of our experience set limits to our science), one of the main impediments to the scientific knowledge emerges from the fact that the science itself is an interaction between us (human beings) and nature. On the one hand, we have developed science while being "endowed" with our bounded cognitive and knowledge capabilities. But on the other hand, the reality of – predominantly –social phenomena is complex.[21] These epistemologically-bounded capabilities lead to the aforementioned past-present-future question and precisely to the explanation-prediction debate. It is worth recalling in this regard that explanation-prediction debate (whether to explain or to predict) – but this time concerning the role of theories – has also been an important factor in shaping opinions on forecasting (see below and in Milton Friedman, 1953).[22]

Nevertheless, Rescher´ s thinking implies that we should explore along reasonable lines of past and present. From the perspective of the possibility of knowledge, an explanation is to some extent attainable, whereas predicting the future is unfortunately inaccessible. Epistemologically, there is much cleaner causal linkage of knowledge in explanation, than in prediction. Methodologically, we can explain some processes (once they occur) that are unpredictable; and from the ontological viewpoint, it is much easier to deal with past and present (recently unfolded futures) developments than with those "not yet" unfolded and still open.[23]

It is clear, however, that the text above sets out to expound some essential philosophical premises – regarding the question of the possibility of knowledge in the realm of societal processes and social studies – which either forecast optimists or pessimists somehow directly or indirectly refer to. To put it very plainly and as Weber´s ideal types, the optimists more or less believe in the possibility of achieving the objective scientific

[20] PRIEST, Graham. The sun may not, indeed, rise tomorrow: a reply to Beall. Analysis, 2012, 72.4: p. 740-741.
[21] GUILLAN, Amanda. Epistemological limits to scientific prediction: The problem of uncertainty. Open Journal of Philosophy, 2014, 4.04: p. 510-511.
[22] FRIEDMAN, Milton. The Methodology of Positive Economics. In: HAUSMAN, Daniel M. The Philosophy of Economics: An Anthology. 3rd ed. Cambridge : Cambridge University Press, 2008. p. 145-178.
[23] GUILLAN, Amanda. Epistemological limits to scientific prediction: The problem of uncertainty. Open Journal of Philosophy, 2014, 4.04: p. 512-513.

knowledge (linearity, causality, laws) and thus also perceive predictions and forecast not only as possible, but also as acts of rationality. Furthermore, making of scientific predictions and forecast should develop into a gold standard of science.[24] Conversely, sceptics challenge the notion of attainable objective scientific knowledge, either in ontological or epistemological terms (see also below). Moreover, the relentless pursuit of the future is both scientifically irrational and fallacious, and the predictions and forecast will always be just a matter of random guess.[25]

Following part introduces the key arguments of forecast and prediction optimists as well as sceptics, but this time from the perspective of international relations. This is particularly important for the nouveau Czech debate about the use of these methods in the political-security realm.

1.2 Optimism: Foresight is Real

The experts and politicians have been assiduously called upon by forecast and prediction optimists to utilize these tools more for the political-security purposes in the domain of international relations. While some first foundations for prediction and forecast* intellectual groundwork were already laid – besides aforesaid philosophical thinking – by Jacob Bernoulli in the 18th century, it was not until the 1950s that some disciplines started to apply them seriously (also with the help of the researchers such as Egon Brunswik or Bruno de Finetti).[26] Arguments concerning the growing amount of information data in our world are among the most powerful ones. That is to say, as the processes of our modern world (e.g. Age of Big Data, globalization dynamics, etc.) continue to perpetuate change and uncertainty, decision-makers are increasingly reliant on attempts to anticipate future dynamics and its possible outcomes. In that sense, predictions and forecasts are richly advise-capable and guidance-capable.[27] Furthermore, they epitomize a necessary endeavour if we attempt to counteract perilous dynamics such as conflicts, arms races or climate change.[28] Other resonant

[24] WARD, Michael D., et al. Learning from the past and stepping into the future: Toward a new generation of conflict prediction. International Studies Review, 2013, 15.4: p. 4.
[25] GARTZKE, Erik. War is in the Error Term. International Organization, 1999, 53.3: p. 567.
* For some authors, there is no difference between predictions and forecasts. However, many scholars claim that the difference is in the type of data (prediction: estimating for the unseen data; forecast: predicting on the basis of time-series data. Nevertheless, this thesis accepts following the definition: both prediction and forecast explore the unknown future, but only the forecasts are conceived with a probability range (see Choucri, 1974).
[26] LICHTENSTEIN, Sarah; FISCHHOFF, Baruch; PHILLIPS, Lawrence D. Calibration of probabilities: The state of the art. In: Decision making and change in human affairs. Springer, Dordrecht, 1977, p. 276.
[27] FREEMAN, John R.; JOB, Brian L. Scientific forecasts in international relations: Problems of definition and epistemology. International Studies Quarterly, 1979, 23.1: p. 113.
[28] SCHNEIDER, Gerald; GLEDITSCH, Nils Petter; CAREY, Sabine C. Exploring the past, anticipating the future: A symposium. International Studies Review, 2010, 12.1: p. 3.

arguments emanate from the respectable track record of various disciplines´ methods and institutions ranging from meteorology and economics to even Sherman Kent´s intelligence analysis.[29]

1.2.1 A Fascination with Economics, and the Role in IR Theories and Decision-making

The optimists´ thirst for prediction and forecast making in the field of international relations (or security) often relates to an admiration for other disciplines widely using them as standard and routine practices. And one of these endowed disciplines is economics. As it is indicated in the introductory part of this diploma thesis, financial institutions and experts as well as national banks seek to eliminate the uncertainty and anticipate the future with predictive inflation reports, forecasted portfolio values, nominal Gross National Product forecasts or stock index predictions. For that purpose, sub-disciplines of economics have also developed many models useful for predicting and forecasting. Some of them are based on assumptions about the predictable and expectable behaviour of individuals and markets. Thus, economists postulate the economic rationality (i.e. also predictability) of individuals and their business choices, precisely through the lens of "optimal gambles" and efforts to maximize expected utility.[30]

The second type of models employed to make predictions and forecasts is a more mechanical one. For example, these include Bayesian vector autoregression (BVAR) and Vector autoregression models (studied by Robert Litterman) enabling – mainly in macroeconomics – among other things to test predictive assumptions by running them through historical data.[31]

Like meteorology or seismology, economics serves as a model for optimists when it comes to establishing predictions and forecasts as a gold methodological standard and to developing various applicable models. Obviously, then there is the main question whether the nature and "standard" dynamics of economics or meteorology can be compared to the political-security processes of international relations. The sceptics´ response to this question has always been unsurprising (see below).

[29] TETLOCK, Philip E.; GARDNER, Dan. Superforecasting: The art and science of prediction. Random House, 2016, p. 54.
[30] KAHNEMAN, Daniel; LOVALLO, Dan. Timid choices and bold forecasts: A cognitive perspective on risk taking. Management science, 1993, 39.1: p. 17.
[31] SIMS, Christopher A., et al. Are forecasting models usable for policy analysis?. Quarterly Review, 1986, Win: p. 5.

On the top of that, optimist´s positive stance towards making forecast and predictions in the realm of international relations asserts that forethought plays a key role in decision-making and should, therefore, be routinely exercised through predictions and forecasts. Moreover, predictions and forecasts have already offered decision-making stakeholders several key advantages.

Predictions and forecasts – and especially the probabilistic ones – have usually acted as an essential part in strategic planning. Specifically, these tools offer guidance for the allocation of resources. Without predicting in the strategic domain, then on what basis can policy-makers make decisions about such allocation? There can be a plethora of courses of actions or outcomes in the "not yet" unfolded future, whilst resources for action (to anticipate or to counteract, etc.) are seldom infinite. In this respect, predictions and forecasts help decision-makers to identify truly "realistic" outcomes (e.g. when preparing for something before it happens) or actions (e.g. when choosing which policy to adopt in order to achieve designed goals in the future), and hence, to allocate resources accordingly.[32] Naturally, knowing what is likely to happen is of great interest to all decision-makers. Furthermore, these allocation and scenario-evaluation functions relate to another role of predictions and forecasts as important requisites for early warning.

Schneider, Gleditsch, and Carey show that forecast activity and models can provide – by evaluation of data and monitoring of trends - early warning of conflicts and other disastrous processes (e.g. with the use of forecasting projects such as Political Instability Task Force, etc.). For instance, structural forecasting approaches aim to predict the risk of certain behaviour of geographical units, the time-series design tries to predict particular conflicts, and for example, game-theoretic approaches based on Bruce Bueno de Mesquita´s research examine the possibilities of predicting the future decision-making steps.[33] Furthermore, political events of particular regions can be forecasted through the dynamics of corresponding financial markets.[34] A certain level of anticipation of processes´ outcomes and counteraction to the challenges is therefore possible. Decision-makers can subsequently better understand continuities in the world and new patterns of global relations.[35]

[32] MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: p. 422-423.
[33] SCHNEIDER, Gerald; GLEDITSCH, Nils Petter; CAREY, Sabine C. Exploring the past, anticipating the future: A symposium. International Studies Review, 2010, 12.1: p. 5-7.
[34] Ibid., p. 11.
[35] FEDER, Stanley A. Forecasting For Policy Making In The Post–Cold War Period. Annual Review of Political Science, 2002, 5.1: p. 111-112.

Besides the aforesaid role of planning, early warning and increasing the understanding of international relations, predictions and forecasts serve as a test of scientific understanding. The goal of theory (or hypothesis), as the explanation-prediction debates and the "Popperian critique" debate suggest, is to both explain and predict. And prediction as such is central to the process of science.[36] Milton Friedman, therefore, points out that scientific hypothesis incorporates not only implications but also meaningful assumptions about future phenomena. That said, the comparison of hypothesis´ assumptions (predictions) with experience and reality is the relevant test of the international relations theories (i.e. what is by theory expected to occur versus what actually occurs). Furthermore, it is the test of how well we understand – based on theories and its hypotheses – the social phenomena and dynamics.[37] Michael D. Ward goes on to state that prediction frameworks also incorporates scientific techniques of cross-validation: firstly at the initial level, where the researcher uses the data from the training set to predict outcomes in the test set, and then at the level of collection of new data and subsequent re-evaluation of the prediction model.[38]

1.3 Scepticism: Prediction is a Fool´s Errand

Anticipating the occurrence of international relations´ dynamics and processes is a matter of doubt and conjecture. This is one of the main and resonant statements of prediction and forecast sceptics.[39] The main impediments to predicting and forecasting the future stem from several propositions. Firstly, our cognitive system is not – while craving the elimination of uncertainty – built enough for making predictions and forecasts – not only – in the area of international relations. Similarly to the philosophical debate, the second obstacle concerns the nature of the international system itself. Due to its complexity and frequent non-linearity, the system as such (and its processes) cannot be predicted or foretold by anybody. Beyond that, predictions and forecasts often fail in those cases when they are needed most. [40] There is no need to mention what consequences this stance has for the decision-makers in terms of strategic planning or early warning. In this respect, Erik Gartzke claims that our ability to predict crucial events such as wars "*will probably prove little better than the naive predictions*

[36] SAREWITZ, Daniel; PIELKE JR, Roger. Prediction in science and policy. Technology in Society, 1999, 21.2: p. 122-123.
[37] FRIEDMAN, Milton. The Methodology of Positive Economics. In: HAUSMAN, Daniel M. The Philosophy of Economics: An Anthology. 3rd ed. Cambridge : Cambridge University Press, 2008. p. 149-153.
[38] WARD, Michael D., et al. Learning from the past and stepping into the future: Toward a new generation of conflict prediction. International Studies Review, 2013, 15.4: p. 3.
[39] ROBBINS, Lionel. The Nature and Significance of Economic Science. In: HAUSMAN, Daniel M. The Philosophy of Economics: An Anthology. 3rd ed. Cambridge : Cambridge University Press, 2008, p. 78-79.
[40] KRISTÓF, Tamás. Is it possible to make scientific forecasts in social sciences?. Futures, 2006, 38.5: p. 564.

of random chance".[41] Accordingly, sceptics often warn about the dangers of predictions, forecasts, and their advocates (political advisors).[42]

1.3.1 Epistemological Barriers, Ontological Barriers, and the "Pseudo-scientific" Role

Epistemologically, prediction and forecast sceptics often base their critique on the psychology literature that is replete with studies showing how predictions are undermined by our biases, prejudices, or unfounded optimism and undue conservatism.[43] According to psychologist and economist Daniel Kahneman, our mental universe for thinking consists of two domains: system 1, which is constantly running "somewhere in the background", performs automatic and cognitive operations, and forces us to make conclusions from little evidence; system 2 represents the conscious thought and performs everything we choose to focus on. The problem is that when making predictions and forecasts, we are usually more prone to utilize system 1. We are primed to detect patterns when there is none, and this self-evidently leads to various sorts of mistakes.[44]

Hence, in our world overloaded with information, we sometimes focus on noises (not signals) showing the world as we would like it to be, but not how it really is. We also tend to mistake correlation for causation, which can entail – especially in decision-making – serious consequences.[45] From the perspective of decision-makers, another research of Kahneman and Lovallo proves the existence of two negative tendencies emanating from their cognitive habits. On the one hand, decision-makers incline toward considering problems as unique (despite statistical evidence of the past). But on the other hand, they tend to make and adopt predictions and forecasts from an inside view of the problem. This means that decision-makers oftentimes anchor predictions and forecasts on their own plans of success, rather than on past results, which subsequently causes overly optimistic attitudes.[46] In summary, literature from the field of psychology gives some researchers a good reason to be sceptical about our innate abilities to make predictions and forecasts.

---

[41] GARTZKE, Erik. War is in the Error Term. International Organization, 1999, 53.3: p. 567.
[42] MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: p. 422.
[43] MCNEES, Stephen K. The role of judgment in macroeconomic forecasting accuracy. International Journal of Forecasting, 1990, 6.3: p. 287.
[44] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 33-35.
[45] SILVER, Nate. The signal and the noise: why so many predictions fail--but some don't. Penguin, 2012. p. 23.
[46] KAHNEMAN, Daniel; LOVALLO, Dan. Timid choices and bold forecasts: A cognitive perspective on risk taking. Management science, 1993, 39.1: p. 17.

Ontologically, for sceptics, there is no place for causal thinking in current international relations. Determinism has been replaced by multiplicity and elements of non-linear, interrelated and dynamic processes acting as brick walls of predicting and forecasting.[47] The main difficulty thus lies in the complexity of the system, and this complexity furthermore imposes additional barriers to the foresight.

One of them is the strikingly rare occurrence of events of most importance for prediction and forecasts in the realm of international relations. According to King and Zeng, economic shocks, coups, wars, and even systemic changes, have proven difficult to predict (their occurrence can barely be used to predict their recurrence).[48] But simultaneously, these non-linear events and often immediate changes can be very destabilizing for decision-makers´ planning (i.e. predictions and forecasts about the future). This is what Nassim Nicholas Taleb calls the "Black Swans". Thereupon, many authors describe predicting and forecasting as an umbrella with holes: when not needed in evident and very estimable situations, these methods appear to serve "perfectly", but when decision-makers need them, predicting and forecasting leak badly.[49]

Another barrier stemming from complexity concerns the impact of modest changes and phenomena on international dynamics. Not only wars, coup and shocks, but also small and unpredictable variables may result in far-reaching effects both affecting and triggering larger processes. Such a premise is indeed an expression of Edward Norton Lorenz´s "Butterfly effect" and frequently quoted assumption that butterfly flaps in Brazil can set off a tornado in Texas. In the context of international relations, no one could have predicted that the self-immolation of Mohamed Bouazizi might lead to the protests in Tunisia, Egypt, and other countries. [50] This logic is contained in the Chaos theory. Based on that theory, sceptics consider the international system to be dynamic and self-organizing, and chaos as its essential characteristic. Interestingly, the system in this respect is deterministic (organized around chaos), but unlike for Laplacian determinism (see above), there is no possibility of predicting and forecasting these international dynamics.[51]

[47] DORAN, Charles F. Why forecasts fail: the limits and potential of forecasting in international relations and economics. International Studies Review, 1999, 1.2: p. 15.
[48] KING, Gary; ZENG, Langche. Explaining rare events in international relations. International Organization, 2001, 55.3: p. 693.
[49] DORAN, Charles F. Why forecasts fail: the limits and potential of forecasting in international relations and economics. International Studies Review, 1999, 1.2: p. 15.
[50] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 8-9.
[51] KRISTÓF, Tamás. Is it possible to make scientific forecasts in social sciences?. Futures, 2006, 38.5: p. 566.

The well-documented failures of many experts to predict key events in the area of international relations - such as aforementioned Arab Spring, or even 9/11 and recent global financial crisis - also make sceptics´ nihilism more attractive. Moreover, sceptical view implies (and Taleb even proposes) that if the world is really as unpredictable as they say, then decision-makers can rather focus on already existing situations and events (no planning, early-warning, etc.). Under the ubiquitous uncertainty, refusal of making predictions would, therefore, prevent their risky adventures. Nevertheless, more "brave" sceptics even refer to Wald´s drastic "minimax principle" (minimizing the maximum loss) proposing the policy-maker to make his/her decisions always based on the worst possible outcome (what can be lost), no matter how likely this outcome is.[52]

Apart from stressing the epistemological and ontological barriers, sceptics also sometimes indirectly highlight the threats of prediction and forecast experts. These advisors can form what Peter M. Haas would probably call an epistemic community – a network of "professionals" with recognized competence in the domain of predicting and forecasting. Moreover, these experts are delegated with responsibility toward decision-making.[53] Based on some thoughts of scepticism literature, "alleged" scientific role of this expertise (because the foresight is impossible in international relations) can be misused – under certain conditions - in order to intentionally securitize certain issues, or to manipulate the forecasts for policies already prepared and ready for implementation (see more about this possible negative role of expertise in Rychnovská, 2017[54] or in Berling, Bueger, 2015[55]).

Altogether, whereas the optimists firmly believe in the possibility of anticipating future phenomena, sceptics see such an attitude as a fallacious vision. For optimists, the established practice in other disciplines, as well as its guidance-capable and advise-capable potential, provide clear evidence that forecasting should play a pivotal role in the decision-making, planning, and early warning in the politico-security area (and international relations as such). Sceptics would certainly disagree with that. For them, neither our biological settings nor the complexity of the international system enables us to foresee future dynamics. In that sense, forecasting is like an umbrella with holes in it – appearing to serve well when we do

[52] MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: p. 425-426.
[53] HAAS, Peter M. Introduction: epistemic communities and international policy coordination. International organization, 1992, 46.1: p. 3-4.
[54] RYCHNOVSKÁ, Dagmar. Bio (in) security, scientific expertise, and the politics of post-disarmament in the biological weapons regime. Geoforum, 2017, 84: 378-388.
[55] BERLING, Trine Villumsen; BUEGER, Christian (ed.). Security expertise: practice, power, responsibility. Routledge, 2015.

not need it, but failing when it would be needful.[56] Furthermore, it could serve the experts and decision-makers to manipulate and "control" the future, rather than to anticipate it. In any case, forecasting – as an exercise in futility – should not play any role in the politico-security realm, and its sporadic success is a matter of coincidence.

However, since the unpredictability and predictability coexist, rather than create an above-mentioned dichotomy, optimists and sceptics are not the only participants in the "forecasting-debate".

1.4 The Golden Mean: Philip E. Tetlock and Nate Silver

Sceptics of making predictions and forecasts in international relations are correct when they highlight poor record of predicting important events. They are even right in accentuating both our natural cognitive biases and dynamism or complexity of the international system as serious impediments to our prediction efforts. Nonetheless, sceptics go too far in several ways. Firstly, most of the optimistic authors were never that foolish to expect perfect accuracy in predicting rare and crucial events. In that sense, these authors are rather "optimistic sceptics". In some situations and to some extent, it is possible to see into the future. The key to this lies in the probability (see below). [57] Secondly, it seems that sceptics enjoy overlooking recent promising works capable of foreseeing important events (such as state failures or civil wars) with reasonable accuracy.[58] Lastly, not making predictions, as sceptics do, at all is not an option. Moreover, such refusal would not prevent risky adventures. According to Charles A. Miller, a decision not to act against an actor (e.g. state), because the forecast may not be accurate, can be the riskiest option of all. Similarly, the "minimax principle" – anticipating always only the worst possible future outcome – can be destructive if one thinks about finite resources. Sceptics simply offer no useful guidance for decision-makers. Contrary to what its opponents say, potential benefits of a good prediction and forecast for planning, early warning, readiness or time to counteract, is too great to refuse it entirely.[59]

There is also room for improving predictions and forecasts for international relations sphere. Two books – Philip E. Tetlock´s "*Superforecasting: The art and science of*

[56] DORAN, Charles F. Why forecasts fail: the limits and potential of forecasting in international relations and economics. International Studies Review, 1999, 1.2: p. 21.
[57] MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: p. 420-422.
[58] SILVER, Nate. The signal and the noise: why so many predictions fail--but some don't. Penguin, 2012. p. 14-15.
[59] MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: p. 423-426.

prediction"[60] and Nate Silver´s "*The signal and the noise: Why so many predictions fail - but some don't*"[61] – brilliantly summarize this room for improvement along with the aforementioned premises of optimistic scepticism. Professor Tetlock´s multi-year research program focused on the feasibility of improving probabilistic forecasts, "The Good Judgment Project", won the forecasting tournament (between 2011 and 2014) sponsored by the Intelligence Advanced Research Projects Activity (IARPA). By making – overall – very accurate forecasts on more than 400 questions concerning the world affairs, Tetlock´s project has proven that with the involvement of rigor, successful high-stakes predictions are achievable.[62] Likewise, Nate Silver – statistician and founder of "FiveThirtyEight" website for probabilistic forecasting – correctly predicted the winner of the 2012 United States presidential election in 50 states.[63] And there are others too.

1.4.1 The Evidence of Improvement

There are many hopeful examples of the remarkable progress of prediction and forecast making, and not all of them directly relates to the field of international relations. This is the case of meteorology. For Silver, the weather is the epitome of a dynamic system full of elements of non-linearity. Thus weather fits into the Chaos theory. Nonetheless, the weather forecasting is one of the success stories of our foresight efforts, and thanks to the alliance between the human-based analyses or judgments and computer models, some big weather centres´ forecasts have become 350 % more accurate in the past 25 years.[64] And even though we may not always be able to predict the weather´s Black Swans (e.g. very rare and huge tornado), as well as those in global affairs, success in other meaningful statements is why it is worth continuing in this endeavour.[65] The National Banks also operates in a complex, non-linear environment, and are still able to accurately predict future phenomena. Do these examples signal a mere coincidence? If you think that coincidence stands behind the successes in the forecasting, join – for instance – a stock market, and randomly invest all of your capital. You would "certainly" not become poorer...

---

[60] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016.
[61] SILVER, Nate. The signal and the noise: why so many predictions fail--but some don't. Penguin, 2012.
[62] MELLERS, Barbara, et al. Identifying and cultivating superforecasters as a method of improving probabilistic predictions. Perspectives on Psychological Science, 2015, 10.3: p. 268.
[63] MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: p. 419-420.
[64] SILVER, Nate. The signal and the noise: why so many predictions fail--but some don't. Penguin, 2012. p. 112-113.
[65] MAKRIDAKIS, Spyros; HOGARTH, Robin M.; GABA, Anil. Forecasting and uncertainty in the economic and business world. International Journal of Forecasting, 2009, 25.4: p. 795.

Furthermore, many experts were not blind to the cited Black Swans. Robert J. Schiller or Paul Krugman did see the recent global economic crisis coming.[66] In the case of the Arab Springs, authors examining the Middle East pointed out that the situation in some countries was about to blow. Of course, no one knew – at that time - Mohamed Bouazizi was the trigger. In our world, the uncertainty is pervasive, but still, so much of our reality is predictable. As Tetlock states: "*...it is one thing to recognize the limits on predictability, and quite another to dismiss all predictions as an exercise in futility*".[67] And the potential for improvement in terms of theory-testing activities is almost endless.

In addition to Tetlock´ s and Silver´s achievements, other successes related to the predicting of international affairs do exist. For instance, Duke University has successfully created a series of geographically informed statistical models for conflict prediction. Furthermore, these predictions have been highly accurate.[68] The Peace Research Institute Oslo, too, seeks to improve its work on forecasting. Hegre and his colleagues created a model for predicting civil conflicts up to the year 2050. So far, this model has been respectably accurate.[69] Another successful model using the statistical data comes from the Political Instability Task Force - with more than 80 percent predictive accuracy with a two-year lead time.[70] Without any doubt, in many processes of international relations, the accuracy of our prediction fells of the further we try to forecast (usually when looking more than one year out). However, even the development of month-to-year forecast can be fundamental for decision-making.[71]

1.4.2 Always Think Probabilistically and Evaluate Your Success

According to both Silver and Tetlock, systematic integration of rigor into the prediction and forecast making is the key tool for current, but also future improvement. This rigor consists of three main steps. The first one is to always use numeric probabilities. While meteorology, medicine, and many sub-disciplines systematically use probabilistic forecasts, much of our attempts – except those fruitful mentioned above - in the area of international relations have been marginalizing such necessary element. Under the omnipresent

[66] SILVER, Nate. The signal and the noise: why so many predictions fail--but some don't. Penguin, 2012. p. 25.
[67] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 10.
[68] WARD, Michael D., et al. Learning from the past and stepping into the future: Toward a new generation of conflict prediction. International Studies Review, 2013, 15.4: p. 4.
[69] HEGRE, Håvard, et al. Predicting armed conflict, 2010–2050. International Studies Quarterly, 2013, 57.2: 250-270.
[70] MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: p. 427.
[71] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 5.

uncertainty, it is foolish and sometimes even dangerous to claim something will certainly happen. This applies also to the cases of strategic planning, allocation of resources or early warning. But providing – based on the data – the forecast with an assessment of probabilities to a range of possible outcomes (e.g. there is a 20 % chance of regime change in Iran) represents the most honest expression of the uncertainty in the real world.[72] Additionally, it enables us to retroactively check the accuracy, thus also the quality, of the forecasts. If we are serious about improving, forecasts must contain an assignment of numeric probabilities. This is in fact, along with the natural statement about more than one outcome, the correct scientific definition of a "forecast".[73] Henceforth, the term "forecast" or "prediction" refers only to this scientific probabilistic definition.

If there is no probabilistic assignment, there is also no measurement and subsequent revision, and without revision, the improvement is not possible.[74] The next step in the bringing of rigor is, therefore, a consistent initial revision during the forecast-making. In this respect, it is all about updating. For Silver, updating of forecasts emanates from the Bayesian thinking – from the philosophical underpinnings of Bayes´ theorem. Leaving aside its algebraic form, Bayes´ theorem steers us to think carefully about the probability, uncertainty and also updating (see more in Silver, 2012, p. 230-236). But the general idea behind Bayesian thinking is that we should update (hence revise) our probability estimates during the process of forecast-making. Given the fact that new evidence presents itself to us every day, updating should be done continuously (before releasing forecasts).[75]

The last step is to judge the accuracy of the forecasts. In this case of forecast verification, the forecast is compared against the corresponding observation of what actually occurred. Undoubtedly, accuracy-checking is a critical element for the future improvement of forecasting. Thanks to the forecast verification, we are able to measure the quality (accuracy) of probabilistic forecasts and predictive capabilities of experts (forecasters), institutions, or even respective predictive models.[76] From the methodological perspective, this can be done through the use of scoring rules (see more in the next chapter). Without forecast verification,

[72] SILVER, Nate. The signal and the noise: why so many predictions fail--but some don't. Penguin, 2012. p. 60-61.
[73] CHOUCRI, Nazli. Forecasting in international relations: problems and prospects. International Interactions, 1974, 1.2: p. 63.
[74] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 14.
[75] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 265-273.
[76] GNEITING, Tilmann; RAFTERY, Adrian E. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 2007, 102.477: p. 1-2.

there would be no improvement in meteorology, medicine or economics. The same goes for the area of international relations.[77]

The ideas and researches of Philip E. Tetlock and Nate Silver further explain why this diploma thesis – besides drawing our Czech decision-makers attention to the possible flawed forecasts and thus avoiding bad decisions – seeks to bring the rigor into our (hopefully) emerging forecast activities. Rigor relates to the revision (updating) and verification. As we already know, revision and verification lead to improvement. The improvement of forecasting is even more crucial in the area of political-security processes, and the forecasts of the European Values thus fit into this requirement. Even though this work cannot update forecasts that have already been made (and many events have already occurred), it can still conduct another important step necessary for improvement: forecast verification. The next chapter presents how to do that methodologically.

