



POLITICAL BIAS OR FAIR PLAY? HOW CONSISTENT ARE THE EUROPEAN COMMISSION'S PROGRESS REPORTS WITH VIEW TO THE RULE OF LAW?

A comparative analysis using the examples of
Croatia and Turkey

Inauguraldissertation zur Erlangung des Doktorgrades (Dr. rer.
Pol.) des Fachbereichs Politik- und Sozialwissenschaften an der
Freien Universität Berlin

vorgelegt von: Jana Bachmann, M.A., M.A., Berlin, 2016

Erstgutachterin: Prof. Dr. Carina Sprungk

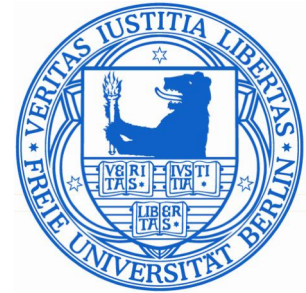
Zweitgutachterin: Prof. Dr. Eva Heidbreder

Datum der mündlichen Prüfung: 01.03.2017

Abstract

The European Commission has been accused of producing politically biased Progress Reports on its candidate countries by Member States, candidate countries and academia. This claim is tested by comparing the Commission's reporting results to those of two other, independent democracy evaluation indices after transposing them onto one, common scale. The predominant consistence of their results speaks against the political bias claim and to the high quality of the European Commission in gathering factual information on the candidate countries. However, the indices' level of independence proves to be lower than expected since they are basically using the same sources, including one another. Thus, more resourceful indices have a bigger impact on the general evaluation tone when less equipped indices draw from their Reports. Also, institutional choices that lead to and reinforce the political bias perception must be addressed by the Commission. These aim at increased transparency of the reporting process by a) publishing the reporting methodology as well including specific checklists for the respective criteria, b) reducing unofficial input opportunities for member and candidate countries and c) dissolving the double role of the Commission as both assessor and distributor of financial support to the candidate countries.

Die Europäische Kommission wird von Seiten der Mitgliedsstaaten, der Beitrittskandidaten sowie von Vertretern der Politikwissenschaften der politischen Färbung ihrer Fortschrittsberichte für Kandidatenländer beschuldigt. Diese Behauptung wurde getestet, indem die Ergebnisse der Fortschrittsberichte der Europäischen Kommission mit denen von zwei unabhängigen Demokratiemessungsindizes verglichen wurden. Dies geschah, nachdem alle drei Indizes mithilfe einer einheitliche Messskala vergleichbar gemacht wurden. Der hohe Grad an Konsistenz der Indexergebnisse schwächt das Argument der politischen Färbung und unterstützt die Wahrnehmung, dass die Europäische Kommission qualitativ hochwertige, faktische Informationen über die Kandidatenländer erhebt. Allerdings ist der Grad der Unabhängigkeit zwischen den Indizes niedriger als erwartet, da alle drei im Wesentlichen dieselben Quellen nutzen, inklusive einander, was wiederum zu zirkulären Zitationen führt. Dadurch haben Indizes mit höheren Ressourcenkapazitäten einen größeren Einfluss auf den allgemeinen Bewertungston, insbesondere, wenn ressourcenärmere Indizes sich ihrer als Quellen bedienen. Darüber hinaus führen institutionelle Entscheidungen seitens der Europäischen Kommission zu einer verstärkten Bias-Perzeption, welche wiederum adressiert werden sollte. Konkrete Politikempfehlungen beziehen sich in erster Linie auf die transparentere Gestaltung des Bewertungsprozesses inklusive der a) Veröffentlichung der Evaluierungsmethoden mit dezidierten Kriterien-Checklisten, b) der Reduzierung von inoffiziellen Einflussmöglichkeiten seitens der Mitglieds- und Kandidatenländer und c) der Auflösung der Doppelrolle der Kommission als Evaluierender und gleichzeitiger Distributor von finanziellen Zuwendungen an die Kandidatenländer.



Name: Bachmann

Vorname: Jana

Matrikelnummer:

Selbstständigkeitserklärung zur Dissertation

Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten schriftlichen Arbeit mit dem Titel

Political Bias or Fair Play? How consistent are the European Commission's Progress Reports with view to the Rule of Law?

A comparative analysis using the examples of Croatia and Turkey

um eine von mir selbstständig und ohne fremde Hilfe verfasste Arbeit handelt.

Ich erkläre ausdrücklich, dass ich *sämtliche* in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet (einschließlich Tabellen, Grafiken u. Ä.) als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken u. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken u. Ä. anderer Autorinnen und Autoren (Paraphrasen) die Quelle angegeben habe.

Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbstständigkeit als Täuschung betrachtet und entsprechend der Prüfungsordnung und/oder der Allgemeinen Satzung für Studien- und Prüfungsangelegenheiten der FU geahndet werden.

Datum:

Unterschrift:

Danksagung

Mein besonderer Dank gilt Professor Dr. Eckart D. Stratenschulte, ohne den diese Arbeit nie entstanden wäre. Es gibt keinen Zweiten, der es so gut wie Sie versteht, das richtige Maß an Geduld, ‚seriösem Pragmatismus‘ und Motivation in Zeiten intellektueller Durststrecken zu finden. Sie sind mein persönlicher Held und ich schätze Ihre Mentorenschaft sehr!

Weiterhin bedanke ich mich bei Prof. Dr. Carina Sprungk, deren kluge Hinweise insbesondere die methodologische Vorgehensweise dieser Arbeit stark geprägt haben sowie bei Prof. Dr. Eva Heidbreder, deren Ermutigung zu klaren Worten die Prägnanz der Untersuchungsergebnisse untermauerte.

Darüber hinaus wäre diese Arbeit nicht ohne die Unterstützung der Friedrich-Naumann-Stiftung für die Freiheit entstanden. Den durch die FNF ermöglichten zahlreichen Seminaren und hervorragenden Gesprächen mit MitstipendiatInnen und MitarbeiterInnen (insbesondere Frau Dr. Kathleen Klotchkov) verdanke ich mein persönliches und wissenschaftliches Wachstum der letzten drei Jahre.

Nicht zuletzt danke ich meinen Eltern für die tagelange Belagerung ihrer Büros, unendliche Stunden am Drucker sowie die Vermeidung der Frage, wann ich mich endlich dem Arbeitsmarkt zuzuführen gedenke.

List of Abbreviations

BTI = Bertelsmann Transformation Index

CC= Copenhagen Criteria

COM = European Commission

EU = European Union

FH = Freedom House

JA = Judicial Accountability

JC = Judicial Council

JI = Judicial Independence

JIm = Judicial Impartiality

MoJ = Ministry of Justice

RoL = Rule of Law

Table of Contents

How consistent are the European Commission’s Progress Reports with view to the Rule of Law?

A comparative analysis using the examples of Croatia and Turkey.

I. Introduction	11
II. Literature Review	21
II.1 Liberal Democracy and the Rule of Law	21
II.1.1 Defining Liberal Democracy and Democratization	21
II.1.2 Defining the Rule of Law	22
II.1.3 Relationship between Liberal Democracy and the Rule of Law	31
II.1.4 The Role of the Rule of Law in External Democracy Promotion	35
II.1.5 The Role of the Rule of Law in Croatia and Turkey’s Democratization Process	38
II.2 External Democracy and Rule of Law Promotion	44
II.2.1 History and Actors	44
II.2.2 Motives	46
II.2.3 Types and Methods	47
II.2.4 The EU as an external Democracy and Rule of Law Promoter	48
II.2.5 EU Monitoring and Progress Reports	56
II.2.6 The EU’s RoL concept within the accession process	57
II.3 Measuring the Rule of Law	73
II.3.1 Actors and Motives behind Democracy and Rule of Law Evaluation	73
II.3.2 Quality Criteria for Adequate Measuring	75
II.3.3 Research Gaps: Indices	79
III. Democracy and Rule of Law Index Comparison	80
III.1 Index Selection	80
III.2 Index Backgrounds	81
III.3 Adequacy of Indices’ Methods and Implications for Across-Index Comparison	86
III.3.1 The European Commission’s Progress Reports	86
III.3.2 Freedom House’s Freedom in the World Reports	99
III.3.3 Bertelsmann Transformation Index	104
III.4 Implications of Indices’ Methodological Set-up for Empirical Analysis	110
IV. Political Bias Claims	112
IV.1 Academia	113
IV.1.1 Category 1 – Accession as a Political Decision	113
IV.1.2 Category 2 – Lack of Aptitude and Usability of Criteria	114
IV.1.3 Category 3 – Inconsistent and Wrongful Application	115
IV.1.4 Category 4 – Double Standards between EU Member States and Non-members	116
IV.1.5 Category 5 – Double Standards among Candidate Countries	117