[77] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 29-30.

2. Methodology

In order to meet Tetlock´s and Silver´s principles and thereby bring the rigor of measurement and verification into the European Values´ - and future - probabilistic forecasts, this diploma thesis uses a scoring rule method as the fundamental and proven instrument for performance (accuracy) analysis evaluation.[78] By assigning numerical values based on the proximity of the forecast to the event when it materializes, it is precisely scoring rules that enable us to test *a posteriori* both how skilful individual forecasters and how accurate resulting forecasts are.[79] This is the evaluative role of scoring rules. Additionally, in terms of elicitation, scoring rules also encourage assessors to take the ubiquitous uncertainty into account and, therefore, to be honest, and make careful assessments. Amongst a plethora of existing scoring rules, this work applies the one both widely used in fields ranging from meteorology to international relations and particularly well-suited for verification of European Values´ probabilistic forecasts of categorical and binary variables: The Brier score.[80]

Developed by Glenn W. Brier in 1950 the Brier score measures - like other scoring rules - the distance between what we forecast and what actually happened (reality).[81] Mathematically, the broadly applied (popular) and binary-related Brier score is the sum of the squared differences (deviations) between event occurrence and probabilistic forecast,

$$B = \frac{1}{N}\sum_{j=1}^{N}\left(p_j - o_j\right)^2$$

where p_j is the forecasted probability of the occurrence of the event – in the range of 0 (0%) and 1 (100%), o_j the actual outcome of the event at instance j (i.e. o_j is a value equal to 1 if the event occurred, and 0 if the event did not occur), and N is the number of forecasting instances.[82] In effect, lower Brier score means a higher level of accuracy. Whereas Brier score of 0 represents a "perfection" (i.e. $p_j = o_j$), a Brier score of 1 tells us that our forecast was a perfect opposite of reality. Let us consider a binary variable/event of interest along with the corresponding question, say: "*Will Kim Jong Un and Moon Jae-in sign a peace treaty – ending officially the Korean War - before 31 November 2019*?" Suppose a forecaster who

[78] BRÖCKER, Jochen; SMITH, Leonard A. Scoring probabilistic forecasts: The importance of being proper. Weather and Forecasting, 2007, 22.2: p. 382.
[79] JOLLIFFE, Ian T.; STEPHENSON, David B. (ed.). Forecast verification: a practitioner's guide in atmospheric science. John Wiley & Sons, 2012. p. 1.
[80] MELLERS, Barbara, et al. The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. Journal of experimental psychology: applied, 2015, 21.1: p. 2.
[81] BRIER, Glenn W. Verification of forecasts expressed in terms of probability. Monthey Weather Review, 1950, 78.1: p. 1-3.
[82] CANDILLE, Guillem; TALAGRAND, Olivier. Evaluation of probabilistic prediction systems for a scalar variable. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 2005, 131.609: p. 2133.

predicts 90 percent chance that both statesmen sign such a treaty before 31 November 2019. If they really sign the peace treaty (reality coded as 1), then the forecaster´s Brier score for this estimate would be impressive $(0.9 - 1)^2 = 0.01$, a number not too far away from the perfection and thus very close to the reality. Otherwise, if both statesmen do not sign the peace treaty (reality coded as 0), the forecaster would get a disastrous Brier score of $(0.9 - 0)^2 = 0.81$.[83]

Nevertheless, the performance (accuracy) analysis of the European Values´ skills is based on the original Brier score (not the popular), too, verifying probabilistic forecasts of dichotomous and binary events. The logic behind the original Brier score is identical to the more "popular" one introduced above. It differs only in its inclusion of squared differences for the complementary ("non-") event.[84] Essentially, when assigning a probability that not-yet-unfolded future event will happen, we are also simultaneously and indirectly passing judgment on the complementary – that means the opposite – situation (i.e. the future event will not happen). Hence, by the inclusion of such a complementary judgment, this original Brier score equation takes values between 0 (best) and 2 (worst).[85] Using the same example of the peace treaty between North and South Korea, a forecaster who predicts 90 percent probability that both countries will sign a peace treaty before 31 November 2019 also – at the same time – implicitly says there is a 10 percent chance it will not be signed. So if they do sign the treaty, it will produce a Brier score of $(0.9 - 1)^2 + (0.1 - 0)^2 = 0.02$, which means almost hitting the bull´s eye. Conversely, if aforesaid statesmen do not sign it before this date, the forecaster would finish with a Brier score of $(0.9 - 0)^2 + (0.1 - 1)^2 = 1.62$, that is, getting closer to be "as far from the truth as it is possible to get".[86] Importantly, the original Brier score enables us not only to better see the European Values´ and alternative players´ accuracy (due to the score range from 0 to 2) but to compare these "yes/no" questions of binary and dichotomous events with three-possible-outcomes questions properly. For a question with three possible outcomes, forecasters always assign a probability to each outcome. These questions typically concern likelihoods of the post-election coalition forming.

Suppose you forecasted chances of three possible coalition formations as follows: coalition A = 70%, coalition B = 20% and coalition C = 10%. If "A" outcome occurred, your Brier score – by the same logic – would be $(0.7 - 1)^2 + (0.2 - 0)^2 + (0.1 - 0)^2 = 0.14$ (very close to the perfection). On the contrary, occurrence of "B" coalition would mean a disappointing

[83] FRIEDMAN, Jeffrey A., et al. The value of precision in probability assessment: Evidence from a large-scale geopolitical forecasting tournament. International Studies Quarterly, 2018, 62.2: p. 414-415.
[84] WILKS, Daniel S. Statistical methods in the atmospheric sciences. Academic press, second edition, 2006. p. 284.
[85] CHANG, Welton, et al. Accountability and adaptive performance under uncertainty: A long-term view. Judgment & Decision Making, 2017, 12.6. p. 617.
[86] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 64.

Brier score of $(0.7 - 0)^2 + (0.2 - 1)^2 + (0.1 - 0)^2 = 1.14$, while if "C" outcome happened, your Brier score would be terribly awry: $(0.7 - 0)^2 + (0.2 - 0)^2 + (0.1 - 1)^2 = 1.34$.[87]

The major advantage of using the Brier score for the verification of European Values´ forecasts is that it is classified as a "strictly proper scoring rule", which means, basically, it incentivizes forecasters to report their true beliefs and to eschew both false positive and false negative judgments. In this sense, the Brier score will always penalize you (your score) for bending your beliefs to political pressure.[88] However, based on the decomposition made by Murphy and Winkler, the Brier score tells us more than just how accurate our forecast is. The Brier score is comprised of three components - variability, calibration and resolution – and some of them give us more information about forecaster´s abilities.[89] Whereas variability has nothing to do with the skill as such (it concerns events and its environments) and is therefore not discussed further, both calibration and resolution certainly have. In our case, especially the resolution is something very useful for more in-depth performance analysis of the European Values´ forecasts and mainly of its invited experts (see below).

Calibration refers to an ability to make forecasts coinciding, in the long run, with the outcomes of the events. If you predict for example the weather, and rain happens 80% of the time when you say there is an 80% chance of rain, your forecast is perfectly calibrated. When things you say are 10% likely actually happen 60% of the time, you are under-confident. In contrast, over-confidence comes if things you say are 70% likely actually happen only 30% of the time. The last component of Brier score, the resolution, regards our ability to be decisive in forecast-making (hence to distinguish signals from the noise).[90] Assigning high probabilities to things that really happen and low probabilities to things that do not, signifies superb resolution. When rain happens 40% of the time when you say there is a 40% chance of rain, you are perfectly calibrated, but also very cautious and poorly decisive. Forecasters are said to have a poor resolution if they assign probabilities around the "maybe" zone (50% chance that something will happen) between 40% and 60%. Thus, the outcome that happens 80% of the time you say there is an 80% chance of its occurrence indicates not only a perfect calibration of your forecast but also its excellent resolution.[91]

---

[87] Frequently Asked Questions. In: Good Judgment Open [online]. [cit. 2019-04-15]. Available from: https://www.gjopen.com/faq

[88] BO, Yuanchao Emily, et al. An IRT forecasting model: linking proper scoring rules to item response theory. Judgment & Decision Making, 2017, 12.2. p. 90.

[89] MURPHY, Allan H.; WINKLER, Robert L. A general framework for forecast verification. Monthly weather review, 1987, 115.7: p. 1330-1338.

[90] MELLERS, Barbara, et al. Identifying and cultivating superforecasters as a method of improving probabilistic predictions. Perspectives on Psychological Science, 2015, 10.3: p. 270.

[91] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 62.

Although the "original" Brier score represents a pivotal tool for performance (accuracy) analysis of the European Values´ probabilistic forecasts, one of its components – resolution – helps us to explore other aspects of forecasting abilities and qualities of both their final forecasts and invited experts (see the introduction of the European Values´ research project in the next chapter).

Having said the above, we can now finally proceed to the nature of our analysis. Given the fact that the European Values did not assign all the probabilities to the future development of the politico-security environment based on their own considerations, but for some questions invited the experts from the field of International relations and aggregated their estimates (again, see the next chapter), the verification process extends several levels of analysis.

At the initial - internal - level, the analysis not only examines the invited experts but also scrutinizes some elements of the think-tank´s research project. Regarding the experts-analysis, this is done firstly by the use of elemental descriptive statistics (e.g. mean, median or standard deviation). More importantly, the internal analysis subsequently inquires some aspects of the probabilistic behaviour related to one of the aforementioned components of the original Brier score – the resolution. Calibration works well for those types of events enabling us to make new forecasts every day. But unfortunately, it works less well for rare events emanating from the politico-security realm (e.g. elections or the EU disintegration). Thus, only the resolution (i.e. how decisive a forecaster is in her/his forecasting) can be analysed without any limitation. Furthermore, resolution, along with the so-called "anti-probabilistic" thinking, is a parameter signalling the potential inexperience of invited experts as well as of the European Values. Answering the question of how resolute are both the invited experts and the think-tank´s research team, and whether they show some signs of inexperience, is very important for the final recommendations and for the future forecast-making in the Czech politico-security area.

The second level of the following analysis ascertains real results of the European Values´ forecasts. This is the key part of the verification process, showing us how accurate the European Values are. Since their forecasts are related to the 2016-2019 period and few events still wait for its denouement, the accuracy assessment - the original Brier score for each event (or process) and the final average Brier score – is performed for the two and a half year horizon (from the end of 2016 to the June 2019). However, does the possible low and disastrous Brier score always indicate poor forecasting abilities or exceptional alarmism? Predicting wrongly a 90 percent probability that something will happen (and reality is actually

the opposite) and getting a Brier score of 1.62 would seem to be, without a doubt, very disappointing. But now imagine that other institutions or forecasters would also fail to predict an outcome of a rare event and would, too, finish with a terrible Brier score. If this is the case, it is harder to criticize forecasting abilities (accuracy) on the basis of dynamics that "no one" was able to anticipate and correctly predict. Accordingly, it is appropriate to compare probabilities and corresponding Brier scores of interest with the accuracy of other institutions or forecasters.[92] This step helps us to decide whether the prospective poor accuracy emanates from the forecasting deficiencies, or, rather, results from dynamics that are – in general – difficult to estimate.

On the basis of the above, the third and last level of analysis brings the comparison into the verification process. Moreover, the third level of analysis not only serves as a tool for checking the accuracy (Brier scores) of the European Values´ forecasts but also shows us the cases in which the European Values´ forecasting abilities were better/worse than the alternative ones. The average Brier score of the European Values – based on selected questions (assigned probabilities) – is compared with the alternative judgments (and an average Brier score) of foreign institutions. Naturally, these questions under comparison are – in most instances – identical for both the European Values and foreign institutions. In cases where identical questions of other actors were not found, very similar and comparable questions to those of the European Values are used. Nevertheless, ideal proper accuracy comparison with other institutions is sometimes impeded by lacking data. Typically, not so many institutions have common interests in estimating the probability of the same events or processes (see more in Chapter 6). As a consequence, the third level of analysis also utilizes other two suitable (and widely used) means of comparison. One of them is an accuracy comparison between the European Values´ research team and the experts invited (and used minimally) by the think-tank. The second comparison utilizes the simulation of randomly generated estimates. Therefore, the accuracy comparison is, similarly to the whole analysis, performed at three levels: 1) The European Values vs. The invited experts, 2) The European Values vs. The Alternative Institution; and 3) The European Values vs. Randomly Generated Estimates.

Aforesaid three-level encompassing analysis makes the verification process complete and should tell us about both good and bad elements arising from the first attempt to make politico-security probabilistic forecasts in the Czech Republic. Such an analysis also

[92] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 65.

represents an essential and necessary foundation for future efforts (i.e. what should be further improved in our future endeavour). Following chapter finally introduces the European Values´ research project.

3. Introduction to the European Values´ Research Project

The European Values Think-Tank has developed probabilistic forecasts to achieve several pre-defined ambitious goals. These can be described as follows: 1) to create an indicative framework for understanding crucial politico-security phenomena which will fundamentally affect the Czech Republic between 2016 and 2019; thus also 2) to provide the Czech policy-making stakeholders with such an analysis of possible development of important trends; and subsequently 3) to formulate recommendations for what decisions the Czech politico-security elite should accept to protect the vital interests of our country (defined by the Security Strategy of the Czech Republic).[93]

Therefore, with the participation of 24 experts from the Czech intelligence and security community, the think-tank´s research project answers the questions dealing with factors which "*will have the main impact on the European politico-security environment from the perspective of the public interests of the Czech Republic between 2016 and 2019*".[94] In other words, the European Values have attempted to forecast the probability of occurrence of the global, European, and sometimes also "Czech" events or processes capable of affecting our security environment as well as the vital strategic interests of the Czech Republic. The think-tank addresses these possible "futures" by two slightly different approaches – firstly, by the introduction of "four disastrous scenarios", and secondly, by the presentation of "partial findings" (all individual probabilistic forecasts) which provided the basis for disastrous scenarios. Whereas those four scenarios are not part of our inquiry, the partial findings are of great importance for us and represent one of the two analytical cornerstones (datasets) of this diploma thesis.

Regarding the disastrous scenarios, the European Values have created four forecasts which would act – in case of coming true – as "game-changers" and would radically and negatively affect both the European and the Czech politico-security system. These are: 1) *President Marine Le Pen Gives the Final Blow to the European Integration (40% probability)*; 2) *Russian Aggression Will Destroy the Legitimacy of NATO and Cause Appeasement in the Czech Republic (30% probability)*; 3) *Uncontrolled Mass Migration Will Cause Political Implosion of Some European States (60% probability)*; and 4) *Western European Democratic Leaders Will Lose the Political Struggle with Islamism (60%*

[93] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 8 and p. 27. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

[94] Ibid., p. 1.

probability).[95] The scenarios have been built more as creative stories (delineating many possible catastrophic effects on several actors) combining the results of partial findings (based also on invited-experts´ estimates), and imaginativeness of the think-tank´s research team. Given the fact that the think-tank assigned relatively high probabilities to the occurrence of such disastrous scenarios, and none of these forecasts happened (or will probably not happen until the end of 2019), they seem to be very alarmist. Nevertheless, these "four scenarios" have been already methodologically analysed and also criticized in Miroslav Kalous´ article "*HOW (NOT) TO PREDICT THE FUTURE?*"[96] - which means there is no need to pay more attention to them.

On the other hand, the think-tank´s partial findings truly represent both the serious attempt and a significant breakthrough in utilizing the probabilistic forecasting in the Czech politico-security area. The logical consequence of such an analytical division is that the term "European Values´/think-tank´s research project" refers, in this work, only to the "partial findings" (all individual probabilistic forecasts). Moreover, as mentioned above, these probabilistic forecasts represent the main dataset for bringing the accuracy assessment – and hence the rigor – into the Czech probabilistic forecasting.

The partial findings consist of 58 forecasted questions and 79 resulting estimates/assigned probabilities (when including not only the binary forecasts with one stated probability but also three-option forecasts where all probabilities have to be stated)*. Compared to the foreign projects (see for example Tetlock´s Good Judgment Open[97]), the European Values´ forecasted questions do not really take the form of "questions" (e.g. what is the probability that something will happen?) but rather the form of statements or judgments (e.g. something will happen with a particular probability). However, the meaning remains the same here. These 58 questions (statements) regard seven "key players" of the global and European politics whose internal development would have a major impact on the Czech political scene during 2016-2019, and seven "systemic threats" capable of jeopardizing the Czech Republic. Among the key players, Germany, France, Russia, the United States, and

[95] Ibid., p. 2-7.

[96] KALOUS, Miroslav. HOW (NOT) TO PREDICT THE FUTURE?: Analysis of several pioneering studies in the field of Czech political and security scenario-building. In: Obrana a strategie [online]. [cit. 2019-07-10]. Available from: https://www.obranaastrategie.cz/en/archive/volume-2018/1-2018/articles/how-not-to-predict-the-future.html

* In forecasts concerning the binary event, there is one stated probability (e.g. 70% probability that the US will rejoin the TTP negotiations). The unstated probability is only a logical completion (30% probability that the US will not rejoin the TTP negotiations) and this 30% estimate is not included in the overall number of estimates. Conversely, in three-option questions (e.g. A: The US will negotiate the same TTP deal, B: The US will negotiate different deal, C: The US will refuse to negotiate), all three probabilities are included in the overall number of estimates, because two out of three estimates cannot be logically completed without knowing the exact probability (e.g. we know that A is 60%, however the B and the C can be both 20% and 20%, but also 30% and 10% etc.).

[97] Good Judgment Open [online]. [cit. 2019-07-10]. Available from: https://www.gjopen.com/

Poland epitomize state actors; the European Commission and the Islamic State represent supranational and non-state players, respectively. Additionally, those seven systemic threats selected by the European Values are: Islamic extremism in Europe, Mass migration, Authoritarianism, Extremist attitudes in society, Economic crisis, Cyber threats, and the Energy threats. The majority of the 58 questions and 79 estimates can be found in the main text of the research project. On top of that, some of these questions and estimates are unveiled only in the final table (primarily illustrating a probable year of occurrence of selected events).

The estimates for both "partial findings (58 forecasted questions)" and "four disastrous scenarios" have been gained through the Delphi method. This research design, developed in the 50s and 60s by RAND, seeks to obtain the "group (experts) response" by the use of repeated questionnaires and subsequent feedbacks. To be more specific, the purpose of the Delphi method is to arrive at something closer to the expert consensus, while preventing well-known problems associated with the "group thinking". Thus, the method utilizes a long-distance questionnaires survey which provides feedback, but also leave the experts anonymous to each other (the experts only know the stance of "unknown others" in a particular question).[98] Based on these principles, the experts - invited by the European Values – participated on two key issues: on the selection of the most important factors and corresponding dynamics, and on the evaluation of the probability of occurrence of these dynamics. The total number of invited experts who really participated was 24. The following brief overview of the European Values´ steps and stages further elucidates how the think-tank got to the individual estimates of invited experts.[99]

- Stage 1 – The European Values´ research team created the "list of Czech interests".
- Stage 2 (the first round of questionnaires) – The think-tank contacted the experts and asked them to define three to five factors in each of the Czech interest.
- Stage 3 – The European Values´ research team evaluated all responses (from those who decided to participate), selected 10 most severe factors, and created the "list of the most important factors".

[98] RAND: Delphi Method. In: RAND Corporation [online]. [cit. 2019-07-11]. Available from: https://www.rand.org/topics/delphi-method.html

[99] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 8 and p. 30-32. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

- Stage 4 – The European Values´ research team added five alternatives of possible developments to each factor, and thus created the "list of the most important factors with development alternatives".
- Stage 5 (the second round of questionnaires) – the experts received the list of the most important factors with development alternatives", and were asked to assess (by a probability) what development alternatives "for each of the factors are most likely to happen".
- Stage 6 – The European Values´ research team evaluated these responses, and formulated about three to five "possible development forecasts".
- Stage 7 (the third round of questionnaires) – the experts received the list of "possible development forecasts", and were asked to evaluate the probability of their fulfilment.
- Stage 8 – The European Values´ research team finalized the forecasts according to the probability of their fulfilment and the severity of their impact on the Czech interests (in case of their fulfilment).

The research project was conducted between April and September 2016 and published in October 2016. Interestingly, in Stage 5, the European Values asked the experts to judge "*what development alternatives before April 2019 for each of the factors are most likely to happen?*"[100] This would imply that the forecasted period of the research project was from the end of 2016 until April 2019. But elsewhere in the text, the think-tank frequently mentions that the time period is from 2016 to 2019 (the whole year 2019 is included). Hence, it is reasonable to assume that the European Values´ really operated with the period from the end of 2016 to the end of 2019.

Nevertheless, based on the think-tank´s research design chapter, the extent to which the invited experts´ probabilities constitute the partial findings is unclear. Elsewhere in the text, the European Values briefly mentions that the percentage estimates are an average of the estimates of all the experts.[101] Does it mean that all 58 questions (statements) along with 79 estimates really emanate from the probabilities assigned by 24 invited experts? In order to find this out, the author of this diploma thesis asked Doctor Radko Hokovský (who is an Executive Board Chairman of the European Values and the head of their Internal Security Program) for the questionnaire-data (in an anonymized form). Mr. Hokovský responded

[100] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 31. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf
[101] Ibid., p. 8.

positively and sent the evaluated questionnaire containing the following data: all the forecasted questions (related to those development forecasts) sent by the European Values which invited experts responded to, and all the estimates assigned (by invited experts) to the probability of fulfilment of forecasted questions. On the whole, this accounts for 51 questions and 530 individually assigned probabilities/estimates (see Appendix B, Sheet 1).

Moreover, questionnaire-data reveals two interesting findings. Firstly, there are strong indications suggesting that the European Values did not employ an "average" as the method for converting many individual experts´ estimates into the one final probability (see their claim above). Short analysis (see Appendix B, Sheet 2) shows that the European Values probably utilized "median" value of individual estimates rather than the "average" value. This is not a big problem however, as the use of "median" would also be in line with the Delphi method. In the research design chapter, the think-tank refers to the Olaf Helmer-Hirschberg´s article about the Delphi method, which recommends "median" as the only method of conversion.[102] Secondly, the European Values used, in the end, only a minimum of invited-experts´ probabilities in their research project. From the total number of 51 questions forecasted by the experts in the questionnaire, the think-tank directly or indirectly used only 11 of them. Yet in two cases, the think-tank´s research team evidently modified 3 out of 11 forecasted questions used in their research project (see Appendix B). Whereas one question forecasted in the questionnaire was split in the research project into two questions, the other two questions forecasted in the questionnaire were combined into one question which occurs in the research project. All in all, the final number of experts´ forecasted questions (based on the questionnaire and used in the think-tank´ research project) is still only 11. It also implies that the think-tank´s research team somehow internally created 47 questions (statements) and its estimates for its research project, without taking any of the questionnaire data (both questions and probabilities) into consideration.

Nonetheless, the data from the questionnaire opens up a great opportunity not only to analyse the estimates assigned by the experts, or their behaviour from the perspective of probabilistic forecasting, but also to check whether or not there are any differences (in the accuracy or behaviour) between the research project´s estimates and experts´ estimates. Thus, the questionnaire data concerning the invited experts represents the second analytical cornerstone (dataset) of this diploma thesis. Although the think-tank divided the questionnaire

[102] HELMER, Olaf. Analysis of the future: The Delphi method. RAND CORP SANTA MONICA CA, 1967. p. 8.

according to ten topics, this diploma thesis comes up with its own division which allows analysing the data in a more interesting and productive way.

Hence, the data about the invited experts consists of following six blocks of forecasted questions (see both Table 1 and Appendix B): 1A) questions used in the European Values´ research project, 1B) non-used questions (questions which are not used by the think-tank in their research project), 2A) questions regarding the Czech Republic, 2B) questions concerning the world (dealing with European or global events/processes but not directly with the Czech Republic), 3A) questions about the one-date events (concerning events which are related to a specific date or period, or – as the post-Brexit dynamics – processes emanating from a specific events framed by time), and finally 3B) questions with loosely-time-framed processes (regarding the dynamics which could occur at any time between 2016 and 2019). Naturally, these blocks are not fully mutually exclusive. A specific question can be both "used" in the research project and – for instance – simultaneously regard the world. Conversely, one question cannot be – for example – "used" and "non-used" at the same time.

Table 1: The overview of each block of questions (based on the questionnaire data - 51 questions and 530 estimates)

Blocks of questions	Number of questions in each block	Number of individual estimates
1A) Used questions	11 questions	111 estimates
1B) Non-used questions	40 questions	419 estimates
2A) Questions regarding the Czech republic	34 questions	336 estimates
2B) Questions concerning the world	17 questions	194 estimates
3A) One-date events	4 questions	48 estimates
3B) Loosely-time-framed processes	47 questions	482 estimates

source: own elaboration (based on the data in Appendix B)

On the basis of the above introduction, the European Values´ research project now refers only to the "partial findings" (58 forecasted questions) which can be found in the think-tank´s article. Leaving aside alarmist and very imaginative "four disastrous scenarios", the partial findings represent the main dataset and the essential foundation for the accuracy verification of the European Values. Nevertheless, the success in obtaining the anonymized data from the questionnaire opens other analytical opportunities. The questionnaire data, together with the fact that the think-tank actually used only a minimum of estimates of invited experts, enable us to explore much more "than" just the predictive capability of the European Values. The analysis based on elementary statistics, indicators signalling possible

inexperience of invited experts, or questions which were raised, would give us valuable information about the experts as well as the European Values. Furthermore, the questionnaire data of invited experts can serve as a comparative tool in assessing the think-tank´s accuracy or potential signs of inexperience (whether or not there are any differences between the think-tank´s research team and invited experts). In any case, the questionnaire data, therefore, represent the second cornerstone for the analysis of the European Values´ research project.

Accordingly, before proceeding to the evaluation of predictive capabilities (accuracy) of the European Values, the following chapter carries out an analysis of both invited experts (questionnaire data) and the think-tank´s research project.

4. The First Level of Analysis

The anonymized form of the questionnaire data precludes the accuracy analysis of individual invited experts. It is simply not possible to find out whether or not the order (position) of each forecaster remains the same throughout the whole questionnaire. However, the data still enables us to scrutinize some other aspects related to the probabilistic forecasting, such as the so-called "anti-probabilistic" thinking or poor-decisiveness. Along with the elementary statistical overview, this is the purpose of the first subchapter.

The second part of the chapter subsequently analyses the main dataset of this diploma thesis – the partial findings of the European Values´ research project. Besides examining the potential incidence of "anti-probabilistic" thinking or poor-decisiveness, the subchapter also presents both positive and negative aspects of the think-tank´s research project.

4.1. Internal Analysis of the Invited Forecasters

The European Values presumably used - in its research project – "only" one fifth (11 out of 51) of the questions asked to the forecasters. Even then, the statistical analysis of more than 500 forecasters´ probabilities assigned to 51 questions (mostly statements) not only offers the opportunity to capture valuable data and information about 24 forecasters and European Values´ project but also – again – provides a hint of the necessary final recommendations. Besides, this can be done without necessarily calculating any of 530 Brier scores. The particular data are shown in Appendix B (Sheet 1 and Sheet 2).

Based on the measures of elementary statistics, the mean of all probabilities (530) assigned by forecasters to the total amount of 51 questions is 0.43, and the median value is 0.4. Eleven questions which have not been finally used in the European Values´ project show lower given average value (probability) than those 40 questions which have occurred in their research project (the mean assigned value of 0.39 from 419 probabilities compared to the 111 probabilities with an average value of 0.59). From the perspective of the different blocks of questions (see these blocks in the previous chapter), forecasters tended to to give - in average - higher probabilities to 17 questions regarding the European or world dynamics than those 34 statements concerning the Czech Republic (the mean assigned value of 0.53 compared to the mean assigned value of 0.37). In terms of the differences in the time framing of events, the mean for both events that should or should not occur in one date (along with similarly raised questions), and processes or dynamics that could or could not happen at any time between

2016 and 2019, is nearly identical (former events with the mean of 0.44, latter ones with the mean assigned value of 0.43). However, the mean has only limited informative capability and simply tells us that in our population of forecasters and across the observed blocks of questions, the average assigned probability is relatively close to 50 percent.

In contrast, standard deviation (SD), as a measure of variability quantifying the amount of dispersion of our values, has greater validity.[103] The standard deviation and its particular fluctuations essentially reveal whether or not and to what degree were forecasters consensual (not accurate) in assigning their probabilities to individual questions or blocks of questions. For the whole dataset (with the mean of 0.43), the standard deviation is 0.28. This value fluctuates only modestly when comparing questions that have been used in the research project (SD = 0.26) with those that have been not (0.27), statements regarding the Czech Republic (SD = 0.26) with those concerning the rest of the world (SD = 0.30), and questions about one-date events (SD = 0.28) with loosely-time-framed processes (SD = 0.28). But some patterns of identifiable consensus can be found within the certain blocks of questions, thus indicating that something in the nature of these blocks could lead to experts´ estimates convergence. Interestingly, three questions with the lowest standard deviation (i.e. lowest spread of assigned probabilities around the mean) pertain to the statements about the future decisions of the Czech Republic and behaviour of its decision-makers (statement number 2.4 "*The Czech Republic will leave the EU*" with SD = 0.11; statement number 10.3 "*Representatives of the executive and legislative power will bring positive aspects of patriotism...into the public debate*" with SD = 0.11; statement number 4.2 "*The Czech Republic will systematically deter hoaxes...*" with SD = 0.12). Substantially low standard deviation can also be found in questions relating to considerable time-framing. Accordingly, invited experts have reached "virtual consensus" (assigning similar probabilities) in estimating one-date-statements 1.3, 3.1 and 4.5 about Brexit dynamics (SD = 0.08*), 2016 United States presidential election result (SD = 0.18) and foreign policy direction of the "next" winner of the 2018 Czech presidential election (SD = 0.16**). It should be mentioned, though, that the context and usually average assigned probability (mean) are extremely important here.

[103] MEIER, Kenneth J.; BRUDNEY, Jeffrey L.; BOHTE, John. Applied statistics for public and nonprofit administration. Cengage Learning, 2011. p. 98-101.

* Here, the estimate of one forecaster – who assigned entirely opposite value probably by mistake – is not taken into account. Still, the "consensus" would be very high even if we include his/her estimate.

** Here, the estimate of one forecaster – who complains of the vagueness of questions in this block (and gives many "zeros" – is not taken into account. Still, the "consensus" would be high even if we include his/her estimate.

Regarding statements with the overall lowest standard deviation (greatest consensus) pertaining to the future of the Czech Republic, invited experts assigned – in average – 10% probability that the Czech Republic would leave the European Union by 2019, 30% probability that our representatives would bring positive aspects of patriotism into a public debate along with defending of freedom, and 30% probability that the Czech Republic would systematically deter hoaxes and pragmatically solve real problems associated with extremism.

However, here, the "concordance" among experts (and even high decisiveness related to the resolution) might be more about concrete phenomena than patterns or aspects belonging to the block of questions about the Czech Republic. More than one-third of the statements related to the Czech Republic, by contrast, show a standard deviation higher than 0.26 (SD of the whole block concerning the Czech Republic). Therefore, although the support for the European Union was significantly low in the Czech Republic in 2016, experts "probably" identically reflected long-standing attitudes within Czech society, persisting rejection of departure from the EU[104], or support for staying in the EU expressed by top decision-making stakeholders.[105] Due to a number of fundamental obstacles (e.g. social media algorithms, "online" nature of hoaxes, the use of hoaxes by parliamentary parties and foreign actors, or top-political-representatives criticism of the "Centre Against Terrorism and Hybrid Threats"), experts perhaps unsurprisingly shared considerable scepticism about the feasibility of deterring hoaxes in the Czech Republic.[106] Analogously, they might not believe that the transformation of the Czech party system (since 2010) and the growing exploitation of aspects of nation and migration (especially during the peak of "Migration crisis") for political populism would entail both spreading of positive patriotism and defending of justice and honour.[107]

Just as in the previous case of "Czech Republic questions", the story behind, and specificity of the particular event, rather than the time framing itself, should be considered as the cause of a very low standard deviation of three one-date-event statements (out of a total of 4 statements in this block). Thus, the consensus among experts in assigning very similar

[104] Brexit a postoje k případnému Czexitu, 13. 6. 2016 – 14. 6. 2016: Výzkum veřejného mínění pro Český rozhlas. In: Median [online]. [cit. 2019-05-20]. Available from: http://www.median.eu/cs/wp-content/uploads/2016/06/1516420_Bleskovy_pruzkum_cerven_2016_N_1049_zprava_v05_BREXIT.pdf

[105] Zeman je pro referendum o českém odchodu z EU a NATO, hlasoval by ale proti. In: Český rozhlas: iRozhlas [online]. [cit. 2019-05-20]. Available from: https://www.irozhlas.cz/zpravy-domov/zeman-je-pro-referendum-o-ceskem-odchodu-z-eu-a-nato-hlasoval-by-ale-proti_201606302005_mkopp

[106] Česko je rejdištěm dezinterpretací a dezinformací, říká rektor. O unii chce šířit objektivní informace: interview with rector Jaroslav Miller (Palacký University Olomouc). In: iRozhlas: Interview Plus [online]. 5. 4. 2019 [cit. 2019-05-21]. Available from: https://www.irozhlas.cz/zpravy-domov/euforka-univerzita-palackeho-olomouc-fake-news-dezinformace-evropska-unie_1904051740_och

[107] NAXERA, Petr; KRČÁL, Vladimír. Populistická konstrukce národa ohroţeného migrací: CAQDAS volebního diskurzu českých parlamentních voleb v roce 2017. Sociológia, 2018, 50.5: p. 492-493.

probabilities – again – "probably" lies in the individual aspects of each event. The average assigned probability that the United Kingdom would not follow the referendum results and would not leave the EU was 10% (20% if we include probability assigned by the forecaster who chose completely different estimation; see the footnotes). The accordance (low SD) likely emanates from the – notably at that time – widely accepted belief and an often mentioned reaction that the United Kingdom´s impressive democratic traditions (and respect for them) preclude undermining of country´s democratic processes.[108] In the case of giving the estimates about the 2016 US presidential election winner, the consensus, in fact, was not so convincing, and three out of twelve forecasters did not "bet on" Hillary Clinton´s victory. Nevertheless, the tendency of the rest of experts to assign higher probabilities to Hillary Clinton´s victory led to the average 60% probability (and still to a significantly low SD) of her winning the 2016 US presidential election. Even though Donald Trump edged ahead of Hillary Clinton in the presidential race several days after the European Values´ expert assessment[109], such experts´ tendencies in summer 2016 (a belief that Clinton will likely win) again corresponded with the predominant expectation that she "will" be the next president of the United States.[110] In the last assessed statement characterised by low standard deviation and hence significant consensus, experts – in average – assigned 50% probability (45% if we include probability assigned by the forecaster who questioned the whole block of questions and gave 0% to three of them) that a pro-Western candidate would win the 2018 Czech presidential election. Some experts assigned a slightly higher probability (than 50%) that pro-Western candidate would win; some assigned a slightly lower probability (than 50%) to the same outcome. Consider, by analogy, flipping a coin. This is precisely what the average assigned probability of 50% expresses. Once again, the consensus among experts – that the result could go both ways – might be caused by the specificity of this event and experts´ reflection of the general mood of the Czech society. Let's leave aside the penetration of the notion of "divided society" into the political and public discourse (mainly through media landscape)[111] and fluctuation of the trust (2013-2016) in the then-incumbent president (who

[108] STONE, Jon. Brexit result must be respected, David Cameron says. In: The Independent [online]. [cit. 2019-05-21].Available from: https://www.independent.co.uk/news/uk/politics/brexit-result-must-be-respected-david-cameron-says-a7105886.html

[109] KAHN, Chris. Trump edges ahead of Clinton in U.S. presidential race: Reuters/Ipsos poll. In: Reuters [online]. [cit. 2019-05-21]. Available from: https://www.reuters.com/article/us-usa-election-poll-idUSKCN1062MC

[110] KATZ, Josh. Who Will Be President?: Hillary Clinton has an 85% chance to win. In: The New York Times [online]. [cit. 2019-05-21]. Available from: https://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html

[111] Miloš Zeman podle premiéra rozděluje společnost a připravuje se na příští kampaň. In: Česká televize: ČT 24 [online]. [cit. 2019-05-22]. Available from: https://ct24.ceskatelevize.cz/domaci/1678477-milos-zeman-podle-premiera-rozdeluje-spolecnost-a-pripravuje-se-na-pristi-kampan

was not perceived as an epitome of a pro-Western statesman).[112] Then, the main reason pushing the experts close to the "I don´t know" estimates in summer 2016 could be a simple fact that almost no one (including the then-incumbent president) had officially declared the candidacy.[113]

To conclude this sub-part, three assessed statements with the lowest standard deviation, therefore with a significantly identifiable consensus among experts in assigning particular probabilities, are part of the block regarding the Czech Republic. Other three statements characterised by very low standard deviation relate to the block of one-date-event statements. At first sight, this concordance seems to be somehow connected with some general aspects or patterns belonging to these blocks. Nonetheless, not only the fact that each of two blocks contains questions with considerably high standard deviation but mainly the context itself shows that the specificity of a particular event, rather than some hidden patterns in the blocks, should be considered as a cause of congruence (in assigned probabilities) in these questions.

Since the European Values´ research project represents the first serious attempt in the Czech Republic to foresee future events by using probabilistic forecasting, it is worth taking a look at two parameters that can signal possible inexperience of invited experts.

The first parameter, and also an indicator, is the so-called "anti-probabilistic" thinking. In the theoretical chapter, we learned about the "golden mean" of the possibility of the knowledge, thoughts of optimistic sceptics, and therefore also about fundamental elements of probabilistic thinking and probabilistic forecasting as such. Due to the complexity of our world and the existence of the Black Swans (ontological barriers), or even due to natural cognitive capacities (epistemological barriers), certainty has its limits and perfect certainty is unattainable. But the chapter also showed us that if we recognize these limits, it is possible to see into the future, at least to some extent. It is precisely probabilistic forecasting that recognizes the existence of ubiquitous uncertainty and the limits of perfect predictability.

Therefore, saying something will or will not happen with 100% or 0% probability seems to be, after all, anti-probabilistic. Such statements admit no possibility of uncertainty and surprise. Furthermore, in some cases, this kind of thinking can have a disastrous impact

[112] ČERVENKA, Jan. Důvěra ústavním institucím v únoru 2016: TISKOVÁ ZPRÁVA. In: Centrum pro výzkum veřejného mínění (CVVM) [online]. [cit. 2019-05-22]. Available from: https://cvvm.soc.cas.cz/media/com_form2content/documents/c2/a2020/f9/pi160301.pdf

[113] Čas ubývá, kandidátů přibývá. Seznamte se s těmi, kteří chtějí být českým prezidentem. In: Česká televize: ČT 24 [online]. [cit. 2019-05-22]. Available from: https://ct24.ceskatelevize.cz/domaci/2286658-cas-ubyva-kandidatu-pribyva-seznamte-se-s-temi-kteri-chteji-byt-ceskym-prezidentem

(e.g. When the National Intelligence Estimate – although intelligence analyses almost always involved uncertainty – said Iraq has the weapons of mass destruction).[114] Obviously, specific forecasted events should be assessed by these probabilities. If a prime minister asks you, for example, "what is the probability that an enemy WMD-attack will cause heavy casualties among our civilians?", and your country does not possess ballistic missile defence, you should probably answer 100%. But, notwithstanding those exceptional cases, aforesaid theoretical postulates clearly say that a higher amount of assigned 100% and 0% probabilities ("anti-probabilistic" thinking) goes against the essence of probabilistic forecasting.

Even in terms of practical use of probabilistic forecasting, frequent occurrence of "anti-probabilistic" estimates does not seem to make much sense. By definition, probabilistic forecasts do not ask questions about events where perfect certainty is anticipated, and the case of the European Values´ research project is no exception. Moreover, many of the European Values´ questions contain more than one element and thus should encourage the experts to "play safer". For instance, the European Values asked the experts to assess statement 2.1: "*The Czech Republic will continue to be in a solid institutional core of the EU and NATO*". This question combines two different elements, thus two different statements (Czech Republic as a part of the institutional core of the EU; Czech Republic as a part of the institutional core of NATO). The literature suggests it should be harder to principally reject or accept two linked hypotheses and hence also hard to assign "1" or "0" certainty to the questions consisting of two elements (yet two experts assigned 100% probability to the 2.1 statement).[115] A further problem is a vagueness in the think-tank's questions, making it even more difficult for invited experts to – besides "honestly" assigning their judgements – assign "anti-probabilistic" estimates. The vagueness (absence of clear operationalization) is a persisting problem related to the European Values (see more in the second part of this chapter). In any case, both the possibility of knowledge (existence of uncertainty) and practical nature of probabilistic forecasts (conceding that uncertainty) suggest that frequent playing with 0% and 100% estimates is, by its nature, "anti-probabilistic". As mentioned above, this phenomenon signals the lack of forecaster´s experience.

More precisely, the source of "anti-probabilistic" thinking can be found – to a certain extent – in Isaiah Berlin´s hedgehogs-foxes division of behaviour tendencies. The hedgehogs are often very confident in assigning probabilities and also – often – tend to see the world

[114] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 84.
[115] See for example: WILKS, Daniel S. Statistical methods in the atmospheric sciences. Academic press, second edition, 2006.

only through two settings on their mental dial (between the occurrence and non-occurrence of the event). They do not respect the existence of uncertainty and are therefore likelier to say particular event definitely will or will not happen ("anti-probabilistic" thinking). Conversely, foxes are more prone towards safer plays and try to think twice before the final assessment. Their less-confident behaviour can push them to choose probabilities around the 50% zone (see below).[116] But what is interesting about the hedgehogs-foxes division is that the experts well trained in probabilities or super-forecasters (those with great Brier scores over a long period) tend to avoid 100% and 0% probabilities whilst inexperienced experts incline towards hedgehog behaviour and tend to use them more often.[117]

The second indicator signalling possible inexperience of invited experts relates to the signs of extreme foxes´ behaviour – the indecisiveness. As we already know, decisiveness is a tendency expressed by the resolution (one of the components of the Brier score, see the methodological chapter). But we do not need to see the final Brier scores of experts in order to say something about how decisive these experts are in assigning the probabilities. In the methodological chapter, we learned that forecasters are said to have poor resolution and hence decisiveness if they never stray out (in assigning probabilities) of the "maybe" zone between 40% and 60%. There are two parts of the "maybe zone": the minor shades of this "maybe zone" (40% probability and 60% probability) which is very close to the "I don´t know" estimate but still expresses the probabilistic thinking, and the 50% probability expressing the coin-flipping (random guessing). Assigning a 40% or 60% probability does not necessarily mean that you are poor at forecasting because still, you can get a good Brier score. It simply means that you are not decisive (you are an extreme fox). Perhaps, assigning probabilities in the minor shades of "maybe zone" (not being decisive) can be a problem in the politico-security area, where decision-making stakeholders need more decisive estimates.

But assigning precisely a 50% probability often represents a reverse act of "anti-probabilistic" thinking. This time, it completely accepts the sceptical thoughts of totally unforeseeable future and goes against the logic of probabilistic forecasting. In a similar vein to the use of 100% or 0% probability, some situations call for a verdict that there is precisely a 50% chance that particular event will happen. Furthermore, a 50% probability can sometimes act as a very informative estimate and hence can be highly meaningful. Imagine a meteorologist specializing in the predictions of extreme weather events who states there is a

[116] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 69-70.
[117] Ibid., p. 141-143.

50% chance that a violent tornado will occur in Prague during the next two days. This would certainly not be an "I don´t know estimate" (simple guess or coin flip). Quite contrarily, it would actually express a relatively strong statement about the likelihood of an – "until then" – unimaginable Prague tornado.[118] However, once again, we do not expect probabilistic forecasting projects in the politico-security realm to create many of these 50% situations for invited experts.

Interestingly enough, the literature again suggests that well-trained experts and super-forecasters – again – try to stray out of the minor shades of "maybe" zone (when possible). It does not mean they avoid the use of 40% or 60% probabilities. These probabilities have its place in their repertoire; however, they try to use them only exceptionally. If well-trained forecasters and super-forecaster face a question about the event loaded with high uncertainty, they may keep initial estimates inside this zone and then – as Philip E. Tetlock recommends (see the theoretical chapter) – seek to find more information and update their probabilities in order to move out of the zone. Following the "legacy" of the Bayesian thinking (see again the theoretical chapter), they approach toward more courageous estimates through systematic tracking of their accuracy and careful updating before the final assessment. The same applies to a 50% probability (provided that such probability is not the best possible estimate or is not highly meaningful, see above). Conversely, inexperienced forecasters tend not to make the distinction between the "maybe" zone and other probabilities. They are naturally more open to "play" with them more frequently. Hence, the higher incidence of these probabilities linked to a poor decisiveness is usually referred to as an indicator of forecasters´ inexperience.[119]

With regard to what is said above, we can now move on to the analysis of these two indicators of inexperience. What does the data show in terms of "anti-probabilistic" thinking and poor decisiveness?

Regarding the "anti-probabilistic" thinking, 70 out of 530 assigned probabilities have the certainty-values of 1 (100% probability) or 0 (0% probability). That is, 13% of all assigned probabilities states something will or will not certainly happen, and this still applies when analysing particular blocks of questions (approximate percentage share of 1 or 0 estimates of total number of answers in each block: questions used in the European Values´ research project = 11%; questions not used in the research project = 14%; questions concerning one-date-events = 14%; questions concerning loosely-time-framed processes =

[118] DOSWELL III, Charles A. Probability, Climatology, and Forecasting: Essay on Blog Page. In: Flame [online]. [cit. 2019-06-04]. Available from: http://www.flame.org/~cdoswell/forecasting/probability/Prob_Clim_Fcstg.html
[119] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 143-144.

14%; questions regarding the Czech Republic = 14% and questions regarding the world = 12%).

When it comes to the poor decisiveness, 154 out of 530 assigned probabilities have a value of 40%, 50%, or 60%. To be more precise, the experts chose to assign the minor-shades-of-maybe-zone probabilities (40% or 60%) to a 106 out of 530 particular estimates (20% of all estimates). In terms of individual blocks, the percentage share of 40% and 60% estimates is lower in statements with loosely-time-framed processes (20%) than in one-date statements (23%); and also in non-used questions (19%) than in the statements used in the European Values´ research project (23%). The differences in percentage shares in the questions concerning the Czech Republic and the questions regarding the world reach almost the same level (21% and 17%). Invited experts also showed signs of the reverse "anti-probabilistic" thinking (coin flipping), and assigned the 50% probability in 48 cases (out of 530 probabilities).

At first glance, it would appear that these are certainly not negligible numbers. To put this differently, our dataset shows that approximately every seventh expert´s estimate has the "anti-probabilistic" value of 1 or 0 (and we know the European Values did not raise questions with obvious heavy-casualties-type answers, see the example above). Moreover, almost 30% of all assigned probabilities by the experts have the value of 40%, 50%, or 60%, so approximately every third estimate can be found within the "maybe" zone of poor decisiveness. Based on the two parameters, invited experts show signs of both hedgehog ("anti-probabilistic") and extreme-fox behaviour (poor decisiveness). Nevertheless, do these numbers really represent "high occurrence" of "anti-probabilistic" and poor-decisiveness probabilities? Unfortunately, no clear conclusion can be made without a proper comparison.

It would not be fair to compare invited experts with a few super-forecasters in tournaments sponsored by IARPA. The fairest, most correct and also commonly used method is to compare assigned probabilities with the random – meaning the theoretical – frequency distribution of those probabilities. If we compare these distributions, we find that the invited experts did remarkably well in assigning 100%, 0% (certainty) and even 40%, 50% and 60% probabilities ("maybe" zone). Note that invited experts used a 10 percentage point scale (see Appendix B). Based on the random frequency distribution, every possible probability is expected to have the same percentage representation across the whole dataset (530 assigned probabilities). Quite simply, the percentage share of assigned probabilities should be 9% (approximately 48 estimates) for each of all eleven (from 0 to 1) possible estimates. For 100% (1) and 0% (0) probabilities, we would expect assigned probabilities to account for 18% (9%

for 1, and 9% for 0, which means approximately 96 probabilities) of all estimates. Accordingly, the anticipated share of 40%, 50% and 60% probabilities would be 27% (9% for 0.4, 9% for 0.5, and 9% for 0.6, which means approximately 144 probabilities). Table 2 illustrates, for ease of reference, both numeral and percentage share of all assigned probabilities and individual blocks of questions.

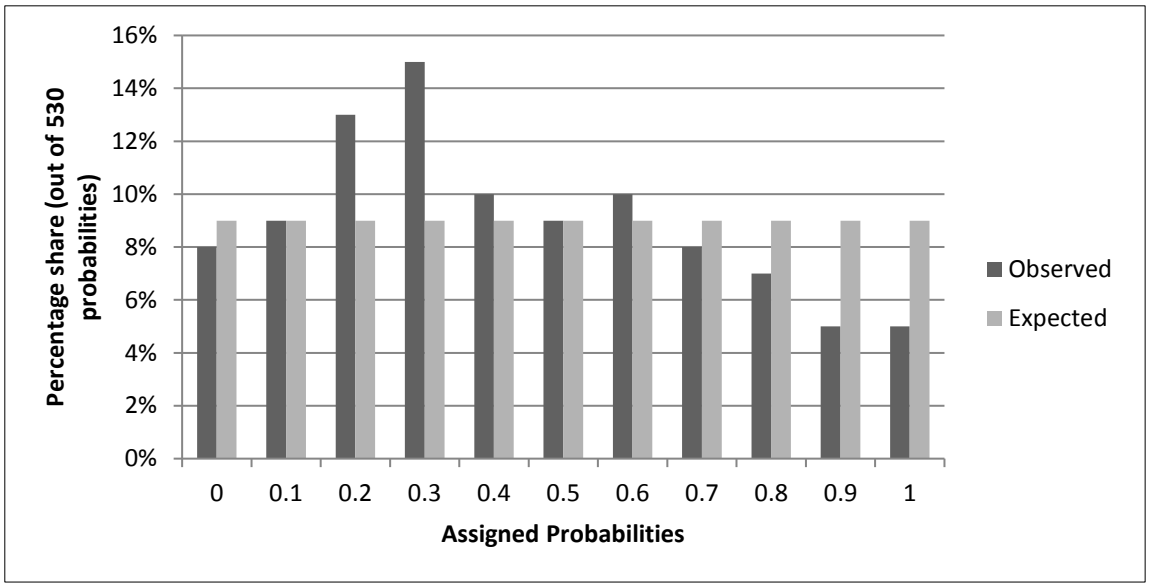
Table 2: Numeric and percentage (approximate) share of all assigned (observed) probabilities

	Numeral and percentage share of each estimate										
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
Data (number of estimates)											
All probabilities (530)	**45x** (8%)	**48x** (9%)	**71x** (13%)	**79x** (15%)	**53x** (10%)	**48x** (9%)	**53x** (10%)	**45x** (8%)	**39x** (7%)	**24x** (5%)	**25x** (5%)
Non-used estimates (419)*	**42x** (10%)	**46x** (11%)	**62x** (14%)	**72x** (17%)	**43x** (10%)	**34x** (8%)	**38x** (9%)	**33x** (8%)	**24x** (6%)	**17x** (4%)	**16x** (4%)
Used estimates (111)	**3x** (3%)	**2x** (2%)	**9x** (9%)	**7x** (7%)	**10x** (9%)	**14x** (13%)	**15x** (14%)	**12x** (12%)	**15x** (15%)	**7x** (7%)	**9x** (8%)
Questions regarding the CR (336)	**37x** (11%)	**30x** (9%)	**54x** (16%)	**59x** (18%)	**38x** (11%)	**28x** (8%)	**35x** (10%)	**23x** (7%)	**20x** (6%)	**2x** (1%)	**10x** (3%)
Questions concerning the world (194)	**8x** (4%)	**18x** (9%)	**17x** (9%)	**20x** (10%)	**15x** (8%)	**20x** (10%)	**18x** (9%)	**22x** (11%)	**19x** (10%)	**22x** (11%)	**15x** (8%)
One-date-statements (48)	**4x** (8%)	**7x** (15%)	**3x** (6%)	**5x** (10%)	**6x** (13%)	**6x** (13%)	**5x** (10%)	**5x** (10%)	**4x** (8%)	**0x** (0%)	**3x** (6%)
Loosely-time-framed processes (482)	**41x** (9%)	**41x** (9%)	**68x** (14%)	**74x** (15%)	**47x** (10%)	**42x** (9%)	**48x** (10%)	**40x** (8%)	**35x** (7%)	**24x** (5%)	**22x** (5%)

source: own elaboration (based on the data in Appendix B). * Non-used estimates are those estimates (and corresponding questions) which were not used by the European Values in their research project (those used are labelled here as "Used estimates".

In addition, Figure 1 shows the whole percentage frequency distribution, both for expected (random frequency distribution) and observed probabilities. This same figure subsequently proves that the proportion of 100% and 0% probabilities assigned by invited experts is, all in all, significantly lower compared to the expected frequency distribution (13% share in observed data, which means 70 probabilities, compared to the 18% in expected data, accounting for 96 probabilities).

Figure 1: Percentage frequency distribution of all expected and observed probabilities



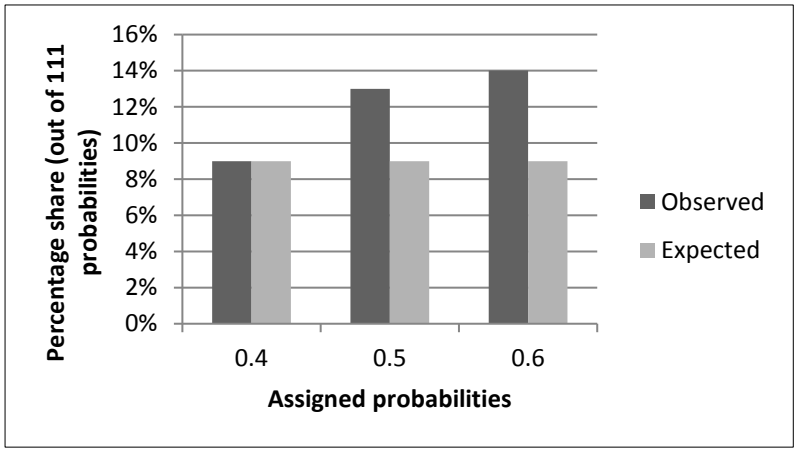
source: own elaboration (based on the data in Appendix B)

Furthermore, Pearson´s chi-squared test validates claim described above about – overall - "significantly" lower share (in real numbers) of "anti-probabilistic" estimates assigned by invited experts (see Table 1). The *p*-value of 0.000814 says that the probability of no statistical relationship between observed and expected data is extremely low. Thus, if we choose the significance level (α value), for example, of 0.05, we can say there is a statistically significant difference between observed and expected data, and that this difference goes in favour of invited experts. Essentially, the share of "anti-probabilistic" values assigned by invited experts is really statistically lower than the expected proportion.