IV.2 Member States' Positions towards Accession.....	119
IV.2.1 Croatia's Membership	119
IV.2.2 Turkey's Membership	120
IV.3 Media Coverage of Political Bias expressed by Political Actors and NGOs.....	121
IV.4 Negative Consequences of Political Bias Claim: Inner and Outer Credibility of the EU	123
IV.4.1 Inner Credibility of the Union	123
IV.4.2 Outer Credibility of the Union	124
IV.5 Repercussions of Political Bias on Member States and (potentially new) Candidate Countries	126
IV.6 Importance of reliability and Validity of the Progress Reports	127
IV.7 Eradicating Suspicion of Political Bias.....	128
V. Conceptual Framework	130
V.1 Dependent Variable I – Political Bias.....	130
V.2 Independent Variables for Political Bias.....	132
V.2.1 Reputation	132
V.2.2 Incentivization.....	135
V.3 Dependent Variable II – Consistency of the Indices' Reports	136
V.4 Independent Variables for Consistency among the Reports.....	136
V.4.1 Sources	136
V.4.2 Path Dependence	138
V.5 Hypotheses	143
VI. Methodology	144
VI.I Qualitative Content Analysis	144
VI.II Expert Interviews	157
VI.III Review of Internal COM Documents	159
VII. Results	160
VII.1 Qualitative Content Analysis Results	160
VII.1.1 Inductive Qualitative Content Analysis Results.....	160
VII.1.2 Scaled, Structuring Qualitative Content Analysis Results.....	162
VII.2 Interview and Internal COM Document Revision Results.....	184
VII.2.1 Progress Report Creation Process	185
VII.2.2 Experts' Opinions on the Political Bias Accusation vis-à-vis the COM	185
VII.2.3 Experts' Explanations for (Dis-)Similarities between their Institution's Results	188
VII.2.4 Reflections on the Independent Variables explaining the (In-)Consistency of Results.....	192
VIII. Discussion	194
VIII.1 Research Question, Hypotheses and Key Findings	194
VIII.1.1 Research Question and Relevance.....	194
VIII.1.2 Contribution to Research	195
VIII.1.3 Key Findings	195
VIII.2 Impact of Independent Variables	201
VIII.2.1 Independent Variables for Political Bias	201
VIII.2.2 Independent Variables for Consistency of the Indices' Results	202
VIII.2.3 Explanations introduced by Interviewees.....	204

VIII.3 Validity of Academia’s Critique of Progress Reports.....	208
<i>VIII.3.1 Political Motives of COM’s Reports</i>	209
<i>VIII.3.2 Aptitude and Usability of Copenhagen Criteria</i>	209
<i>VIII.3.3 Application of Copenhagen Criteria</i>	210
<i>VIII.3.4 Double Standards between Member States and Candidate Countries</i>	210
<i>VIII.3.5 Double Standards between Candidate Countries</i>	210
VIII.4 Embedding and Critical Reflection of Thesis within Academic Debate.....	211
<i>VIII.4.1 Relation to Similar Studies</i>	211
<i>VIII.4.2 Alternative Explanations of the Findings</i>	212
<i>VIII.4.3 The Study’s Limitations</i>	213
<i>VIII.4.4 Finding’s Implications</i>	214
IX. Summary	216
Bibliography	228

List of Figures

Tables

TABLE 1 - THIN VS. THICK RULE OF LAW DEFINITION ASSEMBLED FROM THE CONCEPTS OF MAGEN/MORLINO (2009:9), KLEINFELD (2006: 47), WENNERSTRÖM (2007: 78) AND TAMANAHA (2004/2006)	27
TABLE 2 - ROL VERSIONS ACCORDING TO TAMANAHA (2004: 91).....	29
TABLE 3 - COMPARISON OF CAROTHER'S (2006) AND WENNERSTRÖM'S (2007) ROL CONCEPTS.....	31
TABLE 4 – SIX TYPES OF EXTERNAL DEMOCRACY PROMOTION (CF. GIESENDORF, 2008: 124).....	48
TABLE 5 - CF. KOCHENOV (2008: 58/59).....	50
TABLE 6 - RULE OF LAW ELEMENTS INCLUDED IN GUIDANCE NOTE 2005, ENLARGEMENT PACKAGE TO DG ENLARGEMENT	60
TABLE 7 - SUMMARY OF EU ROL ACCESSION REQUIREMENTS FOR CANDIDATE COUNTRIES	72
TABLE 8 – QUALITY CRITERIA FOR MEASURING INDEX (MUNCK/VERKUILEN, 2002: 8).....	78
TABLE 9 - OVERVIEW OF BACKGROUND INFORMATION ON DEMOCRACY INDICES	85
TABLE 10 - GOALS OF JUDICIARY REFORM (GUIDE TO THE MAIN ADMINISTRATIVE STRUCTURES REQUIRED FOR IMPLEMENTING THE ACQUIS, 2005)	91
TABLE 11 - EUROPEAN COMMISSION (ENLARGEMENT PACKAGE 2005, GUIDANCE NOTE, ANNEX).....	92
TABLE 12 - SELECTION OF JUSTICE REFORM REPORTING GUIDELINE QUESTIONS (INTERNAL COM DOCUMENT, 2015)	96
TABLE 13 – FH'S 'FREEDOM IN THE WORLD' RULE OF LAW QUESTIONS.....	102
TABLE 14 – THE BTI'S RULE OF LAW QUESTIONS.....	107
TABLE 15 - FULFILLMENT OF MUNCK/VERKUILEN'S INDEX QUALITY TEST	110
TABLE 16 - AREAS OF CRITIQUE REGARDING COPENHAGEN CRITERIA FROM ACADEMIA.....	118
TABLE 17 - OVERVIEW OF SSQCA APPROACH	148
TABLE 18 - SPECIFIC ELEMENTS OF ROL SUB-COMPONENTS, GATHERED THROUGH INITIAL QUALITATIVE CONTENT ANALYSIS OF COM, FH AND BTI REPORTS (2005-2012)	152
TABLE 19 - CODING MANUAL	156
TABLE 20 - APPEARANCE OF ROL CATEGORIES IN ALL THREE INDICES	161
TABLE 21 - TABLE OF CROATIA'S AGGREGATED, OVERALL ROL SCORES.....	163
TABLE 22 - TABLE OF TURKEY'S AGGREGATED, OVERALL ROL SCORES.....	171
TABLE 23 – CROATIA AND TURKEY'S ASSEMBLED ROL SCORES (2005-2012).....	179
TABLE 24 - CROATIA AND TURKEY'S ROL SCORES BY SECTOR (2005-2012).....	180
TABLE 25 - CONSISTENCY OF RESULTS FOR CROATIA AND TURKEY	181
TABLE 26 - TURKEY'S SCORE CONSISTENCE BETWEEN THE COM AND BTI	183
TABLE 27 - CROATIA'S SCORE CONSISTENCE BETWEEN THE COM AND BTI	183
TABLE 28 - CROATIA'S SCORE CONSISTENCE BETWEEN THE COM AND FH (FIW)	184
TABLE 29 - IMPACT LEVEL OF INDEPENDENT VARIABLES	208

Illustrations

ILLUSTRATION 1 – THE EU'S ROL REQUIREMENTS IN EXTERNAL DEMOCRATIZATION (ARNULL: 2002)	36
ILLUSTRATION 2 - ELEMENTS OF THE COPENHAGEN CRITERIA (EUROPA.EU)	53
ILLUSTRATION 3 - ELEMENTS OF THE POLITICAL CRITERIA: DEMOCRACY AND THE RULE OF LAW (LUDWIG, 2011: 117)	54
ILLUSTRATION 4 - FIVE ELEMENTS OF THE EU ACCESSION ROL CONCEPT	61
ILLUSTRATION 5 - EU RULE OF LAW REQUIREMENTS GATHERED FROM CHAPTER 23 JUDICIARY AND FUNDAMENTAL RIGHTS	62
ILLUSTRATION 6 - MEASURING TREE ACCORDING TO MUNCK/VERKUILEN (2002)	76
ILLUSTRATION 7 - BTI ROL QUESTION SPECIFICATION	108