From the perspective of poor decisiveness, both Figure 1 and Table 2 illustrate a slightly higher percentage share of 40%, 50% and 60% estimates assigned by invited experts. Compared to the expected 27% share (144 probabilities), the experts assigned these extreme-fox probabilities to 29% of all cases (154 probabilities). However, Pearson´s chi-squared test again plays into the hands of experts invited by the European Values. In this case, the *p*-value 0.594025 leads us to the conclusion that on the same significance level (α value of 0.05); observed frequency distribution does not differ from our theoretical distribution. The share of poor-decisiveness estimates assigned by invited experts is, therefore, not significantly higher than the expected one. This share fluctuates only minimally in the block of non-used questions (27%), statements regarding the Czech Republic (29%), those concerning the world (27%) and in the loosely-time-framed processes (29%). However, two other blocks – questions used in the European Values´ research project and one-date statements – report the

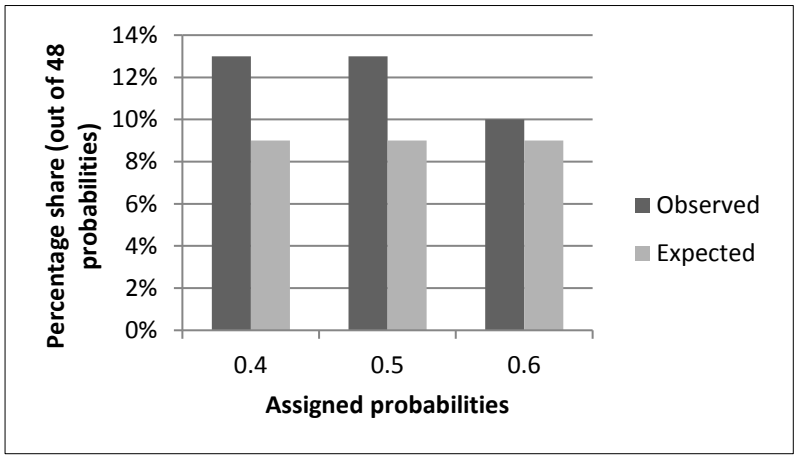
higher share of extreme-foxes probabilities, and thus also increase the overall proportion of these probabilities in all dataset. The expected, theoretical proportion within the block of European Values´ used questions is 30 out of 111 estimates (27%), whilst the observed one is 39 out of 111 estimates (approximately 36%). For the one-date statements, the expected share is approximately 12 out of 48 probabilities (27%) but the real one is 17 out of 48 probabilities (almost 36%). Both blocks are set out in Figure 2 and Figure 3.

Figure 2: Percentage frequency distribution of expected and observed probabilities (0.4, 0.5, and 0.6) in "used" questions



source: own elaboration (based on the data in Appendix B)

Figure 3: Percentage frequency distribution of expected and observed probabilities (0.4, 0.5, and 0.6) in one-date statements



source: own elaboration (based on the data in Appendix B)

Although both blocks evince the highest percentage share of the use of extreme-fox behaviour (each with aforesaid 36%), there is no statistically significant difference between expected and observed real numbers (these differences are not significantly higher). Both the *p*-value of 0.128735 for "used" statements and the *p*-value of 0.324652 for one-date statements are higher than α level of 0.05 (see all *p*-values regarding the "anti-probabilistic" thinking and poor decisiveness in Table 3 and Table 4 below). Within the block of questions

used in the European Values´ research project, the statement about Hilary Clinton becoming the US president (statement 3.1) is the one that contains the highest number of assigned poor-decisiveness probabilities. Notwithstanding this, the fact that experts´ statements chosen – by the European Values – for the research project are at the same time the most indecisive ones, is at least interesting. The Hilary-Clinton-President statement along with its assigned probabilities also negatively affects the incidence of poor decisiveness within the one-date questions block. However, here, the question of whether the pro-Western candidate will win the Czech 2018 presidential elections (statement 4.5) too contributes to that 36% share of extreme-foxes probabilities. But as shown above, it makes sense to assign a 50% probability (summer 2016) when there is almost no official presidential candidate.

In light of the above analysis, neither the hedgehog "anti-probabilistic" thinking nor the extreme-fox poor decisiveness is visible in the behaviour of invited experts. The "fair" and commonly used comparison based on the random (theoretical) frequency distribution clearly demonstrates that the experts´ utilization of 0% and 100% probabilities are statistically significantly lower to what should be expected. The use of poor-decisiveness estimates is slightly higher, yet with no statistically significant difference. And this is really admirable performance when one realizes that two other factors, ambiguous questions, and low granularity, could push the experts into the "maybe" zone of poor-decisiveness.

Table 3: Summary of the Pearson´s chi-squared test in terms of the "anti-probabilistic" thinking (α value of 0.05)

	"Anti-probabilistic" thinking (0 and 1)			
	EXPECTED FREQUENCY (Approximate) 0; 1	**OBSERVED FREQUENCY 0; 1**	***p*-value**	**Statistically significant difference**
Whole Dataset	**48; 48**	**45; 25**	**0.000814**	**Yes, in favour of experts**
Non-used estimates*	38; 38	42; 16	0.000286	Yes, in favour of experts
Used estimates	10; 10	3; 9	0.025347	Yes, in favour of experts
Questions regarding the CR	30; 30	37; 10	0.000109	Yes, in favour of experts
Questions concerning the World	18; 18	8; 15	0.013862	Yes, in favour of experts
One-date-statements	4; 4	4; 3	0.617075	No, but in favour of experts
Loosely-time-framed processes	44; 44	41; 22	0.000816	Yes, in favour of experts

source: own elaboration (based on the data in Appendix B).* Non-used estimates are those estimates (and corresponding questions) which were not used by the European Values in their research project (those used are labelled here as "Used estimates").

Table 4: Summary of the Pearson´s chi-squared test in terms of the poor decisiveness (α value of 0.05)

	Poor decisiveness (0.4, 0.5, and 0,6)			
	EXPECTED FREQUENCY (Approximate) 0.4; 0.5; 0.6	**OBSERVED FREQUENCY 0.4; 0.5; 0.6**	***p*-value**	**Statistically significant difference**
Whole Dataset	**48; 48; 48**	**53; 48; 53**	**0.594045**	**No, but in favour of experts**
Non-used estimates*	38; 38; 38	43; 34; 38	0.58304	No, but in favour of experts
Used estimates	10; 10; 10	10; 14; 15	0.128735	No, but in favour of experts
Questions regarding the CR	30; 30; 30	38; 28; 35	0.212259	No, but in favour of experts
Questions concerning the World	18; 18; 18	15; 20; 18	0.696909	No, but in favour of experts
One-date-statements	4; 4; 4	6; 6; 5	0.324652	No, but in favour of experts
Loosely-time-framed processes	44; 44; 44	47; 42; 48	0.719283	No, but in favour of experts

source: own elaboration (based on the data in Appendix B). * Non-used estimates are those estimates (and corresponding questions) which were not used by the European Values in their research project (those used are labelled here as "Used estimates".

The ambiguity refers to the problem mentioned above with the questions (statements) containing more than one element (see the above example 2.1 about the Czech Republic continuing to be in the EU and NATO institutional core). We already know it should be harder to reject or accept statements with two linked hypotheses principally and to, therefore, assign "anti-probabilistic" probabilities of 0 and 1. This may have contributed to the success of experts in their significantly low use of 0 and 1 probability. On the other hand, and similarly to the 0 and 1 probabilities, many elements could also sometimes prevent forecaster from assigning more resolute (e.g. 20% and 30% or 70% and 80%) probabilities, thus potentially push them into the "maybe" zone. Furthermore, the invited experts had problems with the "compatibility" of some elements. Using the same 2.1 example – as one of the experts wrote into the questionnaire – the parallel membership in the EU and NATO solid institutional cores cannot be simply merged (another expert even claimed that the Czech Republic is not in the constitutional core of the EU so it cannot "continue to be there"). Also, five forecasters identified 5 out of 53 statements as ambiguous, incompatible, or flawed. It is conceivable, again, that unambiguous questions restrain the resolution of experts. It could make it a little harder for experts to stray out of the "maybe" zone.

The second factor that possibly pushes the experts towards poor decisiveness and makes it harder for experts to stray out of the "maybe" zone is the low degree of granularity. In general, granularity reflects the range of potential answers to questions or statements.[120] As Barbara A. Mellers claims, in probabilistic forecasting, high granularity is usually represented by "*the ability to assign finer-grained distinctions along the probability scales*".[121] That is to say, high granularity refers to the use of a single percentage point scale (medium granularity is usually defined by the use of a five percentage point scale). Well-trained forecasters and super-forecasters do not hesitate to exploit full gamut of 0-100 scale. Instead of saying there is 30%, 40% or 80% probability that something will happen, they tend to choose values like 32%, 48%, or 87%. This enables them not only to be more precise (forecasters who use 1% scale are more accurate than those sticking to five percentage point scale, and even more than those sticking to 10 percentage point scale), but also to move out of the zone of poor-decisiveness (the "maybe" zone). [122] Put it simply; these forecasters have more options – if necessary – to avoid this zone (both areas of 40% or 60%, and 50% random-guessing estimate). Unlike the European Values´ statements (assigned mostly without the experts´

[120] SCHEGLOFF, Emanuel. On granularity. Annual Review of Sociology, 2000, 26.1: p. 715.
[121] MELLERS, Barbara, et al. Identifying and cultivating superforecasters as a method of improving probabilistic predictions. Perspectives on Psychological Science, 2015, 10.3: p. 275.
[122] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 145.

estimates) used in their research projects, which utilizes the medium granularity (e.g. 15%, 20%, 25%), the invited experts bet on 10 percentage point scale. Hence, the only possibility for them to escape the area of poor decisiveness is to assign probabilities higher than 60% (70%, 80%, and 90%) or lower than 40% (30%, 20%, and 10%). The problem with low granularity is also that it does away with information. Let's take a 70% probability as an example. When using the 10 percentage point scale, it makes sense to assign 0.7 estimate only when we feel the objective chance that some event will happen, is between 64% and 74% (rounding). We barely need to mention that in reality, there is a difference between saying there is a 65% probability that someone will survive a high-risk surgery or that the probability is 74%.

Unfortunately, one cannot say whether the use of 10 percentage point scale results from a requirement of European Values or the choice of invited experts. But what we can say - leaving aside the problem with information and accuracy – is that low granularity puts experts into a difficult position in terms of poor-decisiveness.

Besides presenting some elementary statistical data (e.g. the concordance among experts), the internal analysis of estimates of the invited experts explores some more intriguing findings. Looking at the two parameters signalling the possible inexperience of invited experts, neither the hedgehog (anti-probabilistic) patterns nor the extreme-foxes poor decisiveness is visible in their behaviour. More interestingly, even though invited experts faced two factors (questionnaire-statements containing many elements, and low granularity) – potentially – pushing them into the "maybe" zone of poor decisiveness, they escaped and managed to stray out of it. The results of the Pearson´s chi-squared test played into the hands of experts invited by the European Values. In this regard, the invited experts demonstrated a certain level of quality, and one could say the think-tank´s selection of experts was a good bet.

Let us look at how the European Values´ research project (made primarily by their own research team) performs in these two parameters as well as at some surprising findings.

4.2. Internal Analysis of the European Values´ Research Project

Before proceeding to the accuracy assessment of the European Values, let us briefly look at some interesting findings, positives and also recurrent shortcomings regarding the 58 questions (statements) and 79 estimates in their research project (see more details in Appendix A).

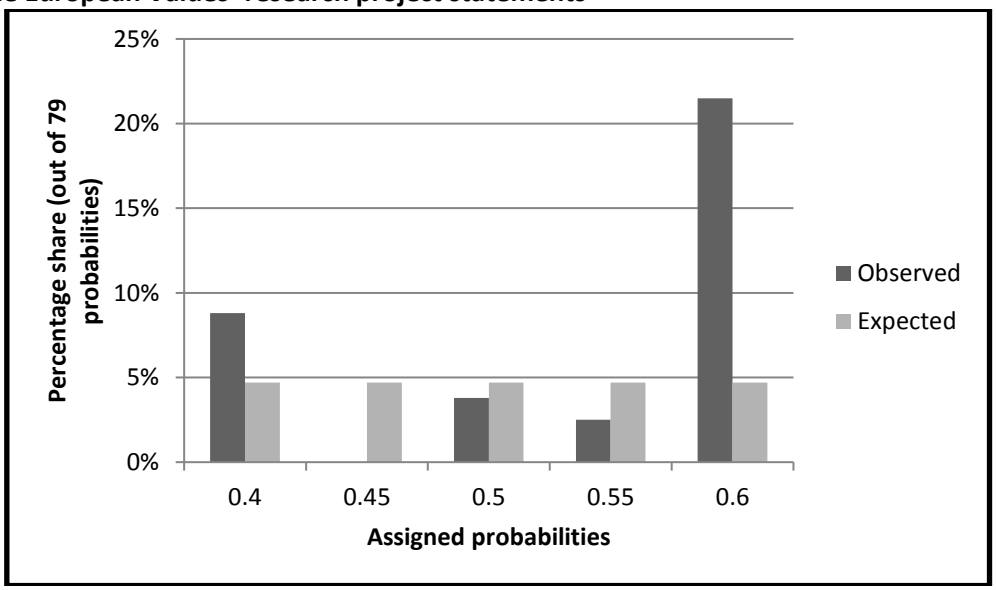
Although this diploma thesis strives to bring the verification (as one of the elements of rigor) and therefore focuses on predictive capabilities (the accuracy) of the European Values, the think-tank deserves to be praised for incorporating some normative aspects into their forecasts. Besides using traditional probabilistic explanatory forecasting ("what is the probability that something could happen?" and "how could something happen" questions), the European Values also frequently employs more creative and constructive thinking of normative forecasting ("what resulting effects could particular events have on defined interest/values/goals?").[123] Thus, for example, we find not only that the probability of CDU-Greens coalition forming was 55% but also that this event could have had both positive (German policy towards the Russian Federation) and negative impacts (strongly liberal asylum policy) on the interests of the Czech Republic.[124] Such a "storytelling" approach also relates to the prefaces. Almost every subchapter of the partial finding of the research project is accompanied by prologues explaining the importance of particular dynamics (concerning the not-yet-unfolded events).

Another positive aspect is the utilization of medium granularity. Unlike experts´ estimates (10 percentage point scale), the research project sticks to the finer-grained distinctions along 5 percentage point scale. This not only potentially encourages more exactness and avoidance of poor decisiveness, but also provides us with – compared to the "tens" – more information. Yet the 5 percentage point scale is used "only" in 11 assigned probabilities (out of 79). Furthermore, the European Values did excellent work in the area of "anti-probabilistic" thinking. None of the 79 assigned probabilities has values of 1 or 0; which also means that there is no hedgehog behaviour signalling the occurrence of inexperience. However, on the other hand, the European Values were not able to sufficiently stray out of the "maybe" zone of poor decisiveness. Based on the same method of theoretical (random) frequency distribution and taking into account the medium granularity, both percentage and

[123] KUOSA, Tuomo. The evolution of strategic foresight: navigating public policy making. Routledge, 2016.
[124] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 10. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

numerical share of observed poor-decisiveness values (0.4, 0.45, 0.5, 0.55 and 0.6) is evidently higher than expected one: approximately 37% (29 probabilities) compared to 22.5% (approximately 19 probabilities). This is caused mainly by the high share of 60% probabilities. The percentage share is set out in Figure 4.

Figure 4: Percentage frequency distribution of expected and observed probabilities (0.4, 0.5, and 0.6) in the European Values´ research project statements



source: own elaboration (based on the data in Appendix A)

Pearson´s chi-squared test only confirms this with the *p*-value of 0.000000000027, and the value remains very low even if we do not include those poor-decisiveness estimates of invited experts which are used in the research project (*p*-value of 0.00000001). Hence, we can state (with the same α value of 0.05) that there is a statistically significant difference between observed and expected data; and that the share of poor-decisiveness probabilities is significantly higher and goes against the European Values. To put this into perspective, even though the European Values´ research project "had an advantage" in the form of granularity, they failed to capitalize on it and, in the end, turned out to be worse in poor-decisiveness than invited experts in their questionnaire. Unlike invited experts, the second parameter – extreme-fox behaviour related to the poor-decisiveness – thus also indicates signs of inexperience in the research project of the European Values. It is important to note here that the estimates of the invited experts are used only sporadically in the research project.

In addition, the European Values´ research project shows some recurrent shortcomings – not only – in the wording of questions (statements) that cannot be overlooked and shed a negative light on their work. The first one concerns several linguistic errors. On the one hand, the European Values fairly often and also correctly use phrases such as "more probable", "less

probable" or "more likely" and "less likely". This approach certainly is a step in the right direction and follows the tradition of Sherman Kent´s intelligence terminology. On the other hand, they just as often begin their judgments with "we expect", "we assume" or "we consider" phrases which are quite non-standard in the field of scenario or probabilistic forecast-making.[125] That is not necessarily a problem as long as these phrases match the assigned probability. However, two statements do not fulfil this logical precondition. The first statement regarding the future United States-Poland relationship is: "*Regarding Poland's stance towards the United States, we expect the preservation of the current state, which is characterized by American political restraint on the background of continuing cooperation in the field of defence (50%)*".[126] The second one, estimating the result of the 2018 United States midterm election and stating that "*...based on previous experience, the party, whose candidate did not succeed in 2016, will win (midterms)*", is complemented by four same 25% probabilities (Clinton wins both elections - 25%; Trump proves himself in the opposition and wins the Midterms - 25%; Trump wins Midterms - 25%, and Trump wins presidential elections but Clinton wins the Midterms - 25%).[127] The inconsistency between the language and probabilistic expression is obvious in both statements. Whereas the former estimate sets the 50% probability ("I don´t know") against the stated confidence ("we expect"), the assigning of four equal (25%) probabilities to four possible "futures" (even odds) in the latter one logically goes against the worded expression ("the one who loses the presidential election, will win the midterms"). If nothing else, these linguistic errors lead one to think about the professionalism of the research project.

Another recurrent shortcoming once again refers to the linguistic errors, but this time more systemic and related to the vagueness and lack of "operationalization". In other words, thirteen questions (statements) contain an ambiguous threshold for either the rejection or acceptance of an event´s occurrence (whether the examined dynamics happened or not). This accounts for almost one-fourth (22%) of all questions. Moreover, these vaguely constructed statements blur vital information about what was supposed to happen. The following example may illustrate this. In a block of statements focusing on the possible outcome of the 2017

[125] FRIEDMAN, Jeffrey A.; ZECKHAUSER, Richard. Handling and mishandling estimative probability: likelihood, confidence, and the search for Bin Laden. Intelligence and National Security, 2015, 30.1: p. 15.
[126] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 15-16. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf
[127] Ibid., p. 14 and (table) p. 23.

French presidential and National-Assembly elections, the European Values states that: "*we expect a close defeat of the National Front (...) president Marine Le Pen (60%)*".[128] But what exactly does the "close defeat" mean? Is a 49% to 51% defeat the only "close one", or, could we also count 4%, 5%, 8% difference as the "close defeat"? Alternatively, and in the context of 2002 difference between Jacques Chirac and Jean-Marie Le pen (82% to 18%), could we think of – similarly to certain media – an actual 34-66 percent defeat as the "close one?".[129] There is simply no clear and specific operational definition of the key concept (close defeat) constituting the European Values´ hypothesis about the election result. The level of allowed abstraction is therefore too high, and the same applies to other statements used in the research project, such as:

- "*...we expect a **dramatic increase** in the presence of US troops in the Middle East (80%)*"
- "*The amount of migrants and asylum seekers in the European Union will not **significantly decrease** from year-to-year (60%)*"
- "*We expect that the percentage of mainstream parties in electoral preferences (...) will fall below 50% in a **significant part** of the European Union (80%)*"
- "*...there will probably be a **noticeable economic downturn** in Russia (70%)*"
- "*We assume that in all areas from which migrants are coming to the EU, the **current situation** will be preserved (Syria – 60%), with the exception of Africa, where we anticipate a **rapid deterioration** (80%)*"[130]

Although the final list of statements with poorly operationalized concepts is longer than this illustration, the issue with vagueness remains the same. What could be the number that sets a threshold for a "dramatic increase" of US troop in the Middle East? Would ten thousand, thirty thousand or rather fifty thousand soldiers fulfil the criterion? And imagine a theoretical situation where there is a president who systematically – and for a long period of time – uses the discourse of military non-increasing in the Middle East. Then, could a sending of "mere" three thousand soldiers be perceived as a dramatic increase? The problem with the

[128] Ibid., p. 11.
[129] THAROOR, Ishaan. Macron defeated Le Pen in France's presidential election. Here's what happens next. In: The Washington Post [online]. [cit. 2019-06-21]. Available from: https://www.washingtonpost.com/news/worldviews/wp/2017/05/07/macron-defeated-le-pen-in-frances-presidential-election-heres-what-happens-next/?utm_term=.f553d6a9c3f2
[130] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 13, 14, 18, 19, and 21. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

concept of significant decrease of migrants and asylum seekers in the European Union is identical. Considering Russia´s economic dynamics, it is unclear whether the "economic downturn" refers to one of several economic definitions (e.g. non-positive growth of real GDP per capita)[131] or – due to the term "noticeable" – to something else. In the case of fall of mainstream parties, it is hard to determine whether the – for instance – Visegrád Group meets the requirement of a "significant part". Finally, the absence of any concreteness in the last statement entails an emergence of numerous aspects and dynamics that could constitute "current preservation of Syria´s situation" (e.g. social, economic or political?) or "rapid deterioration in Africa" (political dynamics, poverty, environmental problems or all aspects together?).

As indicated above, such vagueness coupled with a lack of operationalization obstructs the effort to verify the European Values´ accuracy in these statements (see in the next chapter). More importantly, it partly devalues the relevance of information about future phenomena. The European Values themselves confirm that their probabilistic-forecasting research project should serve policy-makers in the politico-security area as a tool for readiness and well-timed decisions.[132] Frankly, the policy-maker in the Czech Republic does not need to know if the defeat of Marine Le Pen "will" be close or decisive (it is simply a defeat). However, in order to adapt foreign policy to the future events, he or she would probably like to know whether the 80% probability of no dramatic military increase (in the Middle East) of our ally applies equally for maximum of 3000 soldiers or 15 000 soldiers, or similarly, what "will" the rapid deterioration in Africa mean. That is the reason why well-established institutions and platforms for probabilistic forecasting – not only – in the politico-security area try to ensure that their statements and hypotheses will be as much specific as possible. Following statements from Tetlock´s "*The Good Judgment Project*" are technically similar to those of the European Values and provide a perfect example of this approach:

- "*Will the United States experience* ***two consecutive quarters of a negative real GDP growth rate*** *in 2019*?"[133]

[131] THARAVANIJ, Piyapas. Capital Market, Severity of Business Cycle, and Probability of an Economic Downturn. Monash University Economics Discussion Paper, 2007. p. 13.

[132] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 27. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

[133] Question 5 of 11 in Global Judgment Challenge. In: Good Judgment Open [online]. [cit. 2019-06-21]. Available from: https://www.gjopen.com/questions/1094-will-the-united-states-experience-two-consecutive-quarters-of-a-negative-real-gdp-growth-rate-in-2019

- "*Will the winner of the 2016 presidential election win the popular vote **by 10 or more percentage points**?*"[134]
- "*Will **more than 1 million refugees** and migrants arrive in Europe **by sea** in 2016?*"[135]

Individual forecasters on the Good Judgment Open website assigned (and in the first case still assign) their probabilities to these questions. Without showing particular estimates, we can clearly see that these questions have both well-operationalized concepts and specific thresholds. Hence, there is no impediment for proper accuracy verification (occurrence/non-occurrence of the clearly defined event) and for a devaluation of a clear message to policy-makers.

The third recurrent shortcoming involves several questions with unstated judgments. To be more precise, the European Values conceal relevant information in four questions with three possible outcomes (A, B or C can happen) as much as in one four-option question (A, B, C or D can happen). In a section inquiring the future development in Poland, the European Values estimate that: "*We believe that an undermining of the liberal character of the Polish political system is more likely to happen than this remaining just an issue of individual excesses (50% against 40%)*". Without a doubt, this certainly is valuable information, but it says nothing about the remaining 10% judgment (complement to 100%). Therefore, we know that the probability of (A) undermining of liberal character is 50%, the probability (B) that this remains an issue of individual excesses is 40%, and that (C) "something else" has the probability of 10%. The same goes for already mentioned (see above) "United States - Poland relationship" statement: "*...we expect the preservation of the current state, which is characterized by American political restraint on the background of continuing cooperation in the field of defence (50%). We do not consider the strengthening of their mutual relations to be that probable (40%)*". Again, the remaining 10% judgment is unstated. So what do those two unstated judgments account for? Fact is that in these two cases it is possible to decipher the meaning of the concealed third event. The unstated judgment in the first case probably says there is a 10% chance of no undermining of the liberal character of the Polish political system. In the second case, the 10% probability could cover a situation where the mutual United States - Poland defence cooperation deteriorates. But this does not alter the fact that

[134] Monkey Cage US Election 2016 Challenge. In: Good Judgment Open [online]. [cit. 2019-06-21]. Available from: https://www.gjopen.com/questions/275-will-the-winner-of-the-2016-presidential-election-win-the-popular-vote-by-10-or-more-percentage-points

[135] GJP Classic Geopolitical Challenge (2016). In: Good Judgment Open [online]. [cit. 2019-06-21]. Available from: https://www.gjopen.com/questions/106-will-more-than-1-million-refugees-and-migrants-arrive-in-europe-by-sea-in-2016

the probabilistic research project should logically always describe all outcomes. Unstated judgments mostly do not present any obstacles to the verification (calculating Brier scores), but they create a situation where policy-makers are not provided with the "complete picture".

In the other two questions, it is much more difficult to figure out what are the unstated judgments about. The statements in the first one are: "*In Poland, the support or at least the tolerance of policies of PiS will result in the country's political system becoming more authoritarian (60%). The option that the majority of Poles will defy the authoritative tendencies of PiS is rather unlikely (30%)*". The statements in the second question are: "*In case of territorial gains outside Syria and Iraq, the Islamic State will either expand its influence (60%) or will continue to create smaller networks across Europe (30%)*". Now imagine you are a policy-maker. Concerning the first statement, does the unstated 10% probability relate to Poland (its political system) becoming less authoritarian (or even non-authoritarian) or – for instance – to a possibility that another actor (not the majority of Poles) defies these tendencies? Similarly, does the 10% probability in the second statement relate to the maintenance of Islamic State influence, to discontinuation of smaller-networks approach, or rather to both dynamics together? This leads to greater confusion in terms of having a "complete picture". Another problem with this statement is that it pits the possibility of expanding influence of the Islamic state against the possibility of its network-approach continuance. But these two "futures" are not mutually exclusive, and possible simultaneous occurrence of both events during a specific time period precludes verification (see more in the next chapter).

Nevertheless, the European Values´ research project also includes one four-option question which generates a reverse difficulty in providing "complete picture" – we can imagine what the two unstated judgments cover, but we do not know the probability. This question is partially mentioned above and regards the 2017 French presidential and parliamentary elections. The European Values give two estimates of possible results in their final table: (A) "*Neither the National Front nor M. Le Pen will win (60%)*", and (B) "*The National Front will win the parliamentary elections and M. Le Pen will win the presidential elections (30%)*".[136] We can see straight away that remaining 10% probability is left for two unstated judgments that concern a possibility of cohabitation: (C) National Front will win the parliamentary elections and M. Le Pen will not win the presidential elections, and (D)

[136] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 23. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

National Front will not win the parliamentary elections and M. Le Pen will win the presidential elections. Taking into account the medium granularity (5 percentage point scale), how can we know whether both (C) and (D) have 5% probability, or whether one of these events have 10% probability whereas the second one has 0% probability? Table 5 illustrates this confusion.

Table 5: Question regarding the 2017 French presidential and parliamentary elections

		Probability	
		Clearly stated or unstated	**Any other possible distribution?**
Stated Judgments	**Neither the National Front nor M. Le Pen will win**	stated - 60%	(no) 60%
	Both the National Front and M. Le pen will win	stated - 30%	(no) 30%
Unstated Judgments	**National Front will win but M. Le Pen will not win**	unstated - 10%	(yes) 0%; 5%; 10%
	National Front will not win but M. Le Pen will win		(yes) 10%; 5%; 0%

source: own elaboration (based on the data in Appendix A)

Lastly, the European Values have unfortunately not been able to avoid mathematical (and logical) error in their research project. The think-tank tried to forecast the result of the Alternative for Germany (AfD) party in the 2017 election for a new Bundestag. The problem lies in the fact that the judgments used in the main text differ from those stated in the final table. Former says that the probability of *a)* AfD getting up to 10 percent of the votes is 70% and *b)* AfD reaching more than 10 percent is 30%.[137] However, the latter ones estimate that the AfD *a)* will get 10 percent of the votes with 70% probability and *b)* will get 20 percent of the votes with 30% probability.[138] Mathematically, these judgments are incompatible and mutually exclusive (see Table 6). It is obvious that the European Values made a mistake and wanted to state something like:

- a) *AfD will get a maximum of 10 percent of the votes (70% probability)*
- b) *A scenario where the AfD reaches more than 10 percent of votes (with a maximum of 20 percent) is less likely (30 % probability).* It implies simultaneously that they assigned 0% to the possibility of the AfD winning more than 20%.

As for all previous shortcomings, such "careless mathematical mistake" is particularly unfortunate for the European Values, since it depreciates – to some extent – the value of their work.

[137] Ibid., p. 9.
[138] Ibid., p. 23.

Table 6: Mathematical incompatibility regarding the AfD election result

		The meaning of assigned probability	
		- Text – "AfD will get..."	**- Final table - "AfD will get..."**
Assigned probability	**70% chance**	0 - 10 percent of the votes	10 percent of the votes
	30% chance	10.01 - 100 percent of the votes	20 percent of the votes

source: own elaboration (based on the data in Appendix A)

To conclude this brief analysis, the European Values´ research project deserves credit not only for incorporating the normative aspects into their probabilistic forecasts but also for providing us with prologues of key dynamics. Although the think-tank failed to stray out of the "maybe" zone and thus assigned a statistically high share of poor-decisiveness probabilities, it did an excellent job in terms of avoiding "anti-probabilistic" (0 and 1) estimates. The utilization of – at least – medium granularity (5 percentage point scale) certainly is a step in the right direction.

Nonetheless, the above mentioned – and often recurrent – shortcomings, such as linguistic errors (incompatibility between worded and numerical expressions) or mathematical mistake, naturally shed a negative light on the research project. The occurrence of unstated judgments or unstated probabilities leads one to think about the level of professionalism among the leaders of the research project. Moreover, the vagueness and absence of clearly operationalized concepts represent more serious shortcomings. At "best", it obstructs the accuracy verification (no clear thresholds). At worst, it acts as an impediment for delivering a clear message for policy-makers in the politico-security area. These shortcomings cannot be simply overlooked. However, it is the following chapter that shows how the European Values perform in terms of accuracy – which is eventually the most important metric for assessing similar projects.

5. The Second Level of Analysis

This chapter finally brings the verification, and hence one of the Tetlock´s aspects of rigor, into the European Values´ probabilistic forecasts. However, several adjustments regarding the think-tank´s forecasted questions (statements) need to be done before proceeding to the accuracy evaluation. These adjustments are explained in the first subchapter.

The second subchapter checks the accuracy of the research project´s estimates by comparing its forecasts against the corresponding observation of what actually occurred (Brier score), and thus measures the predictive capabilities of the European Values.

5.1. Adjustments

By the end of this chapter, we will be able to see the accuracy of the European Values´ research project along with its predictive capabilities. However, before we proceed to the verification, a few specifying information about assessed forecasts need to be known.

Firstly, a relatively high number (thirteen, see in the previous chapter) of poorly operationalized questions (characterized by vagueness and absence of specific thresholds) negatively affects the ability to determine occurrence or non-occurrence of associated events, and hence also the Brier scores. With little effort, one can easily find robust arguments for simultaneous acceptance or rejection of the realization of the single event. As a consequence and despite all efforts, this leads to an involuntary reduction in the number of assessed forecasts. Firstly, there is a natural elimination of six conditional questions where the conditional event did not occur. Furthermore, when it comes to the question dealing with the Temelin Nuclear Power Plant, postponing of the tender to 2020/2021 means that we are, again, unable to decide about the occurrence of the forecasted event.Secondly, there are other four (out of thirteen) poorly operationalized questions (statements) that have to be voided due to the impossibility to decide the event´s occurrence or non-occurrence. Two of them – regarding the situation in Africa and Syria – were based on the experts´ responds from the questionnaire. As a result, the total number of experts´ forecasts used by the think-thank decreases from eleven to nine. In the rest of those ill-defined questions (9 out of 13), accuracy assessment is – at least to some extent – possible. The last round of elimination is related to the questions characterized by different, or multiple difficulties. One of them regarding the Islamic State has to be eliminated due to both unstated judgment and problem with mutual

exclusivity. The second one is related to the mathematical incompatibility in two AfD-result judgments (see the previous chapter). Whereas the judgment in the text ("*AfD will get up to 10 percent of the votes*") is assessed, the judgment in the final table ("AfD will get 20 percent") is voided. On the contrary, another problematic question which incorrectly operates with mutual exclusivity is split into two separate statements. This step prevents another potential elimination. Thus, the total number of questions included in the accuracy assessment is 46 (compared to the original 58 questions). Table 7 (below) illustrates not only above mentioned – and necessary – changes but also the concrete explanation for such a step.

Secondly, several European Values´ statements require adjustment in assigned probability or clarification of its Brier score calculation. The think-tank has created its probabilistic forecasts for – approximately – three year period (from the end of 2016 to the end of 2019). But the accuracy assessment is performed in the middle of 2019, which means only for a two and a half year period. Therefore, in order to be fair to the European Values, we adjust the estimated probabilities if the occurrence of a given event was provided for the three year time horizon (something will happen between the end of 2016 and the end of 2019) for example: "*There will be a systematic series of major terrorist attacks*"). Such a calculation of our adjusted probability (A_p) requires the following formula, where p_n is the probability of non-occurrence of a particular event, T is an overall forecasted period and t marks the adjusted forecasted period.

$$A_p = 1 - (\sqrt[T]{p_n})^t$$

As an example, when European Values estimates there is a 60% probability that there would be mass civil unrest – instigated by far-right – in France, the calculations are as follows: 1) $\sqrt[3]{0.4}$ (cube root of the non-occurrence of no civil unrest, meaning here "there will be mass civil unrest") = 0.7368, 2) $(0.7368)^{2.5}$ (probability of no civil unrest for the two and a half year period) = 0.47, and 3) 1-0.47 (adjusted probability for the two and a half period) = 0.53. Taking into account the medium granularity, the European Values say that the probability of civil mass unrest in France in the two and a half year period (until the middle of 2019) is 55%. Of course, this procedure is not used for events and corresponding statements where the think-tank set a specific period or year in advance. Moreover, probability adjustment cannot be performed in statements focusing on processes. In these cases, the probability remains always the same, regardless of whether the particular process – in reality – has been seen since 2017 or since, for example, 2018.

Table 7: Summary of thirteen voided questions (eliminated from accuracy assessment) and one split-question

Statement; Assessed probability	Decision	Substantiation
AfD will get 20% (Elections to the Bundestag); 30%	Voided	Mathematical and logical error – it goes against the statement in the text
The possibility of a dramatic escalation of relations between the two strongest European countries arises if Marine Le Pen wins the French election (German-French relations); 30%	Voided	Conditional probability – condition was not fulfilled (Marine Le Pen did not win)
(If Marine Le Pen wins the presidential election)...*it might come to a revision of the Treaties of the European Union*; 30%	Voided	Conditional probability – condition was not fulfilled (Marine Le Pen did not win)
(Noticeable economic downturn in Russia)... *will not lead to a significant political disintegration/or to increasing separatism, but rather to tightening the regime's grip on the country*; 20%	Voided	Conditional probability – condition was not fulfilled (there was no significant economic downturn)
The idea of an Islamic caliphate will survive after ISIS ceases to exist; 80%	Voided	Conditional probability – condition was not fulfilled (ISIS still exists)
(year-to-year declines in GDP growth in the Eurozone)... *will have minimal political implications*; 30%	Voided	Conditional probability – condition was not fulfilled (there were no year-to-year declines)
(year-to-year declines in GDP growth in the Eurozone)... *will lead to the fall of European governments*; 30%	Voided	Conditional probability – condition was not fulfilled (there were no year-to-year declines)
(Completion of the Temelin Nuclear Power Plant)…*construction may lead to a similar situation as in the case of the Hungarian Paks NPP* (Russia´s involvement and Russia´s effort to deepen the political influence); 60%	voided	The tender was postponed until 2020/2021
In case of territorial gains outside Syria and Iraq, the Islamic State will either expand its influence (60%) or will continue to create smaller networks across Europe (30%)	Voided	1. unstated judgment with 10% probability 2. two events are not mutually exclusive (and both happened)
(Areas from which migrants are coming to the EU)... *the current situation in Syria will be preserved*; 60%	Voided	Poor operationalization (extremely high abstraction) – too many aspects can construct "current situation"
(Areas from which migrants are coming to the EU)...*there will be a rapid deterioration in Africa*; 80%	Voided	Poor operationalization (extremely high abstraction) – too many aspects can construct "current situation"
The percentage of mainstream parties in electoral preferences... will fall below 50% in a significant part of the European Union; 80%	voided	Poor operationalization (absence of a specific threshold) – what does "significant part" mean? (4, 10 states, or V4 states?)
Despite the increase of sympathy for the radical political groups, they will suffer from internal conflict and lack of consolidation; 70%	Voided	1. follow-up question to the previous one 2. Poor operationalization – internal consolidation, or consolidation between radical parties?
(Reaction of the European Commission to the migration from third countries; one "binary event")...*There will be slow effort to implement Turkish model with other countries (70%), or there will be an effort to build asylum facilities outside of Europe (30%)*	split into two separate statements	It prevents another elimination – one new question (statement) emerges

source: own elaboration (based on the data in Appendix A)

The last adjustment relates to the only proper four-option-statement in the whole research project, but this time applied to the Brier score calculation. The Brier score formula presented in the methodological chapter cannot sufficiently operate with four options. Nonetheless, the European Values assigned four same (25%) probabilities to each possible

outcome of the Presidential election and Midterms in the US. Due to this fact, the question is split into two statements and resulting Brier scores are subsequently averaged.

From the perspective of unstated judgments in three-option-questions, these have no impact on Brier score assessment, no matter if the think-tank omits to delineate one (e.g. the United States – Poland relationship, see in the previous chapter) or two missing futures (e.g. the result of French presidential and parliamentary elections, see also in the previous chapter). The reason for this is that one of the two clearly delineated futures has always occurred so we do not need to be preoccupied with the unstated one. Bearing this in mind, the results of the European Values´ research project are the following.

5.2. The Results of the European Values´ Research Project

Based on the assessment and results of 46 questions illustrated in Table 8 (see below; more details in Appendix A, Sheet 1) the European Values succeeded, and thus correctly predicted occurrence or non-occurrence of the event or process in 31 of them. In other words, the think-tank failed to predict such occurrence or non-occurrence in 15 cases (including two "I don´t know" cases where the European Values assigned equal – 50% or 25% – probabilities to all selected futures). This number accounts for one-third of all forecasts (precisely 32.6%). Although such share in itself drops some hints about the predictive capabilities of the European Values (if they were right or wrong that something will or will not happen), it certainly says too little about their accuracy. These forecasts are probabilistic, so there are no uncertainty-excluded statements related to the "anti-probabilistic" thinking. Given the probabilistic nature of the research project´s forecasts, which in general emanates from the theoretical debate about uncertainty, the accuracy is therefore what we are really curious about. That is why the Brier score, as an instrument for performance (accuracy) analysis, finally comes into play to indicate how far away from the truth the European Values´ forecasts were. As a reminder, the Brier score can range from 0 to 2 (considered as "ideal types"), where the lowest possible score represents the perfection (bull´s eye) and the highest one means the perfect opposite of reality. This applies for particular Brier scores (of particular questions/statements) and overall average Brier score (of all questions) alike. Indisputably, it is worth examining some of the forecasts with very high or very low brier scores as well as general findings, which subsequently positively or negatively affects the overall result (average Brier score) of the research project.

Table 8: Results of the European Values´ forecasts

Forecasted Question	Probability Estimate	Time Frame	Occurrence of the Forecasted Outcome	Brier Score
Key players: Germany				
The CDU will form a coalition with the Greens	55%	2017	No, the CDU formed a coalition with the SPD.	0.605
AfD will get up to 10 percent of the votes	70%	2017	No, AfD won 12.6% of the vote.	0.980
There will be no major political or economic conflict between France and Germany	70%	process, but as an outcome of the election: perceived until the end of 2018	Yes, there was no conflict (at worst, there was lower consensus in some areas).	0.180
... cooling of relations between the V4 countries and Germany as a result of the lower political interest of Germany in Central Europe	40%	process, but as an outcome of the election: perceived until the end of 2018	Yes, there were (and still are) troubles, but no cooling from the German side.	0.320
If (cooling) does not occur, the relations between the countries will remain similar as the present ones (no fundamental improvement will occur)	40%	process, but as an outcome of the election: perceived until the end of 2018	Yes, there was no fundamental improvement.	0.320
Realisation of terrorist attacks by Islamists on German soil will lead to a reaction of the far-right groups (even in the form of violence	60%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, 1) there have been several terrorist attacks in Germany since the end of 2016 2) anger and violence exploded in Chemnitz (2018) after word spread that an Iraqi and a Syrian asylum seeker killed a German man.	0.320
There will be a systematic series of major terrorist attacks by the Islamists or the extreme right of the right-wing extremists	40%/adjusted probability is 35%	event - the year is not specified (2016-2019)	Yes, Germany has witnessed several attacks from both Islamists and right-wing extremists; however, these cannot be described as "systematic series".	0.245
The relationship between Germany and Russia will take a form of status quo maintenance	60%	2018	Yes, while Germany again engaged in dialogue with the Russian Federation (also, GER-RUS diplomatic activity increased; Germany granted the Nord Stream 2 permit), it still remained a consistent critic of Moscow. Any serious bilateral progress was curtailed by Germany's adherence to NATO and the EU.	0.240
Key players: France				
Marine Le Pen will not win the presidential elections	60%	2017	Yes, Marine Le Pen was defeated by Emmanuel Macron.	0.320
Neither the National Front (parliamentary election) nor M. Le Pen (presidential election) will win	60%	2017	Yes, both the National Front and Marine Le Pen did not win.	0.260

Mass civil unrest will occur, which will be instigated by the far-right against the Muslim minority (an outcome of following terrorist attacks)	60%/adjusted probability is 55%	event - the year is not specified (2016-2019)	No, there was no mass civil unrest (instigated by far-right) after the 2016 Nice attack, the 2017 Carcassonne or the 2018 Strasbourg attack. Paradoxically, the mass civil unrest came with the Yellow vests movement.	0.605
Key players: Russia				
Russian Federation's political leadership will behave more aggressively and offensively than in the present day (towards EaP states, Baltic states, and internal affairs of certain NATO members	60%	process since 2017; probability remains the same for all years/period	No, compared to 2016, Russia does not behave more aggressively. The disinformation campaign and other aspects related to sharp power have been used by Russia long before 2016. The same applies to the number and targets of bigger cyber attacks (long-term tactic). Furthermore, the number of intercepts of Russian aircraft in the Baltic Sea Region in 2017 and 2018 was lower than in 2015.	1.040
Russia would still strive for political subversion in the Baltics and specific EU states, such as the Czech Republic	70%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, the Russian Federation uses subversive measures (e.g. support of populist parties or disinformation campaign) to destabilise political systems/public and to influence the cohesion of EU	0.180
Russia will resort to an effective occupation (even in case of using the nuclear threat) of some of the Baltic states	30%/adjusted probability is 25%	event - the year is not specified (2016-2019)	No, there was obviously no effective occupation.	0.125
In the following three years, there will probably be a noticeable economic downturn in Russia	70%/adjusted probability is 65%	event - the year is not specified (2016-2019)	No, after more than 2 years of recession, the Russian Federation returned to growth in 2017. Although there was a decline in the real GDP growth from July 2017 to January 2018, the growth as such was still positive. The same goes for 2 quarters in 2019.	0.845
Key players: United States				
Hillary Clinton's chances of winning are slightly higher than those of Donald Trump	55%	2016	No, Donald Trump won the presidential election.	0.605
There will be no dramatic increase in the presence of US troops in the Middle East	80%	process (2016-2019), the probability remains the same for all years/period	Yes, there was no "official" dramatic increase in the presence of US troops.	0.080
The victory of Donald Trump could dramatically endanger the existence or relevance of transatlantic structures	30%	conditional process (condition has been fulfilled (Trump became a president)	Yes, Conditional probability: 1) Donald Trump won the presidential elections (condition is fulfilled) 2. Donald Trump criticised allies (broadsides). Leaving aside whether these attacks act as a part of his discourse or really endangers the US remaining in the alliance, Trump´s behaviour - according to some experts - strengthens NATO (increase in spending of European allies, their defensive stance for NATO). Furthermore, Trump Administration is spending far more on NATO than the Obama Administration.	0.180
There will be either a rejection or a significant postponement of the TTIP treaty	60%	process (2016-2019, but the occurrence/non-occurrence can be already detected	Yes, Donald Trump cut off - after taking the office - trade talks with the European Union, and the deal-negotiations were postponed until 2018 (regardless of whether or not Trump considers negotiation reopening).	0.320

US involvement in Central Europe within NATO will be the same	70%	process (2016-2019), the probability remains the same for all years/period	Yes, despite the fact that there is a difference between Obama´s and Trump´s behaviour towards - for instance - Germany and Poland, the "involvement" (overall) is not greater or lesser.	0.135
Midterms and Presidential elections – results	One candidate wins both = 50% One candidate wins only in one of the two elections = 50%	2016 and 2018	In order to calculate the Brier score, these statements are divided into two events and then averaged. Due to the fact that the European Values gave equal probabilities to each event, result as such is "unimportant" (in fact, Donald Trump won the presidential election, Republicans won in the Senate, Democrats in the House of Representatives and in the gubernatorial elections.	0.500
The atmosphere of the congressional elections will be affected by the rhetoric presented by Donald Trump	60%	process, from 2016/ 2017 to 2018	Yes, congressional elections were, indisputably, affected by Trump´s rhetoric (e.g. anti-immigrant rhetoric).	0.320
Key players: Poland				
An undermining of the liberal character of the Polish political system is more likely to happen	50%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, the Law and Justice party (Pis) has been weakening Polish democracy since 2015. The latest assault is the PiS effort to bring the Supreme Court under its control.	0.420
Poland will not implement migration quotas, and will also not try to legitimize its stance by negotiating an opt-out	60%	2017	Yes, Poland rejected migration quotas. In addition, Poland did not try to legitimize its stance by the opt-out mechanism.	0.320
Regarding Poland's stance towards the United States, the current state - characterized by cooperation in the field of defence - will be preserved	50%	process (2016-2019), the probability remains the same for all years/period	Yes, although we can see Donald Trump´s rhetorical support of Poland or - for instance - the new "United States–Poland Strategic Partnership", it is always related to the defence cooperation.	0.420
There will be an escalation of rhetorical attacks between Poland and Germany	70%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, due to the different political trajectories (e.g. PiS populism) conflicting views on NATO´s and EU´s future development, there is a crisis in Polish-German relationship. This crisis is accompanied by many rhetorical attacks (e.g. Second World War reparations, Warsaw´s criticism of Berlin for its dominance of the EU, etc.).	0.180
Key players: European Commission				
In the field of integration, the Commission will press for its deepening to the same degree as today. As for asylum policy, the Commission will promote such a model of asylum system, which will place lesser demands on states. The redistribution of asylum seekers will take place, but probably only on a symbolic scale	60%	process (2016-2019), the probability remains the same for all years/period	No, the most problematic statement concerns the "symbolic scale" of redistribution. On 31 May 2018, the European Commission reported that 34 689 asylum seekers (35%) were redistributed. This cannot be perceived as a symbolic scale.	0.720

The EC will abandon of a uniform system implementation	35%	2017	No, the EC proposed two reforms of the asylum system and procedures, but both of them state the system should be truly "uniform".	0.735
The actions of Hungary and Poland which undermines liberal democracy, will continue to be tolerated and only symbolically criticized by the Commission (more radical measures will not be taken)	60%	process (2016-2019), but the occurrence/non-occurrence can be already detected	No, following Poland's law on the Supreme Court (which entered into force on 3 April 2018), the European Commission launched an infringement procedure and took Poland to the European Court of Justice (for the first time). Despite efforts of Warsaw to drop some controversial provisions, the Commission says Poland has failed to respond in an adequate way. This is certainly not a "symbolic" measure.	0.720
Regarding the migration from third countries - there will be a slow effort of the Commission to implement Turkish model with other countries	70%	2019	Yes, the Turkish model can be perceived as a strict version of the readmission agreement. The European Commission has proposed - in its documents - readmission agreements with Libya, Egypt and other Northern Africa countries (e.g. In 2017 and 2018).	0.180
Regarding the migration from third countries – there will be a faster and tougher approach; an effort to build asylum facilities outside of Europe	30%	2019	No, the Commission, EU leaders (individually) and European Council President Donald Tusk have indeed supported (in two 2018 documents) the development of the concept of regional disembarkation platforms ("hotspots" outside the EU).	0.980
Key players: Islamic State				
Islamic State will not be able to maintain a stable territory in the area of Syria and Iraq	60%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, since 2017, the Islamic State has suffered major defeats, thus lost a huge part of its territory. In March 2019, US-backed forces said the Islamic State group lost the final significant part of its territory. Nevertheless, the Islamic State still holds some minor territorial positions.	0.320
There will be an increase in the activities of ISIS (or rather its successors) in Europe	80%	process, from 2015 to 2018 (due to the data accessibility), the probability remains the same for all years/period	Yes, According to Europol, there has been an increase in several activities of IS and mainly of Al-Qaeda (resurgence in Europe). In terms of completed, foiled and fail attacks, the number has grown between 2015-2016 and 2017-2018 period (from 15/year to 28/year). Europol has also observed a general increase in CBRN propaganda, tutorials, and threats.	0.080
Systemic threats: Islamic extremism in Europe				
There will be growth in the number of terrorist attacks is more probable; caused both by Islamists and as well as by the far-right as a reaction	65%	process, from 2015 to 2018 (due to the data accessibility), the probability remains the same for all year/period	Yes, taking into account Europol´s average number of completed, foiled and failed Islamists and far-right attacks, the number of attacks has grown between 2015-2016 (40 attacks, which means 20 attacks/year) and 2017-2018 period (63 attacks, which means more than 31 attacks/year). The number of far-right attacks has been decreasing; however, it does not affect the overall tendency.	0.245
Between 2016 and 2019, the Czech Republic will be struck by the first terrorist attack caused by Islamist extremists	30%/adjusted probability is 25%	event - the year is not specified (2016-2019)	Yes, the Czech Republic was not struck by the first terrorist - Islamist - attack. Conversely, there was one terrorist attack (in 2017) conducted by the supporter of "right-wing" populist party.	0.125

Islamist extremists will formally engage themselves in politics only to a limited extend	70%	process (2016-2019), the probability remains the same for all years/period	Yes, there has been no indication of an increase in formal engagement of Islamist extremists in European politics.	0.180
There will be deepening radicalisation across the Muslim community, which will lead to violence	70%	process (2016-2019), the probability remains the same for all years/period	No, as many scholars claim, research on radicalisation across the Muslim community in Europe is scarce and empirical-based understanding is weak. This compels us to focus on the "will lead to violence" judgment. In this regard, we cannot confirm there has been substantial evidence of a perceptible wave of Muslim violence or its increase (only individual and local cases). Conversely, in the last two years, there have been reports on the intensification of anti-Muslim violence in some European countries.	0.980
Systemic threats: Mass migration				
The amount of migrants and asylum seekers in the European Union will not significantly decrease from year-to-year	60%	process (2016-2019), but the occurrence/non-occurrence can be already detected	No, the number of asylum seekers has significantly decreased since 2016 (from approximately 1 260 000 applications in 2016 to 712 000 in 2017 and 630 000 in 2018). The same tendency can be seen in the number of "migrants".	0.720
Regarding Turkey's cooperation on solving the migrant crisis, the situation will worsen with negative consequences for the EU	60%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, based on the statements of the European politicians, the deal is dysfunctional. The number of migrants reaching the European Union from Turkey rose in 2018.	0.320
Turkey will continue in its authoritative tendencies and in the weakening of liberal democratic principles (Islamist and dictatorial elements)	70%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, after the Constitutional referendum in 2017 (granting the president new powers) and 2018 victory of Recep Tayyip Erdoğan, he continues to bolster mainly the electoral authoritarianism, neopatrimonialism, and Islamism (as political ideology).	0.180
Systemic threats: Authoritarianism				
Republicans will come to realize that they have to take into account the interests of the lower middle class. They will focus their efforts on more conservative policy. However, the fact that the party is torn apart will remain a problem.	70%	process (2016-2019), but the occurrence/non-occurrence can be already detected	No; this statement can be rejected on the basis of the second judgment: Donald Trump has been able to consolidate the Party and with the re-election campaign, "anti-Trump" block of Republicans has shrunk.	0.980
In Poland, the support or at least the tolerance of policies of PiS will result in the country's political system becoming more authoritarian	60%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, the Law and Justice party (Pis) has been weakening Polish democracy since 2015. The support of PiS has been gradually declining in the last two years, but still, the PiS has - according to the polls - more than 35%. In the European election, the PiS had more than 45% (turnout was 46%)	0.260
In Hungary, Viktor Orbán will be able to radically deepen its authoritative practices (neither an improvement nor a deterioration of authoritative practices)	40%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, Victor Orbán has been able to deepen its authoritative practices. However, Hungary represents a fully developed hybrid regime where a readjustment of the national legislative framework would require a constitutional supermajority (the regime act and deepen the practices in a more gradual and procedurally "right" way). In this case, there has been no "radical" deepening.	0.320

Systemic threats: Extremist attitudes in society				
The Czech Republic and the nearby countries will experience a growth of anti-system parties, which will get into parliaments and governments	50%	specified events (based on the year of Parliamentary elections in nearby countries)	Yes, there was a growth, albeit minimal one, of anti-system parties (PILZ in Austria, We Are Family in Slovakia, FDP - according to the European Values - in Germany and eventually also Pirates in the Czech Republic.	0.500
Systemic threats: Cyber threats				
There will be unexpected attacks on cyber security in the upcoming years	80%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, there have been many - and logically unexpected - attacks on cybersecurity in 2018 or 2019.	0.080
Systemic threats: Energy Threats				
In the upcoming years, we expect that the construction of the gas pipeline Nord Stream II is more likely	70%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, Nord Stream II construction has begun in summer 2018.	0.180
source: own elaboration (based on the data in Appendix A). The references and links are set out in Appendix A. Forecasted questions are usually shortened and do not include the introductory part (and even the normative part). However, questions thus also contain grammatical errors made by the think-tank.	**Average Brier score**			0.410

Since poorly operationalized questions receive special attention in both previous pages and the previous chapter, it seems reasonable to start with them. The forecast concerning the United States military presence in the Middle East is among the top three most – in terms of Brier score – accurate estimates. The European Values estimated that there was an 80% probability of no dramatic increase in the presence of United States troops in the Middle East. At first sight, it might seem like a poor forecast. Based on the figures released by the US Department of Defense in 2017, the number of deployed troops to the Middle East surged by 30% between June and September (from approximately 40 000 to 54 000).[139] Regardless of the unstated threshold for "dramatic increase", such growth would probably entail rejection of the event´s occurrence (because 30% should be perceived as dramatic increase) and thus also poor Brier score. However, fortunately for the European Values, a Pentagon spokesman Eric Pahon – later that year – stated that the Department of Defense data was not accurate and mistakenly operated with quarterly rotations in the region. Moreover, since 2018, the Pentagon has stopped releasing information about the deployment to several Middle Eastern countries.[140] The last information about the troops´ deployment in mid-2019 said – besides planned gradual withdrawal from Syria – that the Department of Defense had approved to deploy around 1500 troops to the Middle East (United States-Iran tensions).[141] This would not be a "dramatic increase". As a result, the judgment was correct (officially no noticeable increase) and the 80% probability thus leads to an outstanding Brier score of 0.08.