ILLUSTRATION 8 - DIAGRAM OF CONCEPTUAL FRAMEWORK 142

ILLUSTRATION 9 - MAYRING'S QUALITATIVE CONTENT ANALYSIS (2000) 149

DIAGRAMS

DIAGRAM 1 - CROATIA'S RoL EVALUATIONS (COM, FH, BTI)163

DIAGRAM 2 - CROATIA'S JUDICIAL INDEPENDENCE EVALUATIONS (COM, FH, BTI)..... 164

DIAGRAM 3 - CROATIA'S EFFICIENCY EVALUATIONS (COM, FH, BTI).....165

DIAGRAM 4 - CROATIA'S CRIME EVALUATIONS (COM, FH, BTI) 166

DIAGRAM 5 - CROATIA'S RIGHTS AND LIBERTIES EVALUATIONS (COM, FH, BTI)167

DIAGRAM 6- CROATIA'S RoL EVALUATION (COM, ALL FOUR SECTORS) 168

DIAGRAM 7 - CROATIA'S RoL EVALUATION (BTI, ALL FOUR SECTORS)..... 169

DIAGRAM 8 - CROATIA'S RoL EVALUATION (FH, ALL FOUR SECTORS)170

DIAGRAM 9 - TURKEY'S RoL EVALUATIONS (COM, FH, BTI) 171

DIAGRAM 10 - TURKEY'S JUDICIAL INDEPENDENCE EVALUATIONS (COM, FH, BTI) 172

DIAGRAM 11 - TURKEY'S EFFICIENCY EVALUATIONS (COM, FH, BTI) 173

DIAGRAM 12 - TURKEY'S CRIME EVALUATIONS (COM, FH, BTI).....174

DIAGRAM 13 - TURKEY'S RIGHTS AND LIBERTIES EVALUATIONS (COM, FH, BTI)..... 175

DIAGRAM 14 - TURKEY'S RoL EVALUATION (COM, ALL FOUR SECTORS)176

DIAGRAM 15 - RoL EVALUATION (BTI, ALL FOUR SECTORS) 177

DIAGRAM 16 - RoL EVALUATION (FH, ALL FOUR SECTORS)..... 178

Political Bias or Fair Play: How consistent are the European Commission's Progress Reports with view to the Rule of Law?

A comparative analysis using the examples of Croatia and Turkey.

I. Introduction

If a country seeks full EU membership it has to fulfill the Copenhagen criteria which comprise of a set of political, economic and legal requirements (cf. II.2.4). Its progress in meeting these criteria within the respective policy fields is annually monitored by the European Commission (COM) who summarizes its findings in so-called Progress Reports (cf. II.2.5).¹ These Reports have been established to make the accession process more transparent and feasible as well as guaranteeing that future Member States are mature enough for accession (Kochenov, 2004).² They are also supposed to be the basis for fair accession decisions.³

“Progress is measured on the basis of decisions taken, legislation adopted and measures implemented. This approach ensures equal treatment across all Reports and permits an objective assessment.”

(European Commission, Croatia Progress Report 2006: 5)

The thesis' vantage point

In order to make well-founded accession decisions, it is necessary for the Progress Reports to adequately describe and evaluate the developments within the candidate countries. However, the adequacy and neutrality of the Reports has been questioned repeatedly by scholars, the

¹ The official title of the Reports is „Regular Report on X-land's progress towards accession“ (European Commission, Guidance note 2005, Version 20 05 2005).

² The COM recommended “to stress the absolute priority of the Copenhagen political criteria before beginning and continuing the accession negotiations with any candidate country” (1999 Composite paper 30, In: Kochenov, 2004: 5).

³ „The Regular Reports are part of a permanent impact assessment loop which covers the entire accession process and which culminates in the final decision concerning the admission and date of admission of the candidate countries to the Union“ (European Commission, Guidance note 2005, Version 20 05 2005).

media, NGOs and parties to the accession process (cf. Chapter IV.). Moreover, the COM has been accused of political bias, ergo an intentional adaptation of evaluation results according to its political preferences regarding the (non-) accession of particular candidate countries. If such bias existed, the results of the COM's evaluations should noticeably deviate from those of other, independent measuring indices. This thesis aims at testing these allegations by examining the consistency of the COM's Progress Reports with the results of other, non-EU affiliated institutions' Reports.

The Thesis' Rooting in Previous Research

The salience of the research topic at hand has been acknowledged in 2011, when Viljar Veebel employed a similar approach to assess the COM's Reports reliability. He and I share the same premise, namely that unless it was politically biased, the Commission should produce similar results to independent Reports evaluating the same aspects. Veebel compared the COM's qualitative findings with the combined quantitative assessments of six other indices. He explained the COM's motives behind delivering partially differing results with political bias against two of his case studies (Veebel, 2011).⁴ However, Veebel's study has a significantly different range, depth and methodological approach than the thesis at hand.

It compares the COM's to other indices' combined findings with view to Bulgaria, Croatia, FYROM and Romania between 2004 and 2006 with view to five Copenhagen Criteria (governance efficiency, existence and quality of the RoL, level of corruption and efficiency of anti-corruption activities, efficiency of the legal system, and economic liberties and freedoms). It reviews pre-dated data for one of the case studies used for this comparison while spreading its attention to multiple Copenhagen Criteria rather than focusing in-depth on one aspect. Veebel's test also differs by treating corruption and efficiency as individual aspects of the legal system rather than mere sub-components of the RoL, as suggested by academia (cf. II.1.2, II.2.6).

This thesis will analyze two case studies in-depth over a longer, more current period of time (2005-2012) and with a clear focus on the RoL criterion and its sub-components, thus enabling a more detailed account of one specific political criterion.

Additionally, Veebel's research design and methodological approach differs significantly from the method chosen for this thesis. Instead of simply stating that the Reports are

⁴ Veebel compares the evaluation results of the IMF, the Worldbank, Freedom House, Bertelsmann, Transparency International, the Fraser Institute, and the Heritage Foundation.

independent, this thesis employs a preliminary analysis regarding their backgrounds, underlying interests and sources (cf. III.2). This turned out to be an important aspect because the preliminary test asserted independence among the indices whereas the later employed extensive qualitative content analysis as well as the expert interviews suggested interdependence, thus destabilizing the initial premise.

Veebel also acknowledges the problem of comparing the COM's qualitative findings with the quantitative data of the other indices. He addresses it by creating an ordinal ranking between the case studies based on the language that is used to describe their progress in the Reports. While in principle, this approach is legitimate, Veebel does not offer an explanation or coding manual describing his process which makes it difficult to track or reproduce his analysis. Consequently, in order to secure inter-coder reliability, this thesis provides a detailed account of how qualitative data is transformed into quantitative data (cf. VI.I). The same is true for his calculations regarding the combined rankings of the six test indices as well as the differences between the EU ranking and that of the other indices. It remains unclear what thresholds Veebel chooses in order to show significant or insignificant differences and why. Consequently, special emphasis is put within this thesis on documenting any methodological decisions and steps as outlined in detail in VI.I.

Veebel then attributes the differences he found in his analysis to the COM's political bias. Moreover, he claims that the causal explanation for the COM using qualitative instead of quantitative assessments is to "offer a possibility for a subjective image of applicant countries" (Veebel, 201: 20). However, he does not test or argumentatively substantiate this explanation. It also remains unclear whether he tested for other potential explanatory variables. Although this was not part of the initial research question, the expert interviews uncovered a range of aspects that could explain such differences. Moreover, in combination with the qualitative content analysis and the revision of internal COM documents, the interviews unearthed the origins of some systematic deviations among the results. By amending the content analysis with the analytical steps of interviewing and document revision, a more comprehensive picture of the political bias phenomenon, its origins, credibility and abatement can be drawn.

The Thesis' Conceptual Approach

In accordance with Veebel, this thesis assumes that if the COM's Progress Reports are politically biased (dependent variable) they should be inconsistent with the findings of

independent institutions evaluating the same aspects (indicator of the dependent variable). Therefore, the study is carried out by comparing the COM's Progress Reports' Rule of Law (RoL) component to those of Freedom House's Freedom in the World Report and Bertelsmann's Bertelsmann Transformation Index for the two case studies Croatia and Turkey (2005-2012). After finding that the Reports' results are mostly consistent – ergo the COM's political bias cannot be substantiated – their consistency is elevated to become the new dependent variable for which explanations are tested (cf. V).