Prediction about the ISIS activities in Europe represents another poorly operationalized but also close-to-bull´s-eye judgment. The European Values estimated with 80% that there "*will be an increase in the activities of ISIS (or rather its successors) in Europe*".[142] The undefined concept of "activities" gives us a lot of manoeuvrability in deciding about the occurrence or non-occurrence of this process. But the main problem is lacking data for 2019. Therefore, the only suitable and fair approach lies in the comparison between 2015-2016 and 2017-2018 periods. According to Europol reports, there was an increase in "activities" between these two periods. In terms of foiled, failed and completed

[139] ZENKO, Micah. US military policy in the Middle East: an appraisal. Chatham House, 2018. p. 11.
[140] DETSCH, Jack. Pentagon reports troop surge in Middle East. In: Al-Monitor [online]. [cit. 2019-06-28]. Available from: https://www.al-monitor.com/pulse/fa/originals/2017/11/pentagon-troop-surge-syria-middle-east.html
[141] SONNE, Paul and Missy RYAN. Trump approves sending more forces to the Middle East amid tensions with Iran. In: The Washington Post [online]. [cit. 2019-07-02]. Available from: https://www.washingtonpost.com/world/national-security/trump-approves-sending-more-forces-to-the-middle-east-amid-tensions-with-iran/2019/05/24/7403ffac-7e2c-11e9-a66c-d36e482aa873_story.html
[142] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 17. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

jihadist attacks, the 2017-2018 periods experienced an increase from 30 attacks (between 2015 and 2016) to 57 attacks. In other words, this was an increase from 15 attacks per year (2015-2016) to 28 attacks per year in Europe. It is problematic to provide direct links of aforementioned attacks to ISIS (or other groups), but Europol goes on to state that even home-grown terrorists were radicalized or supported through various ISIS and also Al-Qaeda networks. In that sense, the increase in activities of – in Europe – resurgent Al-Qaeda could be seen as a "successor" here. Europol also claims there was an increase of CBRN (chemical, biological, radiological and nuclear) threats, propaganda and tutorials in 2016-2017 periods[143], as well as growing use – by ISIS and Al-Qaeda – of online criminal markets systems (e.g. Hawala transfer system) for financing and supporting the activity. Lastly, ISIS increased its activities in recruitment and use of women in order to mobilize other females on European soil.[144] This gives us all good reason to accept the occurrence of the process, which again means a very low Brier score of 0.08.

Conversely, accuracy assessment of poorly operationalized statements also yields several disappointing Brier scores. One of them refers to the dynamics of Russia´s economy. The think-tank forecasted, with 70% probability, that the Russian Federation would experience a noticeable economic downturn. This event is precisely the case of required adjustment, where the probability should be adjusted to the two and a half year period. Based on the aforementioned formula, the probability of a noticeable economic downturn in Russia for our period is 65 percent. If we adopt the definition mentioned in the previous chapter and thus describe the economic downturn as non-positive growth of real GDP per capita, there hardly was an economic downturn. But if we classify slowdown as an economic downturn, raw data would prove the European Values right. Russia´s economy underwent such slowdown in two consecutive quarters. Whereas the first downturn was almost imperceptible (from GDP growth of 2.325% in July 2017 to the growth of 2.200% in October 2017), the second one was obvious (from GDP growth of 2.200% in October 2017 to the growth of 0.595% in January 2018). Since then, Russia´s economy has recovered and climbed to a six-year high in 2018. The growth again decreased in the first two quarters of 2019.[145] Nevertheless, it is reasonable to assume, in view of the period of think-tank´s forecast making, that the European Values believed – by adding "noticeable – in some more serious dynamics. The forecast was made at the end of 2016. By this time, Russia was still affected by the

[143] EUROPOL, TESAT. European Union Terrorism Situation and Trend Report 2019. 2019. p. 9-13, p. 19 and p.35.
[144] EUROPOL, TESAT. European Union Terrorism Situation and Trend Report 2018. 2018. p. 15 and p.33.
[145] Russia Real GDP Growth: 1996 - 2018 (Quarterly; %). In: CEIC [online]. [cit. 2019-06-23]. Available from: https://www.ceicdata.com/en/indicator/russia/real-gdp-growth

recession, serious financial crisis, or depressed domestic demand. Moreover, prospects for improvement were not promising.[146] That is why we should think of "noticeable downturn" as something more severe, and not as two consecutive slowdowns in GDP growth. Yet as a matter of the fact, Russia did not experience a recession in 2017, 2018, or 2019. Thus, the judgment was not correct, and the resulting Brier score is poor 0.845.

Subsequently, the European Values underestimated behaviour of the EU states both during and after the peak of the "refugee crisis". The long statement "*In the field of integration, the Commission will press for its deepening to the same degree as today. As for asylum policy, the Commission will promote such a model of asylum system, which will place lesser demands on states, and which is currently being discussed. It is possible that the redistribution of asylum seekers will take place, but probably only on a symbolic scale and between a group of likeminded states*"[147] is by its very nature problematic. The statement consists of many elements, and the concept of "symbolic scale" is – again – left without a specific threshold. Leaving aside other partial estimates (e.g. promotion of asylum system with lesser demands on states is also questionable), the forecast as such should be seen as mistaken mainly because of the estimate about the "symbolic scale of redistribution". In 2015, the European Commission proposed mandatory relocation of 40 000 asylum seekers. However, in the midst of the crisis, the Council adopted a decision to relocate an additional 120 000 people from Italy and Greece (the total number was later revised to 98 000). The decision has led to the well-known dispute between Visegrád Group countries and the European Commission.[148] Although the distribution was very slow in the first phase, states did, in the end, redistributed more than 34 000 asylum seekers by May 2018. That being said, participating states reached 35% of commitments enshrined in the Council decisions.[149] Hence, the redistribution of 34 000 asylum seekers was by no means symbolic. Given the fact that the European Values said there is only a 30% probability of something more than just a symbolic redistribution, the resulting Brier score is 0.72.

[146] MANKOFF, Jeffrey. The Russian economic crisis. Council on Foreign Relations, 2010. p. 4.
[147] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 17. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf
[148] ŠABIĆ, Senada Šelo. The relocation of refugees in the European Union. The Friedrich-Ebert-Stiftung , 2017 p. 5.
[149] Relocation of asylum seekers from Italy and Greece: Datas. In: European Migration Law [online]. [cit. 2019-06-27]. Available from: http://www.europeanmigrationlaw.eu/en/articles/datas/relocation-from-italy-and-greece.html?fbclid=IwAR2kQYPzhsW_8RoUJa47yELox89nbyNdBiCx2W96drmmg0o5BsFL3U2nSlE

Turning to overall findings not exclusively related to the poorly operationalized statements, the best score achieved by the European Values is 0.08. Apart from the above-mentioned predictions about the United States presence in the Middle East and the activity of ISIS in Europe, the last top-three-score forecast concerns cybersecurity. The think-tank confidently predicted (with 80% probability) that "*there will be unexpected attacks on cyber security in the upcoming years*".[150] While this statement lacks any geographical specification, its nature still enables accuracy evaluation. The evidence from the *Center for Strategic and International Studies* database substantiates this statement. Since the development of the European Value´ research project, there have been lots of unexpected attacks on cybersecurity in many states.[151] But this issue has been ongoing for more than three years. This is a safe bet however, similar to saying the space debris will continue to cause problems for space agencies. Such estimates guarantee good results, rather than express predictive capability.

The second best (lowest) score, achieved on two forecasts with adjusted probability for our examined time period, is – still an impressive – Brier score of 0.125 (again, see table 7). But one these forecasts – estimating a 30% (25% after the adjustment) probability of the Islamist terrorist attack in the Czech Republic – is definitely more intriguing. On the one hand, such statement (and especially its probability) may seem too alarmist, on the other, the Czech Republic was indeed struck by a terrorist attack. However, this attack was not conducted by an Islamist extremist, but a home-grown supporter of the anti-Islamic populist party.[152]

From a broader perspective, the European Values were successful in predicting the occurrence of events and processes related to Poland. They correctly predicted undermining of the liberal character of the Polish political system, Poland´s rejection of migration quotas, preservation of the current relationship with the United States and finally the rhetorical escalation between Poland and Germany. However, the frequent tendency towards extreme-foxes behaviour, and hence the inability to stray out of the maybe-zone, results here in four not so impressive Brier scores ranging from 0.32 to 0.42. Thanks to the Brier score of 0.18 awarded for the Germany-Poland deterioration forecast, the average Brier score for the whole section is 0.335.

---

[150] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 22. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

[151] Significant Cyber Incidents. In: Center for Strategic and International Studies [online]. [cit. 2019-06-26]. Available from: https://www.csis.org/programs/technology-policy-program/significant-cyber-incidents

[152] Vrchní soud potvrdil důchodci Baldovi čtyři roky za teroristický útok na vlaky. In: iRozhlas [online]. [cit. 2019-06-28]. Available from: https://www.irozhlas.cz/zpravy-domov/balda-senior-teroristicky-utok-utok-na-vlak_1904161114_nkr

The example of forecasts related to Germany and subsequently to the United States demonstrates even better how the Brier score punishes you for being less decisive when it is desirable, regardless of how many occurrences or non-occurrences you correctly predicted. In the section dealing with Germany, the European Values successfully predicted the occurrence of 6 out of 8 events or processes. However, the resulting average Brier score is – not so spectacular – 0.401. In the case of the United States, they hit 5 out of 7 occurrences (worse than in the case of Germany), yet the think-tank finishes, in this section, with quite a nice Brier score of 0.303. This is guided by the fact that the six correctly predicted occurrences in the "Germany" section include five poor-decisive estimates, whereas three of those five correct estimates predicting the United States dynamics are highly resolute. They were much closer to the unfolded reality.

More generally, the European Values were quite unsuccessful in forecasting the results of the elections. They failed to predict "future" coalition in Germany (Brier score of 0.605) or the winner of the 2016 United States presidential race (but they were not alone, see the next chapter). Underestimation of Alternative for Germany has generated poor Brier score of 0.98 (on the whole, there are four Brier scores of 0.98). This is the second worst achieved score, and the forecast regarding internal dynamics of the Republican Party is another example that generates such poor result. In this statement, the European Values forecasted with 70% probability that "*Republicans will come to realize that they have to take into account the interests of the lower middle class. They will focus their efforts on more conservative policy and on changing the image of liberal democracy more towards conservatism. However, the fact that the party is torn apart will remain a problem*".[153] The rightness of such a forecast has to be rejected due to the occurrence of the opposite developments. Although the Republican Party truly was divided at the time of think-tank´s forecast-making, this has been slowly changing. Donald Trump has been able to consolidate the party and enlisted key local and regional Republican figures. Moreover, Republican critics of Trump in Congress have not been re-elected or have stepped down, and the only true block of "anti-Trumpers" (e.g. John Kasich) is now the smallest one at the national level.[154] In light of this, Donald Trump has the

[153] European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. p. 20. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

[154] BACON JR., Perry. The Five Wings Of The Republican Party. In: FiveThirtyEight [online]. [cit. 2019-06-26]. Available from: https://fivethirtyeight.com/features/the-five-wings-of-the-republican-party/

loyalty of Party representatives and the Party as such is – now before his re-election campaign – pretty much united.[155]

Nevertheless, the think-tank´s overall Brier score took the biggest hit in the forecast regarding Russia´s behaviour towards Baltic, Eastern Partnership or NATO states. In this three-possible-outcomes question, the European Values confidently stated that Russia would behave more aggressively and offensively than in 2016 ("present-day" at that time). Taking into consideration then-time image of the Russian Federation (2016), the Kremlin – or supported insurgents – was soon after the biggest battles in Eastern Ukraine, and was involved (for one year) in the Syrian Civil War. As a result of such destabilising actions, NATO top actors warned against provocative military activities and expected more offensive behaviour towards the aforementioned partners.[156] Since then, there have been several noticeable actions related to Russia´s behaviour: the poisoning of Sergei Scripal, Kerch Strait incident, cyber attacks, and continuing interceptions – predominantly – in the Baltic Sea Region. However, there is no robust indication of more assertiveness in Russia´s behaviour. Whereas the poisoning of Sergei Scripal was not – unfortunately – the first case of such Russia´s action on the European soil, the seizure of three Ukrainian ships in the Sea of Azov should be seen in the light of continuing tensions between the Kremlin and Kyiv. Furthermore, Russia´s strategy of dismantling Western alliances through the use of disinformation campaigns, sharp power or non-linear warfare as such has been encapsulated already in 2013 (sometimes labelled as the "Gerasimov Doctrine").[157] It is true that Russia´s military exercise such as Zapad or Tsentr break new records for complexity and size, but again, the every-year increase is not a phenomenon of the last two and a half years. From the perspective of Russian aircraft intercepts in the Baltic Sea Region, the numbers were almost the same from 2014 to 2018 (e.g. there were 130 intercepts in 2017, the record was 160 intercepts in 2015; there are no reliable data yet for the year 2019). The only substantial increase in the Baltic Sea Region has been recorded in the number of Russia´s espionage activities.[158] Lastly, Russia has been conducting cyber attacks since 2007 and has also been able to execute larger hacks in 2018, just as in previous years.[159] Absence of empirical evidence for Russia´s more aggressive and

[155] BURNS, Alexander a Jonathan MARTIN. Trump's Takeover of the Republican Party Is Almost Complete. In: The New York Times [online]. [cit. 2019-06-26]. Available from: https://www.nytimes.com/2019/04/03/us/politics/trump-republican-party.html

[156] Relations with Russia. In: North Atlantic Treaty Organization [online]. [cit. 2019-07-03]. Available from: https://www.nato.int/cps/en/natolive/topics_50090.htm

[157] KIRCHICK, James. Russia's plot against the West. In: Politico [online]. [cit. 2019-07-03]. Available from: https://www.politico.eu/article/russia-plot-against-the-west-vladimir-putin-donald-trump-europe/

[158] Russian Activity Tracker: Nordic-Baltic Region. In: The Stockholm Free World Forum [online]. [cit. 2019-07-03]. Available from: https://frivarld.se/russian-activity-tracker-nordic-baltic-region/

[159] SHUYA, Mason. Russian Cyber Aggression and the New Cold War. Journal of Strategic Security, 2018, 11.1: p. 4.

offensive behaviour along with assigned 60% probability, therefore, entails the poorest – across the forecasts – Brier score of 1.04.

All of the above mentioned Brier scores – impressive or poorest, as well as those resulting from poorly operationalized forecasts – significantly affect the overall result of the European Values´ accuracy assessment.

Regarding the first research question, and as Table 8 illustrates, the average Brier score for all 46 forecasts is 0.410. Without including nine forecasts that were based on eleven estimates of invited experts, the average Brier score would be 0.390. This is caused by the fact that from those nine forecasts, the occurrence or non-occurrence of event/process was correctly predicted only in three cases. However, the think-tank voluntarily chose to include those nine experts´ forecast into their research project. To reiterate: 0 means perfection, 0.5 is what we would from get from estimates randomly generated in forecasting programs, and 2 stands for an absolute error, that is, the perfect opposite of reality.

As we can clearly see, predictive (accuracy) capabilities of the European Values are neither excellent nor disastrous. The think-tank´s Brier score of 0.410 does not lean towards either side of accuracy (perfect accuracy or great error). Furthermore, they – overall – are very close to random guessing (0.5). Although the think-tank proves slightly better in accuracy than the Brier scores achieved from coin flipping, this result signals – following the analysis in the previous chapter – extremely foxy behaviour.

Nonetheless, accuracy alone does not provide an overall picture. Even such a mediocre accuracy, if better than the accuracy achieved by alternative institutions and forecasting tools, can mean success. A broader perspective is necessary. Another evaluation (and also part of the rigor) therefore lies in comparing the European Values´ accuracy with the accuracy of alternative institutions, as well as with other means of forecast (accuracy) comparison.

6. The Third Level of Analysis: Accuracy Comparison

The proper accuracy comparison between the think-tank and other institutions (focusing on probabilistic forecasting) would ideally employ at least one alternative player with as many identical (or very similar) forecasted questions to those of European Values as possible. Conversely, comparing the European Values with a plethora of alternative institution in only one forecasted question would not make sense, since the Brier score based on one question would tell us almost nothing about whether the think-tank´s project performs better or worse.

However, the ideal proper accuracy comparison is impeded by several factors. On the part of the European Values, their forecasts rarely share the interest – with alternative actors – in predicting same, or at least similar, events or processes. Typically, a lot of institutions have attempted to predict the elections results too, though not so many have been interested in estimating the probability of political engagement of Islamist extremists or cooling of the relationship between Berlin and the Visegrád Group. On the part of the alternative institutions – and mainly the prediction markets (e.g. Hypermind, PredictWise or PredictIt) –, the main problem stems from the fact that these actors seldom allows searching for previous forecasts (i.e. forecasts for events/processes that have already taken place). Accordingly, only the Tetlock´s Good Judgment enables us to compare at least seven of its estimates, and hence the accuracy, with the European Values´ results. Although this provides us with valuable information about the think-tank´s accuracy, an analysis of only seven forecasted questions (out of 46) would certainly be unsatisfactory.

Nevertheless, there are two other means enabling us to compare the think-tank´s accuracy in a more suitable way. Due to the fact that the invited-experts´ estimates were used only sparingly by the European Values, the experts open another opportunity to compare – at least to a certain extent – the European Values´ Brier score with alternative results. However, making full use of the all non-used experts´ questions is impeded by both the nature of the questionnaire-statements and time-consuming form of such a method. Therefore, the second level of the comparison analysis utilizes the aggregate of randomly selected experts-questions which are not used in the think-tank´s research project (see below). This step is analytically sufficient and also allows us to ascertain whether the think-tank research team´s accuracy beats the one of the invited experts.

Finally, the only viable option to compare the European Values accuracy "as a whole" is to use the randomly generated estimates (sometimes called randomly generated

guesses/forecasts). The lack of comparison-data represents a frequent phenomenon in the forecast verification. Even the biggest prediction markets or intelligence services supplied with the largest budgets face an actual shortage of sufficient amount of comparable alternative forecasts (e.g. from other institutions). Therefore, in order to check whether the accuracy of its forecasters, research projects, or forecasting systems can be regarded as successful or poor, institutions (or forecasting projects) compare its data with a simulation of randomly generated estimates (and its corresponding Brier scores).[160] To put it simply, probabilistic forecasts are – in terms of contribution – worthless if they are not sufficiently more accurate than randomly generated guesses (not to mention unnecessary economic costs of your forecasts in case they would not produce better results than a random simulation).[161]

Accordingly, this chapter examines the accuracy-comparison at three levels: the first part of the chapter compares the think-tank´s accuracy with the one of the invited experts, the second subchapter draws a comparison between the think-tank and Tetlock´s Good Judgment, and the last part compares the accuracy (Brier score) of randomly generated forecasts with the accuracy of the European Values´ research project.

6.1. The European Values vs. The Invited Experts

The data emanating from the questionnaire (sent to invited experts) represent both the easiest and simplest way to inspect whether or not the European Values beat the experts in the accuracy. Moreover, such a comparison does not necessarily require the inclusion of all invited-experts´ forecasts that were not used in the European Values´ research project. The aggregate of randomly selected forecasts is both sufficient and widely used tool for the comparison analysis. Another reason for using an aggregate for the comparison, and not all the data (statements in the questionnaire), is simply a rational one. As the previous chapter clearly shows, the amount of vagueness in many of the think-tank´s statements pose considerable problems for the evaluation. The same applies to the questions created for the questionnaire, and put bluntly, the process of evaluation is relatively time-consuming. An opportunity to use an aggregate of randomly selected forecasted questions is, therefore, the opportunity not to be missed. However, several issues need to be addressed before proceeding to the analysis.

[160] FERRO, Christopher AT. Comparing probabilistic forecasting systems with the Brier score. Weather and Forecasting, 2007, 22.5: p. 1085.
[161] LUCKNER, Stefan, et al. Prediction markets: Fundamentals, designs, and applications. Springer Science & Business Media, 2011 p. 82.

Firstly, compared to the research project, the forecasts in the questionnaire have no unified assigned probability, because at least eight (fourteen at maximum) experts assessed their individual estimate to a particular statement. Regarding the non-used questions (statements), the think-tank – quite logically – did not unite these individual probabilities. Nevertheless, the European Values itself provide us with a detailed guide to convert many individual estimates into one final probability. Based on the Delphi method (see Chapter 3 introducing the think-tank´s research project)[*], all we need to do is to find a median value of all probabilities assigned by experts to a particular question (statement). This is precisely what the European Values did (in most cases) with the expert´s estimates used in their research project.

Secondly, in order to properly get the randomly selected forecasts, those eleven statements, which were already used for nine questions in the research project, have to be eliminated from the original group of forecasts (from all 51 statements in the questionnaire). Otherwise, we risk that this brief analysis would not be a clear comparison between the European Values´ forecasts (including those nine experts-based forecasts voluntarily utilized by the think-tank) and forecasts of invited experts (those not used in the research project). Thus, the initial data, for now, consists of 40 statements.

Lastly, both the nature and focus of the probabilistic forecasts in the think-tank´s research project differs substantially from the experts´ forecasts in the questionnaire. Whereas in the research project, the European Values directly focus on the Czech Republic only in a sporadic manner, the questionnaire includes four sections directly addressing the Czech internal dynamics. Therefore, two sections (nine statements) are voided in order to both achieve better comparability and preclude a situation where all forecasts would deal with the Czech Republic. As a result, the initial data comprises the final 31 statements.

Due to a quite small amount of initial data, the sufficient number of forecast included in our aggregate lies between one half and one third. Thus, the accuracy of the European Values´ research project is compared with an aggregate of twelve randomly selected forecasts from the questionnaire. Table 9 illustrates the individual as well as the overall results of our sample (aggregate).

[*] As set out in Chapter 3 and despite the claim in their article, both the data from the questionnaire and the think-tank´s reference to the Delphi method suggest that the European Values utilized the median value, rather than the average value.

Table 9: Results of the twelve randomly selected forecasts (sample from the questionnaire)

Forecast	Probability based on the Median value of invited experts/adjusted probability	Time Frame	Occurrence of the Forecasted Outcome	Brier Score
Britain´s divorce from the European Union will be complicated, protracted and followed by disputes within both the EU and Britain	70%	process (2016-2019), but the occurrence/non-occurrence can be already detected	Yes, 1) two years after the United Kingdom invoked Article 50 and hence started the exit process, it had been due to leave on 29 March 2019. However, the UK Members of Parliament have been unable – three times – to accept the withdrawal agreement between the EU and the UK (two extensions followed). 2) There have been many disputes – primarily – not only between Tories and Labour but also within the Conservative Party of the EU, there have been disputes between leaders over – for instance - the length of delay offered to the UK.	0.180
The new EU treaty - with a stronger principle of "subsidiarity" and cooperation primarily in the security field - will be negotiated	35%/adjusted probability is 30%	event - the year is not specified (2016-2019)	Yes, no such treaty has been negotiated.	0.180
The Czech Republic will adopt the euro, or rather will set the target date for adoption	20%/adjusted probability is 15%	event - the year is not specified (2016-2019)	Yes, based on the last report regarding the Maastricht Convergence criteria (2018), the Ministry of Finance of the Czech Republic, together with the Czech National Bank, have recommended not to set a target date for adopting the euro yet. The Government of the Czech Republic has accepted this recommendation.	0.045
In the Czech Republic, the anti-European Union sentiment will deepen, potentially to the point of "Czechxit" referendum-date announcement	30%	process (2016-2019), the probability remains the same for all years/period	No, although the data from 2019 Eurobarometer shows that the majority of Czech respondents (51%, highest proportion among 27 states) have doubts when they think of the EU, other surveys indicate that the number of people satisfied with the EU membership is at the highest peak since 2011. The Czech citizens are neither to the EU-sceptics nor to EU-enthusiasts, but they rather demand a reform of the EU.	0.180
China´s ambitions will grow and the country will thus potentially create pressure on Russia	40%	process (2016-2019), but the occurrence/non-occurrence can be already detected	No, due to its long-term positive economic performance, China´s ambitions have been naturally (and gradually) growing. This may not necessarily be associated with assertiveness. China´s rise – and hence also ambitions – put the pressure on "declining" Russia´s power. Thus, for example, China has been – economically – penetrating into Central Asia (Russia´s sphere of influence); or has been exploiting Russia´s strategic assets – through disadvantageous contracts - in the Russian Far East. These dynamics could certainly be seen as a "pressure".	0.720

The next Czech governments will be pro-systemic and pro-European	60%	2017/2018	Yes, 1) despite the Prime Minister´s criticism of the EU (mainly due to the audit related to his problems with the EU subsidies) or his increasing accent on the Visegrád Group cooperation, the ANO-CSSD government is perceived as pro-European (the Prime Minister is a pro-reform figure, rather than EU-sceptic). Moreover, although the government is supported - in the Parliament - by anti-EU parties, the Minister of Foreign Affairs definitely is pro-European. 2) CSSD is not an anti-systemic party. Based on the Sartori´s narrow conception, the ANO is a protest party (movement) - calling only some part of the system (e.g. traditional parties) into question, not the system as such -, rather than anti-systemic party.	0.320
A pro-Western candidate will win the next Czech presidential election	50%	2018	No, in 2018, Milos Zeman was re-elected. Similarly as in his first term, President Zeman – and also other figures - has been gradually undermining existing pro-Western consensus in Czech foreign policy.	0.500
Brexit will lead to an economic recession in Britain and will weaken the EU economy	60%	process (2016-2019), but the occurrence/non-occurrence can be already detected	No, 1) although the UK´s Treasury several economists or even The Bank of England – before the Brexit referendum – stated the country could experience a profound shock to the economy and even the recession, the economy held up better than expected. Until the first quarter of 2019, the GDP growth, foreign investments or unemployment showed no signs of economic decline. Nowadays, the GDP growth has recovered and climbed back to 0.5%. 2) Brexit robs the EU of its second biggest economy. So far, there has been an economic slowdown in the European Union since 2018 (from GDP growth of 2.7% in 2018 to 1.5% in the first quarter of 2019). However, the European Commission claims the EU slowdown is mostly affected by a slowdown in global growth and world trade, or also a slowdown in China´s economy.	0.720
There will be year-to-year growth in GDP in the Czech Republic	65%	process (2016-2019), the probability remains the same for all years/period	Yes, there has been year-to-year growth in GDP in the Czech Republic: 2.5% in 2016, 4.4% in 2017, 2.9% in 2018 and 2.4% growth is predicted for 2019.	0.245
The Czech Republic will increase its defence spending and move closer to the pledge of 2% of the GDP	30%	process (2016-2019), but the occurrence/non-occurrence can be already detected	No, the Czech Republic has been increasing its defence spending, albeit very slowly. However, the Czech Republic has therefore been moving “closer” to the pledge of 2% of the GDP: 1.01% in 2016, 1.04% in 2017, 1.13% in 2018 and approved 1.19% in 2019.	0.980

The Czech Republic will experience a wave of violence or even terrorism instigated by the ultra-nationalist circles	25%/adjusted probability is 20%	event - the year is not specified (2016-2019)	Yes, 1) both the importance and influence of ultranationalist circle – as a whole – has been gradually diminishing. There have been some individual attacks during the last three years, however, this should not be perceived as a "wave of violence". 2) The Czech Republic experienced a terrorist attack (2017) conducted by a supporter of "anti-Muslim", nationalist and populist party. Based on some of his statements, he could potentially be labelled as an "ultranationalist". Nevertheless, this was not a "wave" of terrorist attacks.	0.080
The migrants and asylum-seekers from the Middle East and North Africa will - to a larger scale - start to settle down in the Czech Republic	40%	process (2016-2019), the probability remains the same for all years/period	Yes, data from The Ministry of the Interior of the Czech Republic have indicated no such dynamics. The Czech Republic serves as a transit country, and even the number of detained people (illegal migration) has been decreasing since 2015.	0.320
Average Brier score				0.373

source: own elaboration (based on the data in Appendix C). The references and links are set out in Appendix C. As a matter of interest, the average Brier score would be 0.390 if we use the "average" (not the median value) for the conversion of individual estimates into the one final probability. However, as we already know from Chapter 3, it is reasonable to believe the think-tank used the "median value" as a means of conversion.