Index Selection

The COM's Progress Reports were chosen because its recommendations serve as the basis for the Council's decision regarding the accession application of a candidate country and have yet been followed each time. Freedom House and Bertelsmann were chosen as test-indices because both their basic conceptual understanding of democracy and the RoL as well as their methodological approaches are similar enough to that of the COM to compare 'apples with apples'. Additionally, they are materially, normatively and ideologically independent from the accession process, as has been outlined in III.2. Since all three indices show different degrees of transparency regarding their methods and sources, in-depth expert interviews with their authors and coordinators will illuminate the question of actual independence. If all three indices happen to use the same sources, their chances of interpreting them similarly given their shared RoL understanding are substantial, thus reducing their independence.

Rule of Law Focus

The Copenhagen Criteria consist of a range of different economic, political and judicial aspects. The thesis' RoL focus originates from two discoveries: theoretically, most political scientists understand the RoL to be a precondition for liberal democracy (for details and dissenting opinions cf. II.1.3). Practically, it is most salient within the catalogue of accession criteria. Since formal democracy and democratic stability have to be established prior to being granted the candidate status, the accession negotiations are mostly contingent on the fulfillment of the RoL requirement (Kochenov, 2004).

Case Selection

Croatia and Turkey serve as case studies because their simultaneous start of accession negotiations granted their accession process to take place in the same political and economic

climate which has been crucial for the EU's general attitude towards accessions.⁵ Moreover, strong opinions on Croatia's accession and Turkey's non-accession have been publicly voiced by Member States during their negotiation processes. These give rise to the apprehension that the COM might not have made its evaluation decisions independently.

Research Gap

The findings of this research project advance previous academic inquiries in several areas. As outlined in more detail in **Chapter II- Literature Review**, research has been conducted in the fields of democracy and RoL in terms of their definitions, with view to external democratization and RoL promotion strategies focused on actors and their methods, specifically the European Union, and the types and qualities of indices measuring the progress within external democracy and RoL promotion. Within the theoretical academic debate, the thesis will amend existing research by providing an up-to-date EU accession RoL concept and checklist by conducting expert interviews and reviewing internal COM documents. Furthermore, it will summarize and categorize the main areas of critique vis-à-vis the Progress Reports as issued by academia, candidate countries, and the media. Third, it will examine each indices' methodological quality prior to their comparison based on a comprehensive analytical framework (Munck/Verkuilen).

With view to empirical research, the thesis introduces a method to make multiple methodologically different indices comparable on one, common scale. This method can be used for future comparisons. The analysis generated a surprising finding: highly consistent results among all indices. These results refute the political bias claim vis-à-vis the COM. Based on the findings of the expert interviews, attention is drawn to the actual level of independence between the reports. While this has practically been assumed unscreened both by academia and users of the indices, the thesis proves high degrees of interdependence.

Research Goals

The research goals include the conceptual and methodological understanding of the COM's Progress Report creation process. Furthermore, the thesis addresses the question of the COM's potential political bias. Additionally, it will provide explanations for similarities and

⁵ During the enlargement rounds of 2004 and 2007, one could detect an 'enlargement hype' within the EU which led to the speedy acceptance even of candidate countries who did not yet fulfill the Copenhagen criteria (Romania and Bulgaria). However, the admission of said countries - whose democratic immaturity proved troublesome later - in combination with "the EU's crisis of confidence and accountability after the failed attempt to introduce the Constitutional Treaty" has led to an "enlargement fatigue" which signifies the Union's reluctance to include further new MS (Szołucha, 2010).

differences of results among the Reports. The thesis will also shed light on the variety of aspects that influence the evaluation process and outcomes. It will provide a key to understanding the complexity (and sometimes simplicity) of factors that weigh into final scoring decisions, ranging from hard-to-measure socio-cultural priming to the disdainful copying of results. Furthermore, the goal of this thesis is not to discredit or glorify any of the evaluating institutions, but to show the strengths and weaknesses of their evaluation methods and processes as well as putting their results into perspective by making them comparable to one another. By pointing out the origins of and double standards behind the political bias claim, the thesis aims to introduce policy solutions to address the issue effectively.

Research Design

When seeking to test the COM's Reports' consistency, two options arose: absolute testing and relative testing. Absolute testing would require the reproduction of the research conducted by the European Commission, comparing those results to the ones of the COM's Reports, and drawing conclusions thereof. Relative testing draws on the Reports of other, non-EU institutions and compares their findings. Furthermore, it can examine the origin of (in)consistencies. The latter option was chosen because a greater number of varying sources improves the quality and reliability of the comparative analysis' results. Also, the sheer amount of resources that has been used by the European Commission in producing the Reports cannot be mimicked within the limited bounds of a doctoral thesis.

The thesis originated from the accusation that the European Commission's Progress Reports are politically biased. If this was true, evidence should be found in the Reports themselves which could be uncovered by comparing them to other, independent Reports. Thus, the political bias thesis rests on the premise that the Reports' results issued by the COM differ noticeably from those of independent institutions. This premise has to be tested before examining potential explanatory variables.

In order to test the premise, the evaluations of all three indices were made comparable on one, common scale. Initially, a shared RoL concept of the indices was generated by using an inductive qualitative content analysis. It consisted of four criteria (Judicial Independence, Efficiency, Rights & Liberties, Crime, VII.1.1). These were subsequently tested and scored for each of the Reports on the new scale by using the scaled, structuring qualitative content analysis (for the detailed process see **VI. Methodology** Chapter, specifically the Coding

Manual summarized in Table 19). Once these comparable scores were generated, an overall RoL score and scores for each of the four RoL elements were calculated for each Report, thus creating a basis for comparison (cf. Tables 23 and 24 in VII.1.2). The comparison proved that the three indices deliver mostly consistent results meaning that the initial premise for the COM's political bias could not be validated.

After generating the surprising finding of striking similarities among the indices' results, their level of consistency was elevated to become the second dependent variable. Consequently, potential reasons for the similarities were examined. Explorative, semi-guided in-depth expert interviews with the Reports' authors and coordinators and well as the revision of internal accession-relevant COM documents helped to shed light on the phenomena of both political bias claim and origin as well as the actual similarity of evaluation results. The interviewees were asked to reflect on their own understanding of political bias, their opinion of the political bias claim towards the COM and to provide own explanations for (in-)consistencies between the indices' results. Furthermore, they were asked to comment on the introduced independent variables. By combining theoretic considerations with empirical evidence, the thesis aims to present a comprehensive understanding of the phenomenon of the political bias claim as well as provide an answer to how and why the democracy indices compared produce highly consistent results.

The Material

The empirical analysis is based on the respective official and public annual Reports of the COM (Progress Reports), Bertelsmann (Bertelsmann Transformation Index) and Freedom House (Freedom in the World) for Croatia and Turkey from 2005 to 2012. From these Reports, all sections pertinent to the RoL evaluation of the given case study have been included in the analysis. Once the RoL elements that were shared by all three indices have been identified, sections pertaining to other RoL elements were not included in the analysis any longer. The shared elements were designated by computing how often all possible RoL elements were addressed in the Reports and choosing those that have been represented in at least 75% of all Reports. The overall RoL score was then calculated by adding the scores of the individual RoL elements and dividing it by the number of elements addressed in each report (cf. VI.1).

Summary of Results

The main results of the analysis are that the COM's Reports are – in fact – mainly consistent with the evaluations of FH and BTI. Consequently, a political bias of the COM could not be substantiated. The high level of consistency among the results can be attributed to two aspects. First, the tested indices use mostly identical sources and circle-reference one another. Due to its superior access to (re-)sources in terms of finance, personnel and access to authorities, the COM's reports are more often referenced by FH and BTI than vice versa. Secondly, all indices base their evaluations on their previous reports' assessments which leads to path dependence in their scoring. Combined with the circle-referencing, all indices' scores become increasingly more similar and stable. Some of the inconsistencies found are systemic, for example when FH and BTI borrow their factual information from earlier COM Reports. Other times, they are unsystematic and can be explained through different interpretative emphasis according to the aims of the respective indices.

While politically biased evaluations by the COM could not be substantiated through the analysis, the thesis uncovered certain aspects that comprehensibly explain why there is such a strong perception of politically biased assessments in the first place. These include the intransparency of the COM's evaluation procedures, the informal consultations with Council staff and the involvement of politically appointed officials in the finalization of the Reports.

On the meta-level, the thesis proves that there persists a dangerous double standard. While political institutions are expected to make politically biased decisions, other independent institutions are perceived to be neutral and therefore more trustworthy. First, it cannot be proved that political institutions automatically generate politically biased evaluations. Such bias could not be substantiated for the COM. Secondly, the superiority of independent indices cannot be automatically deducted, it strongly depends on the quality of their methodology and the independence of their sources. The great interdependence of the COM, FH and BTI shows that comparisons among political and independent institutions' indices may be misleading and that - if politically biased assessments were made by the COM - they would most likely be unconsciously imported into the independent indices' assessments, too.