Similarly to the research project, some of the statements crafted by the European Values would definitely deserve better operationalization. Leaving aside the results, one would hardly find what the concepts of "larger scale" (regarding the asylum-seekers) or "pressure" (regarding the Moscow-Beijing relations) supposed to mean. Nonetheless, as can be seen from Table 9, the experts – by "combining forces" – correctly predicted the occurrence or non-occurrence of events/processes in 8 out of 12 forecasts. The best (lowest) individual score is a great Brier score of 0.045. Furthermore, the experts – jointly – predicted with a very resolute 80% probability that the Czech Republic would not experience a wave of ultra-nationalist violence or terrorism. This estimate earns them the second-best Brier score of 0.080. On the other hand, the experts also delivered some blows to the overall Brier score. The biggest one, with a very poor Brier score of 0.980, arises from doubts about the ability of the Czech Republic (government) to increase its defence spending.

On the whole, the aggregate of 12 randomly selected forecasts finishes with the average Brier score of 0.373, which is slightly better (lower) accuracy than the one of the European Values achieved in their research project. Besides the higher share of correctly predicted events/processes, the higher resoluteness in the questionnaire plays a role in better result. In 6 out of those 8 correctly predicted futures, the questionnaire-forecasts were resolute and thus able to stray out of the "maybe" zone (40-60%). In other words, the findings from two previous chapters (compared to the invited experts, the use of poor-decisiveness

probabilities is significantly high in the case of the European Values´) do affect the accuracy of the European Values.

The European Values´ research project proved less accurate than the aggregate of experts´ forecasts which were not selected (and used) by the think-tank. But to be fair, the difference between the score of 0.410 (0.390 when excluding 9 forecasts based on experts´ estimates) and 0.373 is not staggering.

6.2. The European Values vs. The Good Judgment Forecasters

The Good Judgment (GJ), or rather its open platform GJ Open, has proven to be the only viable option enabling us to compare the European Values accuracy based on at least seven similar forecasted questions.

As mentioned in the theoretical chapter, the Good Judgment Project was and still is Tetlock´s (and his partner Barbara Mellers´) multi-year research program focusing on the feasibility of improving probabilistic forecasts. When the US intelligence community wanted to know how good they are in probabilistic forecasting – which is critical to the national security – the Intelligence Advanced Research Projects Activity (IARPA) created the first forecasting tournament (2011-2014) for five scientific teams.[162] This was a response of the US intelligence to the well-known debacle regarding the "Iraqi-WMD" decision. By demonstrating great accuracy and outperforming even the intelligence analysts, the Good Judgment Project emerged as an undisputed victor in these forecasting tournaments. What is more, the Tetlock´s team was composed of a few hundred ordinary people. Thus, he has also proven that with the involvement of rigor and cultivation of the forecasters (and even giving and extra weight to the estimates of top or super-forecasters), successful high-stakes predictions are achievable. In other words, Tetlock´s winning method was based on combining the judgments of a large group of people. This is called "the wisdom of the crowd".[163] The GJ, as a platform emanating from the Good Judgment Project, nowadays focuses on improving the forecasting capabilities of governments or corporations.[164]

In addition, the GJ Open subsequently serves as an open website platform where anyone can improve forecasting skills, or – based on her or his long-term accuracy – become a super-forecaster and hence a member of the GJ team. Since 2015, the GJ Open has been

[162] MELLERS, Barbara, et al. Identifying and cultivating superforecasters as a method of improving probabilistic predictions. Perspectives on Psychological Science, 2015, 10.3: p. 268.
[163] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 91.
[164] Good Judgment. In: Good Judgment [online]. [cit. 2019-07-13]. Available from: https://goodjudgment.com/

regularly and very professionally (in terms of operationalization, see in the analytical chapter) asking people to forecast the probability of particular events or processes. The data from these probabilistic forecasts have been utilized for instance by The Economist, BBC, or the Early Warning Project.[165] From the perspective of our analysis, each forecasted question – open for estimates for several months or even years – shows the data about the "Crowd Forecast" for every single day (from the period which was set for the estimates assignment). Similarly to Tetlock´s "wisdom of the crowd" which beat the US intelligence community, the "Crowd Forecast" expresses the average probability assigned by all forecasters (usually ranges from five hundred to several thousand people) to a particular question.[166]

Therefore, the probability based on the "Crowd Forecast" on the GJ Open website represents the data for the seven-question comparison between the European Values and the GJ Open. Thanks to the availability of the "Crowd Forecast" data, the selected alternative probability of the GJ Open is always as close as possible to the date of think-tank´s forecast making.* Table 10 (and Appendix C, Sheet 2) illustrates the comparison between the European Values accuracy and the one achieved by the GJ Open. Due to the think-tank´s major achievement in the first question concerning the US Presidential election, this question is complemented by other probabilistic forecasts of a few American institutions. As a matter of interest, some other alternative probabilistic forecasts are also shown in the question regarding the Presidential election in France. However, its Brier scores are not taken into account in the comparison as such, and the GJP Open remains the only alternative for the Brier score (accuracy) comparison.

[165] Good Jugment Open: Challenges. In: Good Judgment Open [online]. [cit. 2019-07-13]. Available from: https://www.gjopen.com/challenges

[166] Good Jugment Open: Frequently Asked Questions. In: Good Judgment Open [online]. [cit. 2019-07-13]. Available from: https://www.gjopen.com/faq

* The research project was published on 12 October 2016, and until this date, the think-tank still had an opportunity to update and thus adjust their forecasts.

Table 10: The seven-question comparison between the European Values and the GJP Crowd Forecast

Pairs of Forecasted Statements	Date of Forecast	Assigned Probability	Brier score
EV: Hillary Clinton will win the presidency	12 October 2016	55%	0.605
GJ: A Democrat will win the presidency**	1 October 2016	70%	0.980
NYT: Hillary Clinton will win the presidency	8 November 2016	85%	1.445
FTE: Hillary Clinton will win the presidency	1 October 2016	65%	0.845
HP: Hillary Clinton will win the presidency	3 October 2016	98%	1.921
EV: Afd will reach more than 10%	12 October 2016	30%	0.980
GJ: Afd will reach more than 10%	8 June 2017	29%	1.008
EV: Clinton/Democrats will win the Midterms	12 October 2016	50%	0.500
GJ: Democrats will gain control in both chambers	17 December 2017	50%	0.500
EV: Russia will effectively occupy a Baltic state	12 October 2016	11% (adjusted)	0.0240
GJ: NATO member will invoke Article 4 in response to actions taken by Russia before 1 January 2017	12 October 2016	1%	0.0002
EV: TTIP deal will be either rejected or postponed	12 October 2016	60%	0.3200
GJ: TTIP negotiations will not be completed before 1 January 2017	2 October 2016	99%	0.0002
EV: Marine Le Pen will win the presidency	12 October 2016	40%	0.320
GJ: Front National will win the presidency***	13 October 2016	13%	0.034
HM: Marine Le Pen will win the presidency	13 February 2017	15.5%	0.048
WH: Marine Le Pen will win the presidency	17 November 2016	40%	0.320
EV: The amount of migrants and asylum seekers will significantly decrease	12 October 2016	40%	0.720
GJ: Less than 100 000, or less than 300 000 refugees and migrants will arrive in Europe by sea in 2017	14 February 2017	81%	0.072
Average Brier score		**European Values**	0.496
		GJ Open Crowd Forecast	0.370

source: own elaboration (based on the data in Appendix C). The references and links are set out in Appendix C.
explanatory note (abbreviations of the institutions): EV = The European Values; GJ = The Good Judgment Open; NYT = The New York Times; FTE = FiveThirtyEight; HP = The Huffington Post; HM = Hypermind; WH = William Hill. ** the Democratic National Convention officially nominated Clinton on July 26, 2016. The crowd forecast, therefore, knew the "Democrat" refers to Hillary Clinton. *** Marine Le Pen announced the candidacy on April 8, 2016. The crowd forecast, therefore, knew the "Front National" refers to Marine Le Pen.

Looking at the final average Brier score of both the European Values and the GJ Open crowd forecast, the think-tank proved to be less accurate and hence also lost – after the first defeat in "experts-European Values comparison" – its second battle. This time, the difference is even more striking. Compared to the crowd forecast Brier score of 0.370, the think-tank finished with a coin-like Brier score of 0.496.

Nonetheless, at least some of the European Values´ forecasts met with partial success. Accordingly, the above table demonstrates one huge victory of the European Values over – not only – the GJ Open crowd forecast. This relates to the think-tank´s Brier score (0.605) regarding the US presidential election. In light of the Upshot´s election model developed for the New York Times, which suggested in November 2016 that "*Mrs. Clinton's chance of*

losing is about the same as the probability that an N.F.L. kicker misses a 37-yard field goal"[167], the think-tank´s judgment seems to be significant. Moreover, compared to the Huffington Post´s 98% probability, the European Values victory is crushing. Without a doubt, the think-tank was not able to correctly predict the future, but how many institutions or political scientists were? The European Values did a great job in accepting – by assigning 45% probability – relatively high possibility that the Trump would, in the end, win the presidency. That is why their Brier score reaches much lower value than the one of the GJ Open. Only the FiveThirtyEight project established by Nate Silver (see above in the theoretical chapter) "keeps pace" with the think-tank result. It is also noteworthy that the European Values´ research team used this question from the experts´ questionnaire, but did not apply the median value of individual expert´s estimates (see Appendix B). The research team lowered the median value by 5 percent. In hindsight, this was a wise move.

The think-tank defeated the crowd forecast also in the question regarding the AfD election result, but the difference between the European Values Brier score and the one of GJ Open is almost negligible. However, the position of GJ Open crowd forecast was more favourable since the forecasters firstly assigned the probability – of AfD getting more than 10% – in June 2017 (8 months later than the European Values). Compared to the European Values, the crowd forecast had four options to choose from, but by including only the two options covering the chance of AfD ten-plus result, the assigned probability was 29 percent. Hence, the think-tank beat the GJ Open by only 0.028. The GJ Open could have benefited from the time advantage (of being closer to the event, thus potentially have more information) also in the question concerning the US Midterms. But even in mid-December 2017, the crowd forecast still estimated the 25% probability that the Democrats would gain control (and thus win) in both chambers. As in the case of the European Values, estimating there is a 25% probability in four-option questions is like saying there is a 50% chance in binary questions. Thus, both actors achieved a Brier score of 0.500. Nonetheless, the successes of the European Values end here. Subsequent forecasted questions brought some heavy defeats to the European Values.

At first sight, the alternative question dealing with Russia´s potential assertive behaviour may seem incomparable to the think-tank´s statement about the potential occupation of a Baltic state. The opposite is true. The GJ Open question is just more encompassing. Based on the North Atlantic Treaty, the Article IV says the parties "*will*

[167] KATZ, Josh. Who Will Be President?. In: The New York Times [online]. [cit. 2019-07-08]. Available from: https://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html?mtrref=undefined

consult together whenever, in the opinion of any of them, the territorial integrity, political independence or security of any of the Parties is threatened".[168] In other words, the GJ Open question asked whether there was even a possibility that Russia´s behaviour would lead to a sense of deep insecurity amongst the Baltic countries. Here, the "behaviour" does not refer to the occupation as such, but even – for instance – to Russia´s border manoeuvres which would be perceived as a signal of a potential invasion. By assigning only 1% probability even to this "threat-perception" scenario (and not necessarily to the occupation as such), the crowd forecast showed high resoluteness.

Moreover, the potential Russian occupation is the only event that can be – due to the fairness – adjusted to the one-year horizon. The rest of both following and above mentioned forecasted questions act rather as processes, or as events with a strictly fixed date of occurrence. Nevertheless, the GJ Open asked already in 2015 whether a NATO member would invoke Article 4 in response to actions taken by Russia before January 2017. Therefore, the European Values probability needs to be adjusted to at least one-year horizon (the period from October 2016 only until 31 December 2016 falls under the first forecasted year). By following the formula for the probability adjustment (see in the previous chapter), the European Values probability of Russian occupation for the one-year horizon is 11%. Still, the higher accurateness of the GJ Open crowd forecast leads to a 0.024 difference in the Brier score.

The European Values suffered more substantial defeats in the accuracy comparison – again – due to the high resoluteness of the GJ Open crowd forecast. Regarding the TTIP deal, the crowd forecast more accurately predicted with 99% probability that the whole process would be postponed even until the beginning of 2017. Furthermore, the crowd forecast gave a significantly lower chance to the Marine Le Pen´s victory in the 2017 French presidential election (13% compared to 40% assigned by the European Values). However, the think-tank´s judgment was certainly not a poor decision. In November 2016, William Hill (betting company) also estimated a 40% probability of Le Pen victory (odds of 6/4).[169] In any case, each of both think-tank´s Brier scores lags behind by approximately 0.300. This is a vast difference in terms of accuracy.

[168] The North Atlantic Treaty: Washington D.C. - 4 April 1949. In: NATO [online]. [cit. 2019-07-14]. Available from: https://www.nato.int/cps/en/natolive/official_texts_17120.htm

[169] BETTING ODDS: : Marine Le Pen is the second favourite to win power in France. In: Business Insider [online]. [cit. 2019-07-08]. Available from: https://www.businessinsider.com/r-frances-le-pen-second-favorite-to-win-power-juppe-evens-bookmaker-2016-11

Nonetheless, the crowd forecast unambiguously defeated the European Values in the question concerning the number of asylum-seekers in the European Union. Once more, although the question created by the GJ open may appear to be very different from the think-tank´s statement, both questions are in fact comparable. In addition to that, the GJ Open case perfectly demonstrates what a proper operationalization should look like (see Table 10). The GJ Open was more specific and conditioned the potential increase/decrease by sea arrivals. Since the majority of "migrants" and asylum seekers really arrived by sea (e.g. through the Eastern Mediterranean Route or the Central Mediterranean Route), such operationalization does not impede the comparison.[170] Neither is there a problem with the fact that the GJ Open question is the four-option one. On closer examination (see Appendix C, Sheet 2), the crowd forecast gave the highest probabilities to two options related to a significant decrease of asylum-seekers. In February 2017, the crowd forecast estimated with a 65% probability that the number of asylum seekers arriving by sea in 2017 would be between 100 000 and 300 000 (inclusive). Compared to 2016, and taking to account the upper limit, this would mean – approximately – a 20% decrease. Such a decrease certainly is significant. Furthermore, the crowd forecast said with a 16% probability that less than 100 000 asylum seekers would arrive (at least 72% decrease, based on the upper limit).[171] Converting it to a binary statement and leaving aside the real outcome, we can say the crowd predicted with an 81% probability that there would be a significant decrease (regardless of whether the 20% decrease or the 72% decrease) in the number of asylum seekers arriving in the European Union. This entails the most obvious difference in the accuracy. Whereas the European Values wrongly predicted the future dynamics and thus have a Brier score of 0.720, the GJ Open crowd forecast finishes with a superior Brier score of 0.072.

On the whole, the crowd forecast´s Brier score of 0.370 is not among the best possible results, and mainly the failure to predict the real outcome of the French and the US presidential election harmed its final result. However, the general – and average – resoluteness of the GJ Open forecasters was enough to beat the foxy European Values in a seven-question comparison. Moreover, the think-tank Brier score of 0.496 is almost precisely what we would get by coin flipping (random guessing). In other words, as Tetlock suggests, this is identical to know nothing and saying 50/50 chance in every particular question.[172]

---

[170] Desperate Journeys: Executive Summary. In: UNHCR [online]. [cit. 2019-07-14]. Available from: https://www.unhcr.org/desperatejourneys/

[171] UNHCR Data - Mediterranean Situation: Sea arrivals monthly. In: UNHCR [online]. [cit. 2019-07-09]. Available from: https://data2.unhcr.org/en/situations/mediterranean

[172] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 309.

Let us see how well or poorly the European Values perform in the last verification test – in the comparison with randomly generated estimates.

6.3. European Values vs. Randomly Generated Estimates

As mentioned above, in the absence of data, randomly generated estimates provide a widely used means of comparing the accuracy of probabilistic forecasts. This accuracy-verification procedure indicates how good the overall – not only that seven-question – European Values´ Brier score of 0.410 really is. More precisely, it tells us in how many cases is the European Values´ Brier score able to produce a better result – and hence outperform the randomly generated forecasts (see below).[173] Nevertheless, several preparatory steps are required in order to conduct a proper comparison.

Initially, we need to choose the number of randomly generated forecasts intended for the accuracy verification. Based on the literature, such comparison usually relies on a simulation of several thousand randomly generated estimates (from 2000 to 5000 forecasts). However, researches sometimes use more robust test set comprised of 10 000 random guesses.[174] Nothing prevents us from utilizing the more robust version. Therefore, the Brier scores of 10 000 randomly generated guesses are compared with the final result of the European Values in those 46 questions.

Furthermore, it is possible to adjust the test set (with 10 000 randomly generated guesses) by employing the same distributional properties as those found in the original data. In other words, the proportion of occurrences and non-occurrences (1 and 0) of events/processes in the test set can be the same as that of the observed reality.[175] Following the forecasted questions in the European Values´ research project, 26 events or processes (out of 46), in reality, have really occurred. This accounts for 57% of all forecasted questions. Thus, in order to achieve more realism in the test set data, 5 700 randomly generated estimates are assigned with the code "1", which means – as always – that precisely 57% of the forecasted events in our comparison-reality have really occurred. Nonetheless, as we already know from the previous chapter, the assigned probability is what most affects the final Brier score. After such procedure, the simulation of 10 000 randomly generated forecasts can be

[173] MASON, S. J. Understanding forecast verification statistics. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 2008, 15.1: p. 31.
[174] Ibid., p. 33.
[175] Ibid., p. 31 and p. 33.

finally run, and – based on the same formula – corresponding Brier scores easily calculated (see Appendix C, Sheet 3).

If one looks at the Appendix C (Sheet 3), the figures clearly highlight that approximately 46%* out of 10 000 Brier scores based on the randomly generated estimates, were better (lower) and thus more accurate than the final score of 0.410 achieved by the think-tank in its research project. Put another way, the European Values were able to produce better results than the randomly generated forecasts in 54% cases. Hence, the question is: what does it say about the European Values predictive (accuracy) capabilities?

In the IARPA tournaments, the Good Judgment Project was able to beat the internal intelligence control group, but also university teams, by 60% in year one, and by 78% in year two. Basic training exercises along with the explanation of how scoring rules work were enough to achieve such good results. It should, however, be stressed that the Good Judgment Project ´s crowd forecast greatly outperformed analysts with access to classified data, not a simulation of randomly generated forecasts (thus randomly generated Brier scores).[176] Nevertheless, one can only assume the defeated teams finished with better results than those that would be obtained through randomly generated forecasts. In that sense, it is uncertain whether the Good Judgment Project could have defeated randomly generated forecasts by more than – or even by same – 60% or 78%. But it is still reasonable to regard this result of the European Values as very poor performance- and here is why.

The European Values have aspired to provide the Czech policy-making stakeholders with the probabilistic forecast of possible developments, and – based on their estimates – to formulate recommendations for what decisions the Czech politico-security elite should do to protect the vital interests of our country. This aim is no doubt worthy. Now imagine being one of the top policy-making stakeholders in the Czech Republic whose task is to find an institution that would predict important events in the politico-security realm. Such probabilistic forecasts would – as the European Values indicate – become an essential part for strategic planning of the Czech Republic, and would act both as a guide for the allocation of resources and an early warning tool (see also in the theoretical chapter). Naturally, accuracy-

* It is important to note that the "Percentage share of randomly generated forecasts with the lower Brier score than 0.410" in Appendix C (Sheet 3) can show you a slightly different number. This is caused by Excel´s "Automatic Calculation" function. When turned on, this function will always (even when you open the Appendix) recalculate 10 000 randomly generated estimates (not the occurrences). As a consequence, the final percentage share in your downloaded Appendix C can range from 0.44 (44%) to even 0.50 (50%). However, the core information is that the think-tank´s final Brier score was able to produce better results than the randomly generated forecasts in approximately 50% cases.

[176] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 17-18.

capabilities of the institution are the most important criteria for your selection. Would you hire the European Values? Given the fact that the think-tank accuracy is quite close to the one achieved through randomly generated estimates, you would probably not. From this perspective, if our decision-making stakeholders are satisfied with the overall Brier score of 0.410, they do not need to ask the European Values about the future. By running a simulation of randomly generated estimates (and thus Brier scores), they would have a good 50% chance of achieving a better score than the one obtained by the think-tank. The degree of the think-tank´s accuracy is higher only by 4%.

In order to bring the forecast verification (as a part of rigor) into the European Values´ research project, this chapter has analysed and compared the think-tank´s results at three levels. Answering the second research question, the overall European Values – close to a random guess – Brier score of 0.410 lost against the invited experts, GJ Open crowd forecast, and also against randomly generated estimates.

Based on the comparison between the research project and the aggregate of twelve randomly selected questionnaire-statements, the invited experts finished with a better Brier score of 0.373. Subsequently, the seven-question comparison between the think-tank and the GJ Open crowd forecasts has once again led to the defeat. The difference between the think-tank´s Brier score of 0.496 and the alternative Brier score of 0.370 was even more noticeable. Moreover, the European Values result was almost identical to that potentially achieved by random guessing. Finally, the think-tank´s overall brier score of 0.410 was only sufficient to beat approximately 50% of randomly generated estimates.

In that sense, the think-tank´s results are unconvincing but this might change soon. The improvement in accuracy certainly is possible. However, it rests upon honest evaluation of weak aspects of the European Values´ research project and systemic implementation of key recommendations. The following chapter, therefore, concerns both of these aspects.

7. Final Recommendations

Besides the fact that the European Values´ accuracy is unconvincing, the research project evinces (as demonstrated in the Analytical chapter) many recurrent shortcomings such as the vagueness, absence of thresholds, or devaluation of a clear message to decision-makers. All in all, the European Values research project could hardly serve as a means of strategic planning in the Czech politico-security area. This is not the result that would both bolster the Czech forecast-making and calm the critical voices of potential forecasting pessimists. However, an identification of the think-tank´s mistakes is of great value as it paves the way for eventual recommendations. Therefore, before proceeding to the conclusion of this diploma thesis, this chapter briefly introduces several key recommendations that could improve value added of any future forecasting enterprise.

One of the main problems devaluating the relevance of the European Values´ research project is the omnipresent "systemic" shortcomings. These do not concern the accuracy, but the preparation of forecasting questions per se. Both the questionnaire and the research project contain too many ambiguous statements (i.e. questions with more than one element, incompatibility of these elements; absence of specific thresholds, unstated judgments, mathematical and logical flaws). In general, the consequences arising from the ambiguity are threefold.

Firstly, the invited experts (and any research team) lose the understanding of what is required of them. As indicated in the analytical section, ambiguity can compel you to play excessively safe in your estimates, rather than enable you to make an honest, effective, and useful judgment about future phenomena. Extreme foxiness (forecasting around 50:50) logically reduces the strength of the information about the future.

Secondly, ambiguity, and especially the poor conceptualization (absence of specific thresholds for occurrence and non-occurrence) frustrates both guidance-capable and advise-capable potential of the probabilistic forecasting. Even though you achieve a superb Brier score in a particular set of questions, the decision-maker can hardly utilize it as an early-warning tool if he or she cannot recognize what dynamics the prediction precisely relates to (see the examples in the analytical chapter). Put simply, it acts as an impediment for delivering a clear message for policy-makers – not only – in the politico-security area.

Lastly, ambiguity obstructs accuracy verification. If you cannot determine whether or not the forecasted event/process – in reality – truly occurred, you also cannot evaluate your success or failure and hence use the Brier score as a feedback of your forecasting behaviour.

1. Recommendation: make sure you always prepare unambiguous and well-operationalized questions (statements). You will only be able to utilize all the unquestionable benefits of forecasting – and your research will only be taken seriously enough – once you do the proper operationalization.

From the perspective of accuracy, there is certainly room for improvement. Good judgment and accuracy skills are not about a talent we are born with. As Tetlock argues, it is about practice and feedback.[177] The key to the improvement lies in your attitude.

Therefore, it is initially crucial for your forecasters or invited experts to understand how probabilities work in time, what a scoring rule tells you about your result, and when "anti-probabilistic" estimates (0, 1, and 0.5) should be used. Moreover, estimates assigned around the "maybe-zone" of poor decisiveness do not often fulfil the role of a decision-making tool in the politico-security realm.

2. Sharpen the forecaster´s understanding of both scoring rules and probabilities.

Furthermore, the accuracy improvement relates to one of the steps in bringing the rigor into the forecast-making – to updating. By following the Bayesian thinking (see the theoretical chapter), super-forecasters update their estimates more often. The three-round questionnaire does not sufficiently allow your forecasters to update their true beliefs. Tetlock proves that those who gather the data and update (lower or raise their judgment) when spotting new information are more accurate.[178] The same rule applies, as we already know, to those using the finer-grained distinctions along the probability scales – the single percentage point scale. On top of that, "sticking to the ones" helps us not only to avoid the "anti-probabilistic" estimates but also to evade the "maybe-zone" of poor-decisiveness.

3. Enable your forecasters (when possible) to gather information and update their forecasts in real-time.

[177] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 252.
[178] Ibid., p. 92.

4. *Do not hesitate to exploit full gamut of 0-100 scale* (high granularity).

Aforementioned "feedback" is another means of accuracy improvement. The great advantage of the scoring rules is that it gives your research team (or invited experts) valuable information about your (or their) forecasting performance. Accordingly, the Brier score always shows you whether you are too under-confident or over-confident in particular topics.[179]

5. Thus, teach your forecasters to modify their forecasting behaviour in response to previous results. Next time, your final score could be lower.

The last recommendations regard the process of selection and cultivation of your forecasters. From dozens of experts invited by the European Values, 24 of them decided, in the end, to participate in the research project. Although experts´ estimates were used only minimally by the think-tank, the analysis in Chapter 4 demonstrates that despite all the problems, the invited experts – unlike the think-tank´s research team – were able to stray out of the foxy-like and "anti-probabilistic" forecasting. Furthermore, they beat the European Values in the twelve-question comparison. This gives the think-tank an opportunity to further cultivate the experts and utilize their full potential in future probabilistic-forecasting projects. However, the same goes for other institutions. Nevertheless, In order to follow the last four recommendations, you naturally need to:

6. Create many opportunities for your forecasters to improve and cultivate themselves.

Inviting the forecasters to assign their estimates only in the time of final research projects would be neither forethoughtful nor efficient. One option for the improvement of the accuracy skills of your forecasters is to test them with forecasting questions (statements) regularly throughout the year. The other one involves a series of small internal (or even open) tournaments focused on topics of interest. Both possibilities enable your forecasters to hone their forecasting skills sufficiently (and both are used in the Good Judgment Open).[180] Testing

[179] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 272.
[180] Good Judgment Open: Frequently Asked Questions. In: *Good Judgment Open* [online]. [cit. 2019-07-22]. Available from: https://www.gjopen.com/faq

your forecasters on a more regular basis essentially opens up an opportunity for the future improvement of the results of your institution. The data based on the results from tournaments or series of questions serves you as a means of "super-forecasters" identification. By identifying those who frequently achieves significantly higher results than others, you establish your top group which could magnify your overall results of the next research project. This can be done – for instance – by a weighted averaging. After calculating the average (or median as in the case of Delphi method) final probability based on the individual estimates of all of your forecasters (the wisdom of the crowd), you give extra weight to the estimates of the top group.* As a consequence, those most accurate forecasters get more influence in the collective conclusion. This is indeed one of the methods that helped the Good Judgment Project to win the IARPA tournament.[181] Finally, the last recommendation is:

7. Exploit the results of the forecasters´ past performances, identify your top group, and give them an extra weight in calculating the final, collective probability estimates.