Limitations and Future Research

The thesis discovered a range of the COM's structural decisions regarding the evaluation process which feeds the political bias perception. At the level of Report-drafting, the COM's DG NEAR ENLARGEMENT officials gave the impression of working in a very structured

and procedural way seeking objective evaluations – just like the political scientists at Freedom House and Bertelsmann did. But, the dual capacity of the Commission as both an evaluator and distributor of financial development support may indirectly affect certain evaluations. Consequently, conducting a detailed analysis of said process and the actual influence on evaluation results could further deepen the understanding of the political bias claim.

As has been discovered during the expert interviews, changes are being made to the COM's Reports along the hierarchy chain. However, it remains unclear if those changes are part of the necessary harmonization process - as is carried out by the other indices as well - or if they may be attributed to informal or indirect forms of external influence. However, since the final COM Reports serve as a guidepost for FH and BTI - if politically biased evaluations were included - they could “spill over” to their Reports, too. Representatives of both indices emphasized during the interviews that they only use the factual information provided by the COM while presenting their own interpretations thereof. Nevertheless, there remains leeway for unconscious and inadvertent biases which could be tested in subsequent studies.

Structure

The thesis consists of nine chapters. This introduction is followed by **Chapter II - Literature Review** where the role of the RoL in the EU's external democracy promotion activities will be illustrated based on previous research. Also, a EU accession RoL concept is developed which serves as a benchmark for the subsequent analysis. Furthermore, an overview of existing research on measuring RoL developments is offered and existing research gaps are highlighted which this thesis aims to close. Additionally, academic literature on the perceived political bias of the Commission's reporting process is reviewed and categorized, thus proving the salience of the thesis at hand. In **Chapter III - Democratization and Rule of Law Index Comparison** all three evaluation institutions (COM, Bertelsmann Foundation, Freedom House) are scrutinized with regards to their origins, goals and/or ideological frameworks and their financial set-up, their underlying RoL concept and methods used to evaluate it. This is necessary to detect whether an ideological bias of some sort is a pre-existing condition based on the indices' backgrounds. **Chapter IV** scrutinizes the **Political Bias Claims** against the COM both with view to its propagators as well as its consequences for the European Union. Afterwards, **Chapter V - Conceptual Framework** outlines all relevant variables, explains their selection criteria, defines them and clarifies how they are tested throughout the analytical process. **Chapter VI - Methodology** elaborates on the research design and operationalization

of the analytical categories. It also explains why and how Mayring's scaled, structured, qualitative content-analysis (SSQCA) has been used to conduct the comparison. Additionally, the case selection is explained. Furthermore, the set-up of the semi-structured in-depth expert interviews and the body of internal COM documents is outlined. In **Chapter VII - Results** the findings of the qualitative content analysis as well as the expert interviews and the internal document revision are presented. **Chapter VIII - Discussion** provides the reader with a structured interpretation of the comparison's results based on the hypotheses outlined in Chapter V. Last, **Chapter IX - Summary** recapitulates the key findings, answers the research question, and outlines how other scholars can build upon the research at hand. Moreover, it will provide specific policy recommendation on how to address the political bias perception and thus improve the EU accession process with view to transparency, feasibility, and most importantly: credibility.⁶

⁶ While the author acknowledges that in EU *realpolitik* specific interests of Member States matter more in decision-making than the outside perception of staying true to one's public assertions vis-à-vis candidates, the EU's biggest political capital is the promise of membership. Consequently, as will be argued in further detail in the Introduction, the EU jeopardizes its political capital if it justifies its decisions to (not) accept a country into its ranks based on "wrongful" Reports.

II. Literature Review

This literature review consists of three parts. First, democracy and the RoL will be defined and their relationship will be examined (II.1.). Because this thesis analyzes the rule of law, it will further be conceptually expanded and categorized before carving out the actual EU accession RoL definition. Secondly, external democracy and RoL promotion will be scrutinized in terms of actors, motives and instruments, once again focusing on EU activities (II.2). Third, RoL measuring tools will be analyzed (II.3).

II.1 Liberal Democracy and the Rule of Law

II.1.1 Defining Liberal Democracy and Democratization

This thesis seeks to examine and explain possible differences between Rule of Law (RoL) evaluations of the European Commission, Freedom House and the Bertelsmann Transformation Index for Croatia and Turkey between 2005 and 2012. The RoL development is fostered and tested as part of the EU's accession strategy which entails the Copenhagen criterion Democracy and the RoL. Therefore, a brief introduction into the concepts of liberal democracy and democratization will be offered at this point. A liberal democracy is a type of government that grants its citizens certain civil rights and liberties. The government is elected in free, fair and competitive elections. There persists a separation of powers between the three state branches. The protection of said freedoms, rights and power limitations is ensured and guarded by the rule of law. According to Diamond and Morlino (2005: xi), the quality of a democracy can be measured in terms of procedures (e.g. elections), results (e.g. satisfaction of electorate with governance), and content (e.g. extensive liberties and political equality). The higher the degree of transparency and sovereignty and the broader the range of freedoms, the higher the overall quality of the liberal democracy becomes. The process of arriving at such an institution is called democratization which conceptually envelopes the "stabilization, routinization, institutionalization and legitimization of patterns of democratic behavior" (Pridham, 2006: 378). During the course of democratization, democratic institutions and processes are being established, the adherence to democratic norms and values is practiced, and ideally, a democratic culture is installed (Tunkrova/Saradin, 2010: 36). Democratization can evolve 'naturally' from within a country or can be imported from another actor, as is the

case with external democracy promotion.⁷ By demanding of its candidate countries to implement strategies that foster democracy and the RoL in order to gain full membership (conditionality principle), the EU employs an external democratization approach. Although the process of external democratization is not the focus of the thesis, a short overview of the academic field will be provided in II.2 in order to familiarize the reader with the accession process and the principles underlying the creation of the Progress Reports.

II.1.2 Defining the Rule of Law

As will be explained and emphasized later on, the conceptualization of a term is crucial when trying to empirically evaluate its quality. In order to carve out and embed the EU accession RoL definition, a general overview of the academic RoL debate will be provided beforehand.

“The rule of law is almost never carefully defined as a concept; users of the expression allude to meanings they assume to be clear and objective but that are not so (...) naturally, this would be a perfectly innocent and common phenomenon, not worth an inquiry, were it not for the weighty political implications of the phrase in different contexts.”

Mattei & Nader, 2008 (In: Sannerholm, 2009: 65)

Even though Shklar and Mattei/Nader illustrate how difficult it is to pinpoint what the RoL really means, a myriad of scholars have dedicated their research to the phenomenon and come up with definitions, concepts and categorizations which, at the core, reflect similar ideas on the matter. Hence, this part will expose the reader to said ideas in a structured manner, moving from the broad definitions to the specific, operational elements.

General Definition(s)

Even though basic ideas on the RoL have been expressed already by Aristotle⁸ or Montesquieu, English Law Professor A.V. Dicey’s definition of the concept from 1885 has supposedly coined

⁷ „Naturally“ occurring democratization processes can be caused by contagion, diffusion or Zeitgeist as suggested by Giesendorf (2008: 124) (for details see II.2)

⁸ “It is better for the law to rule than one of the citizens...so even guardians of the law are obeying the laws” (English translation, In: Bingham, 2010: 3)

the term. In his view, the RoL contains the three elements clarity, equality, and supremacy of Common Law as defined below:

Clarity: *Only a “distinct breach of law established in the ordinary legal manner before the ordinary courts” can be punished.*

Equality: *“No man is above the law” in the sense that “every man, whatever be his rank or condition, is subject to... the same law, administered in the same courts.*

Supremacy of Common Law: *The Rule of law has to be considered a “special attribute of English institutions” because “general principles of the constitution are...the result of judicial decisions” which means that there is no set of laws which is above the courts as is the case with formal constitutions*

(Bingham, 2010: 3-5)

Whereas most other scholars agree with the first two criteria, they do not concur that RoL systems of countries with written constitutions are inferior to the British model.

Instead, they focused on adding to and clarifying Dicey’s definition. Fallon (1997: pp. 9-19) adds to Dicey’s clarity aspect that the RoL “allow[s] people to plan their affairs with reasonable confidence that they know in advance the legal consequences of their actions”. Raz (1977: 198-201) concretizes this notion by demanding all laws to be “prospective, open and clear” as well as “relatively stable”⁹ while Fuller adds among others the criteria of public promulgation, consistency with other laws and the possibility of abidance.¹⁰ Hayek (1944: 75) emphasizes the results of said principles, namely that the governments ligation to fixed rules “make[s] it possible [for the people] to foresee with fair certainty how the authority will use its coercive powers in given circumstances”. All additional aspects as proposed by O'Donnell (2004) and Carothers (2007: 4) result from those basic principles such as the “right to a fair, prompt hearing and [...] presumed innocence until proved guilty” as well as the competence and efficiency of the judicial institutions and the impartiality and independence of the judges. When looking at all

⁹ Raz created a list of eight principles which also include open and clear procedures for rule-making, the independence of the judiciary, open and fair hearings, the courts’ revision powers, equal and easy access to the courts as well as adequate discretion of the crime preventing agencies (Bradley/Ewing, 2001: 98). With view to RoL categorizations, Raz was a representative of the formal RoL concept.

¹⁰ For a comprehensive list which partly overlaps with Raz’ ideas see Sannerholm (2009: 34).

the elements of the RoL suggested by its scholars, Carothers' (2007: 19) 'shorthand form' provides a good general understanding of the concept:

"[The rule of law means that the law is] applied fairly, uniformly, and efficiently throughout the society, to both public officials as well as ordinary citizens, and to have law protect various rights to ensure the autonomy of the individual in the face of state power in both the political and economic spheres."

Categorizations

The abundance of different RoL definitions can be categorized in various ways. This part will suggest two categorizations: the thin vs. thick dichotomy (enhanced by Tamanaha's thin to thick spectrum) and the RoL promotion-relevant depth-of-reform approach.¹¹ It will provide a structured synopsis of the academic debate by sub-dividing each category (for summaries see Tables 1, 2 and 3). Regarding the thin vs. thick comparison, the category will be examined based on the labels it has been given by different scholars, its general meaning, its specific elements, potential variations (Tamanaha's spectrum) and their respective proponents. With view to the reform-depth approach, two scholars' models will be contrasted who use different labels but whose analysis is consistent content-wise.

Thin vs. Thick Rule of Law

The main proponents of the *thin* (also referred to as *formal, negative, institution-based or rule-book*) RoL approach are Dicey and Raz.¹² In their opinion, the RoL merely provides specific principles based on which laws are being made, promulgated, enforced, contested and reviewed. General elements of the thin RoL conception include that laws are *clear, publicly known* and

¹¹ As Sannerholm (2009: 32) correctly asserts, a dichotomous division does not do justice the different characteristics of the concept and the smooth transition from one pole to the other. Moreover, Wennerström (2007: 77) explains how even formal judicial procedures can have normative functions as is the case with "fundamental rights pertaining to criminal procedure" which also refutes the dichotomy-thesis. Consequently, I will refer to it as a spectrum.

¹² The terms *thin* and *thick* were proposed by Magen/Morlino (2009: 8) and correspond to the following labels: the "formal" vs. "substantive" divide was coined by Craig (1997), Selznick (1999) distinguished "negative" from "positive" RoL, Dworkin separated into "rule-book" and "rights". A third conception option is the "functional" one which can either be accepted as an independent or a mere variation of the substantive definition (Wennerström, 2007: 60). It focuses on the measuring quality, quantity and "specific functions of a legal system" and was developed as an evaluation method to assess the "output and capacity in concrete legal systems" based on the contextual definitions of the RoL (Stephenson, 2001, In: Wennerström, 2007: 83).

settled as well as *not retroactive*. The judiciary is supposed to be *trained, efficient and independent* and able to *review government acts*. Additionally, there exists an authority that is able to *enforce the laws* (Kleinfeld, 2006: 47, and Wennerström, 2007: 78). They oppose the notion that the RoL contains any moral values. Specifically, Raz (1977: 96) argues that the RoL can and must not be equated with a tool that fosters democratic or otherwise morally-normative societal progress.

“The Rule of Law is just one of the virtues which a legal system may possess and by which it is to be judged. It is not to be confused with democracy, justice, equality (before the law or otherwise), and human rights of any kind or respect for persons or for the dignity of man. A non-democratic legal system, based on the denial of human rights, on extensive poverty, on racial segregation, sexual inequalities and religious prosecution may, in principle, conform with the requirements of the Rule of Law better than any of the legal systems of the more enlightened western democracies.”

While this might sound intuitively wrong to a scholar socialized in a western liberal democracy, Raz' statement becomes more comprehensible when looking at his definition of the RoL. He formed two groups of elements that constitute the (*thin*) rule of law: the first aims at predictability and the second one proposes safeguarding mechanisms against arbitrary exercise of power. Group one reiterates some of the characteristics including *prospective, clear and open laws; stable legislation, and a stable legislative process*. In fact, authoritarian regimes or dictatorships may possess such attributes and due to infrequent changes of government may even be able to provide more stability than democratic regimes whose power alterations cause more legislative flexibility per definition.

However, when looking at Raz' second group (*judicial independence, respect for principles of natural justice including fair hearing and freedom of bias, review powers of the judiciary, access to justice, limitations to discretionary power of law-enforcing agencies*) it becomes apparent that his observation may be misleading. In fact, most suppressive political regimes have very little interest in an independent judiciary in order to protect their own interests. Countries like Russia or China are known for law enforcing agencies with wide discretion, for not granting certain society groups the same access to the judiciary or holding show trials where the judges are biased. Even though said nations are internationally criticized for violating basic

human rights, which in Raz' view does not have to be protected by the rule of law, they could not even fulfill his *thin, formal* rule of law conception.

In fact, most RoL scholars argue the exact opposite, namely that even the fulfillment of formal conditions is a value in itself and therefore “good” and that the RoL is able to “safeguard against totalitarian tendencies” (Hayek, 1944: 54). The perception that the RoL is tied to “elements of political morality...with institutional practicality to protect political and civil liberties” (Magen/Morlino, 2009: 9) is referred to as a *thick (substantial, material, positive, ends-based, rights)* definition. The five key elements of this perception are that the *state (officials) abide by the law*, that *equality is ensured before the law*, that *law and order* are granted, that the judiciary is *efficient and impartial*, and that *civil liberties and human rights* are upheld (Kleinfeld, 2006: 35).¹³

As becomes visible when looking at the different elements of the *thin* and *thick* rule of law definitions, they overlap to some degree (*independence/impartiality of the judges, enforcing laws/law and order*) which supports the previously mentioned notion that the RoL must be understood as a spectrum rather than a dichotomy.

Waldron, Fallon and Dworkin all agree that the great advantage of the (*thick*) RoL definition is that it creates certain formal guidelines that citizens, executive and judiciary all have to abide by. This in turn “protects against anarchy and the Hobbesian war of all against all” (Fallon, 1997: 19) and at the same time can “correct abuses of power by...governing through law – which is far more apt to protect us against abuse than (say) managerial government or rule by decree” (Waldron, 2008: 11). Fuller emphasizes that the rule of law is in itself a “moral good” because if formal rules are “upheld, a state is more likely to have fair and just laws because people can participate in, act and speak on issues of law-making” (2004: 46).

A synopsis of the thin vs. thick comparison as conducted above can be found in the following Table 1.

¹³These five goals were suggested by Rachel Kleinfeld (2006: 35) who emphasized that this notion correlated with RoL building approaches by NATO, EU and USAID but “cannot simply be stated as given [because] rule-of-law ends are so contested and historically determined”. She asserts that “they must be understood as varying greatly by context, culture, and era” (Ibid.). However, since the thesis will specifically address one RoL promoter in one cultural context within a very limited time frame, the ends assembled by Kleinfeld will be used as guiding posts.