After overcoming the problem with ambiguity and operationalization, even such a basic above-mentioned tutorial can boost your accuracy. Modest improvements stack up and materialise after some time. Tetlock himself, in his book, offers much broader instruction definitely enhancing your accuracy capabilities. Proved in real-world forecasting tournaments, his 60-minutes tutorial improved the accuracy of the crowd forecast by roughly 10% in one year.[182] What is more, he proves it is possible to improve the foresight simply by measuring, analysing, and verifying your results.

Moreover, the accuracy improvement, as well as the necessary professionalism of your projects, could definitely move the Czech probabilistic-forecasting attempts one step closer towards becoming a valuable tool for good decision-making – not only – in the politico-security realm.

[181] TETLOCK, Philip E.; GARDNER, Dan. *Superforecasting: The art and science of prediction*. Random House, 2016, p. 90-91.
[182] Ibid., p. 18 and p. 278-285.
*see various methods of estimate-weighting here: KARVETSKI, Christopher W., et al. Probabilistic coherence weighting for optimizing expert forecasts. *Decision Analysis*, 2013, 10.4: 305-326.

CONCLUSION

Ideally, probabilistic forecasts play – and in the Czech Republic, too, should play – an indispensable role in the decision-making processes. Probabilistic forecasting enhances and facilitates anticipation of threats to the national security and assists the decision-making stakeholders in policy advising or strategic planning. Accordingly, the European Values Think-Tank itself aspired to provide – with its forecasts – the Czech policy-makers and the security elite with the development of possible "futures" and recommendations for protecting the vital strategic interests of the Czech Republic. The European Values Think-Tank unquestionably deserves our recognition for bringing the probabilistic forecasting into the Czech politico-security debate.

However, poorly executed probabilistic forecasting may cause more harm than good. At best, unsuccessful forecasting attempts may discredit the forecasting enterprise in the eyes of decision-making elites. At worst, wrong predictions may lead to the misallocation of scarce resources or to unnecessary securitization of the public debate with all the negative consequences. The dividing line between the positive and negative outputs of the forecasting enterprise, as well as between the successful and poor forecasting, lies in the accuracy. The accuracy, however, is inherently connected with the verification. Without systemic effort at measuring our accuracy, we cannot improve the quality of our forecasts. In that sense, it would be one thing to kick-start the new and potentially very useful method for political-security inquiry in the Czech Republic, and quite another to offer both solid and valuable forecasts.

If we, in the Czech Republic, are serious about utilizing the probabilistic forecasting – not only – in the politico-security area, every attempt must be accompanied by the measurement of forecasting accuracy. Therefore, this diploma thesis has measured and verified the performance (accuracy) capabilities of the European Values forecasts.

The excessive occurrence of vagueness and absence of the proper operationalization – negatively affecting invited experts, decision-makers, and the measurement – have impeded us from verifying all think-tank´s forecasted questions. Nevertheless, an evaluation of predictive capabilities based on 46 forecasts (out of 58) provided sufficient ground for responding our research questions. Regarding the first research question, the European Values finished with a "neither excellent nor disastrous" Brier score of 0.410, and hence do not lean towards either perfect accuracy or great error. However, such a result is relatively very close

to the one you would obtain by random guessing (0.5) or coin-flipping. It only confirmed think-tank´s extremely-foxy behaviour and thus also the inexperience found – with a statistical significance – during the analysis of the European Values. Nonetheless, accuracy alone does not provide an overall picture. Even the mediocre accuracy, if better than the accuracy achieved by alternative (forecasting) institutions and other means of comparison, can mean success. A broader perspective is necessary.

Yet the comparison did not make a change for the better. Answering the second research question, the think-tank proved to be worse in two consecutive comparisons and did not subsequently amaze in the last one. Firstly, the European Values were beaten – compared to the aggregate of twelve randomly selected questionnaire-statements – by the invited experts. Simply put the European Values had been better off had they decided to bet on median guesses of their forecasters Secondly, they recorded even more substantial defeat in the seven-question comparison with the Good Judgment Open crowd forecasts. However, the third result is more worrying. The think-tank was able to beat only approximately 50% of 10 000 randomly generated estimates.

In light of our result, the Czech decision-makers should be careful when engaging with the European Values´ research project and its policy recommendations. The overall accuracy runs counter to their aspiration to help the Czech security elite with strategic planning. Decision-making stakeholders would get almost the same results by running a robust simulation of randomly generated guesses. This is not what the think-tank had hoped for.

Nonetheless, these results do not prove the forecasting sceptics right and should not be read as a rejection of the forecasting enterprise in general – quite contrary. The think-tank´s accuracy is poor but this might change soon. The improvement, however, rests upon honest evaluation of weak aspects of the European Values´ research project and systemic implementation of key recommendations as developed in the previous chapter. Numerous examples of forecasting in complex systems, ranging from weather to economics, clearly show that the room for improvement is really huge. To seize it, all we have to do is to – similarly to this work – bring the rigor into the probabilistic forecasting. If we are to reduce the final recommendation to a single sentence, we would close with the following call: revise and update your forecasts, cultivate your forecasters, verify your results, and repeat – regularly – this whole circle. If the diploma thesis can persuade any future forecasting enterprise to keep the rigor in mind, the journey towards proper probabilistic forecasting in the Czech Republic can begin.

SUMMARY

Probabilistic forecasting, despite the forecasting sceptics, can play an indispensable role in policy advising, provision of possible scenarios and strategic planning. In light of that, the European Values Think-Tank´s research project is both a unique attempt and a significant breakthrough in utilizing probabilistic forecasting. However, the accuracy of probabilistic forecasts is required in order to fulfil its valuable advise-capable and guidance-capable potential – not only – in the politico-security realm. Therefore, the aim of this thesis was to evaluate, by using the Brier score, the think-tank´s accuracy and thus to bring the crucial element of verification into the Czech probabilistic forecasting.

Nonetheless, the additional data provided by the think-tank itself enabled us to analyse and compare the invited experts with the think-tank´s research team in other aspects related to forecasting behaviour. The analytical section demonstrates that unlike the European Values´ research team, the invited experts show no signs of "anti-probabilistic" thinking or poor-decisiveness. Conversely, the think-tank´s research team was unable to sufficiently evade the "maybe-zone" of 40%, 50%, and 60% probabilities. This also affected the overall foxy-like accuracy of the European Values. Moreover, the analysis also reveals the excessive occurrence of vagueness and absence of the proper operationalization in the think-tank´s research project.

Based on the verification assessment, the European Values finished with neither excellent nor disastrous accuracy. However, their accuracy capabilities proved to be poor when compared with alternative foreign forecasts as well as with other means of the accuracy comparison. The European Values were beaten by both the invited expert and the Good Judgment Open crowd forecast. What is more, the think-tank was able to defeat only 50% of 10 000 randomly generated estimates. This certainly is an unconvincing result, running counter to think-tank´s aim of helping the Czech security elite with strategic planning. Nevertheless, numerous examples that with the involvement of the rigor, the room for improvement is really huge.

Thus, this diploma thesis finally introduces several key recommendations that could improve both accuracy and value added of any future forecasting enterprise. The key lies in revising and updating your forecasts, cultivating your forecasters and verifying your results. Such a brief tutorial could definitely move the Czech probabilistic-forecasting attempts one step closer towards becoming a valuable tool for good decision-making – not only – in the politico-security realm.

List of References

1) BACON JR., Perry. The Five Wings Of The Republican Party. In: FiveThirtyEight [online]. [cit. 2019-06-26]. Available from: https://fivethirtyeight.com/features/the-five-wings-of-the-republican-party/

2) BALDUS, Bernd. Positivism's twilight?. Canadian Journal of Sociology/Cahiers canadiens de sociologie, 1990, 149-163.

3) BERLING, Trine Villumsen; BUEGER, Christian (ed.). Security expertise: practice, power, responsibility. Routledge, 2015, 266 p. ISBN 978-1-138-81911-5

4) BO, Yuanchao Emily, et al. An IRT forecasting model: linking proper scoring rules to item response theory. Judgment & Decision Making, 2017, 12.2.

5) BRIER, Glenn W. Verification of forecasts expressed in terms of probability. Monthly weather review, 1950, 78.1: 1-3.

6) BRÖCKER, Jochen; SMITH, Leonard A. Scoring probabilistic forecasts: The importance of being proper. Weather and Forecasting, 2007, 22.2: 382-388.

7) BURNS, Alexander a Jonathan MARTIN. Trump's Takeover of the Republican Party Is Almost Complete. In: The New York Times [online]. [cit. 2019-06-26]. Available from: https://www.nytimes.com/2019/04/03/us/politics/trump-republican-party.html

8) CANDILLE, Guillem; TALAGRAND, Olivier. Evaluation of probabilistic prediction systems for a scalar variable. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 2005, 131.609: 2131-2150.

9) CHANG, Welton, et al. Accountability and adaptive performance under uncertainty: A long-term view. Judgment & Decision Making, 2017, 12.6.

10) CHOUCRI, Nazli. Forecasting in international relations: problems and prospects. International Interactions, 1974, 1.2: 63-86.

11) ČERVENKA, Jan. Důvěra ústavním institucím v únoru 2016: TISKOVÁ ZPRÁVA. In: Centrum pro výzkum veřejného mínění (CVVM) [online]. [cit. 2019-05-22]. Available from: https://cvvm.soc.cas.cz/media/com_form2content/documents/c2/a2020/f9/pi160301.pdf

12) DE FINETTI, Bruno. Theory of probability: A critical introductory treatment. John Wiley & Sons, 2017. 596 p. ISBN 9781119286349.

13) DETSCH, Jack. Pentagon reports troop surge in Middle East. In: Al-Monitor [online]. [cit. 2019-06-28]. Available from: https://www.al-monitor.com/pulse/fa/originals/2017/11/pentagon-troop-surge-syria-middle-east.html

14) DORAN, Charles F. Why forecasts fail: the limits and potential of forecasting in international relations and economics. International Studies Review, 1999, 1.2: 11-41.

15) DOSWELL III, Charles A. Probability, Climatology, and Forecasting: Essay on Blog Page. In: Flame [online]. [cit. 2019-06-04]. Available from: http://www.flame.org/~cdoswell/forecasting/probability/Prob_Clim_Fcstg.html

16) FEDER, Stanley A. Forecasting for policy making in the post–cold war period. Annual Review of Political Science, 2002, 5.1: 111-125.

17) FERRO, Christopher AT. Comparing probabilistic forecasting systems with the Brier score. Weather and Forecasting, 2007, 22.5: 1076-1088.

18) FREEMAN, John R.; JOB, Brian L. Scientific forecasts in international relations: Problems of definition and epistemology. International Studies Quarterly, 1979, 23.1: 113-143.

19) FRIEDMAN, Jeffrey A., et al. The value of precision in probability assessment: Evidence from a large-scale geopolitical forecasting tournament. International Studies Quarterly, 2018, 62.2: 410-422.

20) FRIEDMAN, Jeffrey A.; ZECKHAUSER, Richard. Handling and mishandling estimative probability: likelihood, confidence, and the search for Bin Laden. Intelligence and National Security, 2015, 30.1: 77-99.

21) FRIEDMAN, Milton. The Methodology of Positive Economics. In: HAUSMAN, Daniel M. The Philosophy of Economics: An Anthology. 3rd ed. Cambridge : Cambridge University Press, 2008. p. 145-178. ISBN-13 978-0-511-37141-7.

22) GARTZKE, Erik. War is in the Error Term. International Organization, 1999, 53.3: 567-587.

23) GNEITING, Tilmann; RAFTERY, Adrian E. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 2007, 102.477: 359-378.

24) GUILLAN, Amanda. Epistemological limits to scientific prediction: The problem of uncertainty. Open Journal of Philosophy, 2014, 4.04: 510.

25) HASS, P. Epistemic Communities and International Policy Coordination-An Introduction. International Organization, 1992, 46.1: 1-35.

26) HEGRE, Håvard, et al. Predicting armed conflict, 2010–2050. International Studies Quarterly, 2013, 57.2: 250-270.

27) HELMER, Olaf. Analysis of the future: The Delphi method. RAND CORP SANTA MONICA CA, 1967.

28) HOSNI, Hykel; VULPIANI, Angelo. Forecasting in light of big data. Philosophy & Technology, 2018, 31.4: 557-569.

29) JOLLIFFE, Ian T.; STEPHENSON, David B. (ed.). Forecast verification: a practitioner's guide in atmospheric science. John Wiley & Sons, 2012, 254 p. ISBN 0-471-49759-2.

30) KAHN, Chris. Trump edges ahead of Clinton in U.S. presidential race: Reuters/Ipsos poll. In: Reuters [online]. [cit. 2019-05-21]. Available from: https://www.reuters.com/article/us-usa-election-poll-idUSKCN1062MC

31) KAHNEMAN, Daniel; LOVALLO, Dan. Timid choices and bold forecasts: A cognitive perspective on risk taking. Management science, 1993, 39.1: 17-31.

32) KALOUS, Miroslav. HOW (NOT) TO PREDICT THE FUTURE?: Analysis of several pioneering studies in the field of Czech political and security scenario-building. In: Obrana a strategie [online]. [cit. 2019-03-03]. Available from: https://www.obranaastrategie.cz/en/archive/volume-2018/1-2018/articles/how-not-to-predict-the-future.html

33) KARVETSKI, Christopher W., et al. Probabilistic coherence weighting for optimizing expert forecasts. Decision Analysis, 2013, 10.4: 305-326

34) KATZ, Josh. Who Will Be President?: Hillary Clinton has an 85% chance to win. In: The New York Times [online]. [cit. 2019-05-21]. Available from: https://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html

35) KING, Gary; ZENG, Langche. Explaining rare events in international relations. International Organization, 2001, 55.3: 693-715.

36) KIRCHICK, James. Russia's plot against the West. In: Politico [online]. [cit. 2019-07-03]. Available from: https://www.politico.eu/article/russia-plot-against-the-west-vladimir-putin-donald-trump-europe/

37) KRISTÓF, Tamás. Is it possible to make scientific forecasts in social sciences?. Futures, 2006, 38.5: 561-574.

38) KUOSA, Tuomo. The evolution of strategic foresight: Navigating public policy making. Routledge, 2016.

39) LAI, Tze Leung, et al. Evaluating probability forecasts. The Annals of Statistics, 2011, 39.5: 2356-2382.

40) LICHTENSTEIN, Sarah; FISCHHOFF, Baruch; PHILLIPS, Lawrence D. Calibration of probabilities: The state of the art. In: Decision making and change in human affairs. Springer, Dordrecht, 1977. p. 275-324.

41) LUCKNER, Stefan, et al. Prediction markets: Fundamentals, designs, and applications. Springer Science & Business Media, 2011, 156 p. ISBN 978-3-8349-3358-4.

42) MAKRIDAKIS, Spyros; HOGARTH, Robin M.; GABA, Anil. Forecasting and uncertainty in the economic and business world. International Journal of Forecasting, 2009, 25.4: 794-812.

43) MANKOFF, Jeffrey. The Russian economic crisis. Council on Foreign Relations, 2010, 44 p.

44) MASON, S. J. Understanding forecast verification statistics. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 2008, 15.1: 31-40.

45) MCNEES, Stephen K. The role of judgment in macroeconomic forecasting accuracy. International Journal of Forecasting, 1990, 6.3: 287-299.

46) MEIER, Kenneth J.; BRUDNEY, Jeffrey L.; BOHTE, John. Applied statistics for public and nonprofit administration. Nelson Education, 2011, 608 p. ISBN 13: 978-1-111-34280-7.

47) MELLERS, Barbara, et al. Identifying and cultivating superforecasters as a method of improving probabilistic predictions. Perspectives on Psychological Science, 2015, 10.3: 267-281.

48) MELLERS, Barbara, et al. The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. Journal of experimental psychology: applied, 2015, 21.1: 1.

49) MILLER, Charles A. Prediction and its discontents: guidance for Australia from the debate over social science forecasting. Australian Journal of International Affairs, 2014, 68.4: 418-432.

50) MURPHY, Allan H.; WINKLER, Robert L. A general framework for forecast verification. Monthly weather review, 1987, 115.7: 1330-1338.

51) NAXERA, Petr; KRČÁL, Vladimír. Populistická konstrukce národa ohroţeného migrací: CAQDAS volebního diskurzu českých parlamentních voleb v roce 2017. Sociológia, 2018, 50.5: p. 491-523.

52) PRIEST, Graham. The sun may not, indeed, rise tomorrow: a reply to Beall. Analysis, 2012, 72.4: 739-741.

53) ROBBINS, Lionel. The Nature and Significance of Economic Science. In: HAUSMAN, Daniel M. The Philosophy of Economics: An Anthology. 3rd ed. Cambridge : Cambridge University Press, 2008, 70-99. ISBN-13 978-0-511-37141-7.

54) RYCHNOVSKÁ, Dagmar. Bio (in) security, scientific expertise, and the politics of post-disarmament in the biological weapons regime. Geoforum, 2017, 84: 378-388.

55) SAREWITZ, Daniel; PIELKE JR, Roger. Prediction in science and policy. Technology in Society, 1999, 21.2: 121-133.

56) SCHEGLOFF, Emanuel. On granularity. Annual Review of Sociology, 2000, 26.1: 715-720.

57) SCHNEIDER, Gerald; GLEDITSCH, Nils Petter; CAREY, Sabine C. Exploring the past, anticipating the future: A symposium. International Studies Review, 2010, 12.1: 1-7.

58) SHUYA, Mason. Russian Cyber Aggression and the New Cold War. Journal of Strategic Security, 2018, 11.1: 1-18.

59) SILVER, Nate. The signal and the noise: why so many predictions fail--but some don't. Penguin, 2012, 628 p. ISBN 978-1-101-59595-4.

60) SIMS, Christopher A., et al. Are forecasting models usable for policy analysis?. Quarterly Review, 1986, Win: 2-16.

61) SLOVIC, Paul; LICHTENSTEIN, Sarah. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational behavior and human performance, 1971, 6.6: 649-744.

62) SONNE, Paul and Missy RYAN. Trump approves sending more forces to the Middle East amid tensions with Iran. In: The Washington Post [online]. [cit. 2019-07-02]. Available from: https://www.washingtonpost.com/world/national-security/trump-approves-sending-more-forces-to-the-middle-east-amid-tensions-with-iran/2019/05/24/7403ffac-7e2c-11e9-a66c-d36e482aa873_story.html

63) STRAUSS, Harlan J.; ZEIGLER, L. Harmon. Delphi, political philosophy and the future. Futures, 1975, 7.3: 184-196.

64) STONE, Jon. Brexit result must be respected, David Cameron says. In: The Independent [online]. [cit. 2019-05-21].Available from: https://www.independent.co.uk/news/uk/politics/brexit-result-must-be-respected-david-cameron-says-a7105886.html

65) ŠABIĆ, Senada Šelo. The relocation of refugees in the European Union. 2017.

66) TETLOCK, Philip E.; GARDNER, Dan. Superforecasting: The art and science of prediction. Random House, 2016, 352 p. ISBN 9781847947154.

67) THARAVANIJ, Piyapas. Capital Market, Severity of Business Cycle, and Probability of an Economic Downturn. Monash University Economics Discussion Paper, 2007.

68) THAROOR, Ishaan. Macron defeated Le Pen in France's presidential election. Here's what happens next. In: The Washington Post [online]. [cit. 2019-06-21]. Available from: https://www.washingtonpost.com/news/worldviews/wp/2017/05/07/macron-defeated-le-pen-in-frances-presidential-election-heres-what-happens-next/?utm_term=.f553d6a9c3f2

69) VAN DER HEIJDEN, Kees. Scenarios and forecasting: two perspectives. Technological forecasting and social change, 2000, 65.1: 31-36.

70) VAN VUGHT, Frans A. Pitfalls of forecasting: fundamental problems for the methodology of forecasting from the philosophy of science. Futures, 1987, 19.2: 184-196.

71) WARD, Michael D., et al. Learning from the past and stepping into the future: Toward a new generation of conflict prediction. International Studies Review, 2013, 15.4: 473-490.

72) WILKS, Daniel S. Statistical methods in the atmospheric sciences. Academic press, second edition, 2006, 648 p. ISBN 13: 978-0-12-751966-1.

73) ZENKO, Micah. US military policy in the Middle East: an appraisal. Chatham House, 2018.

74) EUROPOL, TESAT. European Union Terrorism Situation and Trend Report 2018. 2018, 70 p. ISBN 978-92-95200-91-3.

75) EUROPOL, TESAT. European Union Terrorism Situation and Trend Report 2019. 2019, 82 p. ISBN 978-92-95209-76-3.

76) European Values Think-Tank Report: Forecast of the Development of the Politico-Security Environment from the Perspective of the Interests of the Czech Republic 2016 – 2019 [online]. 21. 11. 2016 [cit. 2019-06-19]. Available from: https://www.europeanvalues.net/wp-content/uploads/2016/11/Scenarios-for-the-Development-of-the-Politico-Security-Environment-from-the-Perspective-of-the-Interests-of-the-Czech-Republic-2016-%E2%80%93-2019_FINAL.pdf

77) Good Judgment. In: Good Judgment [online]. [cit. 2019-07-13]. Available from: https://goodjudgment.com/

78) Good Jugment Open: In: Good Judgment Open [online]. [cit. 2019-07-13]. Available from: https://www.gjopen.com/

79) RAND: Delphi Method. In: RAND Corporation [online]. [cit. 2019-07-11]. Available from: https://www.rand.org/topics/delphi-method.html

80) Russia Real GDP Growth: 1996 - 2018 (Quarterly; %). In: CEIC [online]. [cit. 2019-06-23]. Available from: https://www.ceicdata.com/en/indicator/russia/real-gdp-growth

81) Russian Activity Tracker: Nordic-Baltic Region. In: The Stockholm Free World Forum [online]. [cit. 2019-07-03]. Available from: https://frivarld.se/russian-activity-tracker-nordic-baltic-region/

82) Significant Cyber Incidents. In: Center for Strategic and International Studies [online]. [cit. 2019-06-26]. Available from: https://www.csis.org/programs/technology-policy-program/significant-cyber-incidents

83) UNHCR Data - Mediterranean Situation: Sea arrivals monthly. In: UNHCR [online]. [cit. 2019-07-09]. Available from: https://data2.unhcr.org/en/situations/mediterranean

84) Desperate Journeys: Executive Summary. In: UNHCR [online]. [cit. 2019-07-14]. Available from: https://www.unhcr.org/desperatejourneys/

85) The North Atlantic Treaty: Washington D.C. - 4 April 1949. In: NATO [online]. [cit. 2019-07-14]. Available from: https://www.nato.int/cps/en/natolive/official_texts_17120.htm

86) Relations with Russia. In: North Atlantic Treaty Organization [online]. [cit. 2019-07-03]. Available from: https://www.nato.int/cps/en/natolive/topics_50090.htm

87) Relocation of asylum seekers from Italy and Greece: Datas. In: European Migration Law [online]. [cit. 2019-06-27]. Available from: http://www.europeanmigrationlaw.eu/en/articles/datas/relocation-from-italy-and-greece.html?fbclid=IwAR2kQYPzhsW_8RoUJa47yELox89nbyNdBiCx2W96drmmg0o5BsFL3U2nSlE

88) BETTING ODDS: : Marine Le Pen is the second favourite to win power in France. In: Business Insider [online]. [cit. 2019-07-08]. Available from: https://www.businessinsider.com/r-frances-le-pen-second-favorite-to-win-power-juppe-evens-bookmaker-2016-11

89) The road not taken: Brexit has not caused much economic damage. Until now. In: The Economist [online]. [cit. 2019-03-03]. Available from: https://www.economist.com/britain/2019/02/16/brexit-has-not-caused-much-economic-damage-until-now

90) Ideologové, nebo odvážní bojovníci s propagandou? Think tank Evropské hodnoty dráždí politiky. In: iROZHLAS [online]. [cit. 2019-07-20]. Available from: https://www.irozhlas.cz/zpravy-domov/ideologove-nebo-odvazni-bojovnici-s-propagandou-think-tank-evropske-hodnoty-drazdi-politiky_201701131300_dbernardy

91) Brexit a postoje k případnému Czexitu, 13. 6. 2016 – 14. 6. 2016: Výzkum veřejného mínění pro Český rozhlas. In: Median [online]. [cit. 2019-05-20]. Available from: http://www.median.eu/cs/wp-content/uploads/2016/06/1516420_Bleskovy_pruzkum_cerven_2016_N_1049_zprava_v05_BREXIT.pdf

92) Zeman je pro referendum o českém odchodu z EU a NATO, hlasoval by ale proti. In: Český rozhlas: iRozhlas [online]. [cit. 2019-05-20]. Available from: https://www.irozhlas.cz/zpravy-domov/zeman-je-pro-referendum-o-ceskem-odchodu-z-eu-a-nato-hlasoval-by-ale-proti_201606302005_mkopp

93) Miloš Zeman podle premiéra rozděluje společnost a připravuje se na příští kampaň. In: Česká televize: ČT 24 [online]. [cit. 2019-05-22]. Dostupné z: https://ct24.ceskatelevize.cz/domaci/1678477-milos-zeman-podle-premiera-rozdeluje-spolecnost-a-pripravuje-se-na-pristi-kampan

94) Čas ubývá, kandidátů přibývá. Seznamte se s těmi, kteří chtějí být českým prezidentem. In: Česká televize: ČT 24 [online]. [cit. 2019-05-22]. Available from: https://ct24.ceskatelevize.cz/domaci/2286658-cas-ubyva-kandidatu-pribyva-seznamte-se-s-temi-kteri-chteji-byt-ceskym-prezidentem

95) Česko je rejdištěm dezinterpretací a dezinformací, říká rektor. O unii chce šířit objektivní informace: interview with rector Jaroslav Miller (Palacký University Olomouc). In: iRozhlas: Interview Plus [online]. 5. 4. 2019 [cit. 2019-05-21]. Available from: https://www.irozhlas.cz/zpravy-domov/euforka-univerzita-palackeho-olomouc-fake-news-dezinformace-evropska-unie_1904051740_och

96) Vrchní soud potvrdil důchodci Baldovi čtyři roky za teroristický útok na vlaky. In: iRozhlas [online]. [cit. 2019-06-28]. Available from: https://www.irozhlas.cz/zpravy-domov/balda-senior-teroristicky-utok-utok-na-vlak_1904161114_nkr

LIST OF APPENDICES

List of External Appendices (Excel Spreadsheets)

Appendix A: The European Value´s Research Project Data

Appendix B: Questionnaire Data (invited experts)

Appendix C: Comparison Data

See the Appendices as attachments to the thesis (ZIP format), or

See the Appendices as attachments in the Charles University Digital Repository: Search the Item Title: *The Art of (In)Accuracy: A Meta-Analysis of the European Values Think-Tank´s Forecasts*), or

Download the Appendices on independent.academia.edu/MatějŠtěpánek (link)
https://independent.academia.edu/Mat%C4%9Bj%C5%A0t%C4%9Bp%C3%A1nek