Thin vs. Thick Rule of Law Definition		
Thin Rule of Law	Criterion	Thick Rule of Law
Thin, formal, negative, institution-based, rule-book	Labeling	Thick, substantial/material, positive, ends-based, rights
Mere focus on process by which norms are enacted in laws and promulgated, morally neutral regarding the goals the instruments strive to fulfill	Meaning	Value-laden Rule of law concept containing combining elements of political morality with institutional practicality to protect political and civil liberties
<ul style="list-style-type: none"> • Laws publicly known and settled • Retroactive legislation prohibited • Trained, efficient and independent judiciary • A force able to enforce laws • Judicial review of government acts 	Elements	<ul style="list-style-type: none"> • Making the state abide by the law • Ensuring equality before the law • Supplying law and order • Providing efficient and impartial justice • Upholding human rights
<ol style="list-style-type: none"> 1. Rule by law 2. Formal legality 3. Democracy and legitimacy 	Tamanaha's corresponding Versions rising with content density¹⁴	<ol style="list-style-type: none"> 1. Individual rights 2. Right of dignity and/or justice 3. Social welfare rights
Dicey (1959), Raz (1979)	Proponents	Hayek (1944), Shklar (1987), Fuller (2004), Waldron (2008), Dworkin (2001)

Table 1 - Thin vs. Thick Rule of Law Definition assembled from the concepts of Magen/Morlino (2009:9), Kleinfeld (2006: 47), Wennerström (2007: 78) and Tamanaha (2004/2006)

¹⁴The six different models have been proposed by Brian Tamanaha in his seminal work “On the Rule of Law” (2006) and describe a range from the thinnest to the thickest type of formal RoL and respectively the thinnest to thickest version of the substantial RoL concept.

Acknowledging the *thin* vs. *thick* divide to be a spectrum, Brian Tamanaha (2004) created a graded RoL model that shows how – by adding individual elements – the *rule by law* can evolve into a *social-democratic rule of law*. In his model, there are respectively three qualitative types of both the *thin/formal* and the *thick/substantive* version. The thinnest formal version is the *rule-by-law* which merely means that laws are the instruments of government action. This version does not even fulfill Dicey's and Raz' rule of law criterion because no procedural rules or limits are set for the government regarding the making of said laws. Consequently, the next, thicker formal version is the *formal legality* which adds the criteria of laws having to be *general, prospective, clear and certain*. This version is consistent with Dicey's and Raz' RoL concept. At this point, a dictatorship or authoritarian regime can still fulfill the RoL definition. Adding the element of *consent to determine the content of the law* leads to Tamanaha's thickest formal version, the *democracy and legality* one which focuses on a democratic procedure but may include 'good' or 'bad' content. Naturally, Dicey and Raz would not have included this version into their RoL definition because democracy is already a value-laden concept whereas the rule of law is supposed to be value-free.

Tamanaha continues with the thinnest substantive version, where *individual rights* are added (*property, contract, privacy and autonomy rights*). When the *human dignity* becomes a supreme value heading and limiting all government and judicial actions, he refers to the thicker substantive version of *right of dignity and/or justice*. The thickest substantive version he introduces is the *social welfare* one which implies a “positive duty on the government to improve the citizens' conditions for living” (Wennerström, 2007: 96) and thus help foster *substantive equality* within the community.

As an analytical tool, Tamanaha's model is very helpful in that it enables students to assess countries based on specific criteria on a “graded scale” rather than in a black-and-white manner like the thin vs. thick dichotomy. However, it brings up questions on how to evaluate dictatorships or authoritarian regimes with sophisticated individual rights or social welfare programs (e.g. Singapore, former GDR).

RoL Version according to Tamanha (2004)			
Thin -----→ Thick			
Formal Versions			
Formal Versions	1. Rule-by-Law	2. Formal Legality	3. Democracy + Legality
Meaning	Law as instrument of government action	General, prospective, clear, certain laws	Consent determines content of law
Substantive Versions			
Substantive Versions	1. Individual Rights	2. Right of Dignity and/or Justice	3. Social Welfare
Meaning	Property, contract, privacy, autonomy	Human dignity as supreme value	Substantive equality, welfare, preservation of community

Table 2 - RoL Versions according to Tamanaha (2004: 92)

Reform depth categorization

The RoL developments that are measured and assessed by the indices tested and compared in this thesis are influenced by the reforms prescribed by the European Commission prior to accession. Therefore, the following RoL categorization will address the level of rule adaptation and implementation by the target states. Thomas Carothers' RoL categorization (2006: 7) focuses on the external reform promotion process and makes divisions based on the depth of the reforms carried out in the target state. His “thinnest” reform depth is the pure *law focus*, superseded by the *institution-focus* and headed by the “thickest” reform depth, *government compliance*. Content-wise, this model correlates with Wennerström's (2007: 63-66, 100-101) division into *declaratory*, *persuasive*, and *obligatory* operational RoL conceptions wherefore both models are compared in Table 3 below.

Wennerström describes the assertion of the government to *install certain rules and procedures* as *declaratory* because it does not require any change of legislation yet but signals the wish for adaptation. Carothers' *law-focus* reform is similar to this concept in that it addresses the *revision and redrafting of laws* without leading to final implementation results yet. At the center of the

law-focus reforms are economic law (“bankruptcy, corporate finance, taxation, intellectual property, financial markets”) whose revision is important to attract foreign investors as well as criminal law (“expanding the protection of basic rights in criminal procedure codes, modifying criminal statutes to cover new problems such as money laundering and electronic-transfer fraud, and revising the regulation of the police”) which usually caters to the interests of RoL donors (all quotes Carothers, 2006: 7).

The second level of reform depth is the *persuasive* (Wennerström) or *institution-focus* (Carothers) one. Here, the government moves from making declaratory statements and creating laws to *implementing the new rules and procedures* in the law-related institutions. In detail, this includes the *training of law professionals* thus increasing their competence, efficiency and accountability. Furthermore, *judicial independence* is strengthened through higher salaries, reform packages that address potential enmeshment with other state branches as well as the creation of ethics codes and professional standards. Further reform goals are to strengthen the “legislature, tax administration, and local governments” (Ibid.).¹⁵

Step three towards a deep RoL reform is to make rules *obligatory* (Wennerström), that is to ensure institutional (particularly *government*) *compliance* (Carothers) with the new laws. In order to *create judicial independence in practice*, the selection, evaluation, promotion and denunciation process of judges has to be relieved of executive influence, the judiciary's budget must be sufficient and administered by a Council consisting mostly of representatives of the judiciary and the Ministry of Justice's influence on staff decisions, budgeting and the judicial process as such must be limited. The following Table 3 summarizes both concepts:

Carothers	Depth of reform	Wennerström
<ul style="list-style-type: none"> • Revision/redrafting • Focus on economic domain • Focus on criminal law • Revising police regulation 	I. Law-focus vs. declaratory	Assertion of future installation of certain rules and procedures
<ul style="list-style-type: none"> • Focus on competence, efficiency and accountability 	II.	Contractual obligations aimed at installing certain rules and procedures

¹⁵ For a detailed concept see Chapter II.2.6, Judicial Independence and Judicial Impartiality.

<ul style="list-style-type: none"> • Professional training • Increase staff and salaries • Create prof. Standards/ ethics code • Strengthening legislatures, tax administrations and local governments 	Institution-focus vs. persuasive	
<ul style="list-style-type: none"> • Increase judicial independence through better salaries and selection process 	III. Government-compliance vs. obligatory	<ul style="list-style-type: none"> • Existence of mechanisms for control of law implementation • Mechanisms able to impose responsibility for non-compliance (political, legal, economic)

Table 3 - Comparison of Carother's (2006) and Wennerström's (2007) RoL Concepts

As the preceding part has shown, there are significant conceptual similarities between the RoL definitions and categorizations within the academic debate although they have been termed differently. There is a trend towards the more substantive understanding of the RoL, particularly within democratic societies where the facilitation and protection of civil rights and liberties became a *raison d'être* for the state. Against this background, the EU's democracy and RoL conceptualization will be embedded in the subsequent sections (cf. II.2.4 and II.2.6).

II.1.3 Relationship between Liberal Democracy and the Rule of Law

This part will substantiate that the RoL is - by most scholars - considered a precondition for liberal democracy and therefore particularly relevant for the democratization process, thus supporting the focus of the thesis.

“The rule of law bakes no bread, it is unable to distribute loaves or fishes (...), and it cannot protect itself against external assault, but it remains the most civilized and least burdensome conception of a state yet to be devised.”

Today, democracy is internationally acknowledged to be the most civilized, advanced and prominent conception of a state.¹⁶ Hence, the citation above can only be true if the RoL was inseparably intertwined with democracy. In fact, the salience of the RoL within the democracy promotion process has been emphasized for the last twenty years by scholars and practitioners alike. It was exaggerated up to a point where one of its biggest advocates and most important academics in the field – Thomas Carothers (2007) – warned students of the matter to not expect it to be the “silver bullet” in solving all the world’s problems. Even though most scholars agree that both concepts are strongly interrelated, there are also some critics who describe their coupling as a priori artificial.¹⁷ While Rosenfeld criticizes that it remains unclear what role the RoL plays in sustaining constitutional democracy “or how it might ultimately contribute to [its] legitimacy”, Kochenov and Kelsen argue that while democracy describes the relationship between the governing and the governed, the RoL focuses on that between “the governing and the law” (Kochenov, 2008: 112) and therefore addresses completely different relations. This line of argumentation can be challenged based on the fact that democracy requires the trust of the governed in the government to abide by the rules and not to abuse their executive powers to limit the granted freedoms of the governed or for promoting their own interests. Thus, both relationships as described by Kochenov and Kelsen are actually intertwined and create a ‘trust-triangle’. Although the critics deny a causal relationship between the RoL and democracy, most scholars agree that both concepts enrich each other. Consequently, their ideational and practical links will be deduced below in order to show that the academic focus particularly within the EU accession process is justified.

The Council of Europe believes that democracy can originate from adequate RoL enforcement (Piana, 2010: 2). But how exactly does a functional RoL contribute to democratization? According to Hobbes (1651), the right of existence for a state is the protection of its citizens. In order to overcome anarchy, a political state without rules where everyone can be considered everyone else’s enemy, a social contract was formed between citizens and their

¹⁶ Even most authoritarian or dictatorial regimes call themselves democratic republics.

¹⁷ O’Donnell (1998), Bratton/Chang (2006), and Thier (2007) all agree that the RoL is *primes inter pares* among the dimensions of democratic quality (Magen/Morlino, 2009: 7)

rulers which legitimizes the power monopoly of the state.¹⁸ Within the liberal democracy, this protection encompasses the guarantee and defense of the aforementioned rights and freedoms. Consequently, if a state cannot protect those rights, it loses the moral and theoretical foundation for exercising power over its constituents. The role of the RoL within this equation is to secure that both the procedural guarantees (e.g. voting procedures, the making of the law, due process etc.) as well as the actual individual rights (e.g. freedom of speech, expression and assembly) are being enforced. By creating a dependable track record of predictability regarding how the state will act vis-à-vis the citizens and how it will settle disputes among them, a culture of law and order can be created. For example, treating everyone alike before the law and punishing crimes non-dependent of ethnicity, religion, age, or position, as well as holding everyone including state officials and powerful non-state actors accountable for wrongful behavior will help to generate trust in state institutions. Only if state institutions are generally trusted, people will abide by and subdue themselves to them, which in turn is necessary for the state to have effective, democratic control over its citizens. The longer such a culture is intact, the more stable it becomes and the stronger the trust of the citizens in the state's power monopoly grows. The increased compliance with state-made rules generates a higher level of stability and security within the society. If the state can be trusted to act in the interest of its citizens and reliably but fairly punishes those who don't, this also decreases instances of crime, corruption, and the usage of vigilante justice.

The RoL not only serves the establishment but also the consolidation of liberal democracies.¹⁹ It does so by actively protecting its values against people who use the guaranteed freedoms to infringe on the freedoms of others, or destroy the system altogether. Freedoms that are prone to being abused are usually political rights such as the freedom of speech, press, assembly and petition. Throughout history, individuals or organizations abused those rights to disseminate discriminatory demands or propagate political ideologies targeted at abolishing democracy. In the light of Germany's National Socialism, the concept of a militant democracy was introduced by Loewenstein and Mannheim (1943). The idea behind it is to manifest constitutional guarantees that secure the elements of freedom and democracy by not allowing them to be changed legally, not even through democratic majorities. The salience of such a concept

¹⁸ Hobbes' war of all against all.

¹⁹ A concept introduced by Linz/Stepan (1996) which proclaims that there is a certain level of maturity that keeps new democracies from relapsing into authoritarian regimes.

becomes evident when looking at recent political developments even within Europe.²⁰ Additionally, a militant democracy enables preemptive strikes against parties or institutions seeking to endanger said elements. Without the RoL, there would neither be the legal foundation for the interpretation nor the decision-making power regarding the limitation of political rights (e.g. the prohibition of demonstrations or even political parties).²¹

In order to establish a functioning RoL, a liberal democratic state needs an institution to interpret and implement the canon of rules and regulations that make up the law, namely the judiciary. Courts are the state institutions that hold all state and non-state actors accountable to the laws passed by the legislative. They restrict and review other state branches' actions and enable the individual citizen to assert his rights vis-à-vis other, even more powerful actors. Only based on a democratic RoL, the judiciary can become the independent actor to fulfill said functions.

The strength of the RoL within a given country mostly depends on how much the judiciary is to be trusted by the citizens. The more independent, impartial, professional and efficient the judicial system is perceived, the higher the trust in and thus, usage of, legal remedies becomes.²² In contrast, the RoL becomes dysfunctional when the judiciary is unduly (politically) influenced, weak, qualitatively poor and/or inefficient. In these cases, civil rights and liberties cannot be upheld and protected by the state. The separation of powers may deteriorate, and non-state actors or private parties may exert undue influence on state branches. This can create a downward spiral which promotes nepotism, corruption and crime. Once the culture of law and order is destabilized and the deterrence potential of criminal proceedings is eroded, the structural building of crime syndicates is enabled. All of the above pose a major threat to

²⁰ Hungary's Prime Minister Viktor Orbán succeeded in passing amendments to the constitution using a parliamentary two thirds majority which limit the powers of the constitutional court in revising laws with view to their content as well as increasing the possibilities of state interventions with the judiciary and higher education. Although the amendments re-shift the balance of power between legislature and judicative in favor of the former, this does not yet equal a limitation of the liberal rights and freedoms. However, should the fundamental values of the EU be breached, it can suspend the respective country's rights within the Union for the time being (Thomas. 2013).

²¹ Although it is being argued by pertinent sources that the limitation of political rights is being abused by the state to suppress undesired political opinions, Germany's constitutional court's treatment of the NPD-prohibition case illustrates how carefully the responsibilities invested in the highest judges are being exercised in weighing the clashing interests of freedom of political association and the protection of the free democratic basic order (Tretbar, 2013).

²² Interestingly, the perception of the aforementioned aspects is more relevant for the citizens' trust in their country's rule of law than the actual output of the judiciary. This is also due to the fact that it is difficult to quantitatively measure categories such as independence or impartiality. For the same reason, Transparency International does not scrutinize the actual, but perceived level of corruption.

democracy, particularly when tolerated or even supported by the ruling political and economic elites. Due to the power monopoly of the governing elites in countries that have previously been ruled by authoritarian regimes, creating a functioning RoL became a cornerstone in external democracy promotion, as will be discussed subsequently.

II.1.4 The Role of the Rule of Law in External Democracy Promotion

Within the democratization context, most countries are in transformative processes from previous authoritarian or dictatorial regimes wherefore democracy promoters need to address five policy areas: *democratic elections, a functioning executive, legislative, and judiciary*; and *anti-corruption measures* (as outlined in Illustration 2 below). Each of the five fields addressed by external democratizers will be explained below. Special emphasize is being put on the RoL and anti-corruption. Since the RoL is considered a prerequisite for democracy, it gained salience in external democracy promotion. Because most transitioning countries struggle with the creation of independent judicial system, anti-corruption measures will also be emphasized.

First, open, free and fair *elections* need to be established to grant everyone an equal voting right. The elections process generates two of the three state branches whose checks and balances-system is characteristic for a democracy: the executive and the legislative. Therefore, the goal of the democratization process with view to the *executive* (political administration) is to conduct decentralization and structural reforms to weed out old authoritarian structures as well as holding it accountable to the other state branches and the public. Furthermore, consultation processes with other interest groups (political, economic, cultural) need to be implemented and a unified civil service system must be created which implements the new policies. Third, the *legislative* has to be reformed in order to be able to work without external influences and have its own respected powers in the legislation process. Additionally, the opposition must be given a realistic chance to contribute to the political debate and the decision-making process, and the inclusion and respect for minority interests must be secured. Fourth, the *judiciary* has to be reformed. Before democratization, the judiciary (just like the legislative) was often a tool controlled and employed by the administration to implement its policies and ideologies. Consequently, judicial independence needs to be granted which includes the filling of vacancies with democratic legal experts that are trained in democratic and European law. Providing them with the necessary resources

(technological, financial, personnel) will increase their capacity to handle all cases and ensure the enforcement of their rulings. Also, the access to courts must be ensured for all citizens. Finally, in order to prevent the newly reformed state from slipping back into intransparent, random and undemocratic patterns, *anti-corruption* measures have to be established across all state branches. After the summarizing Illustration 2 below, it will be explained why a functioning judiciary (based on the RoL) is vital for the sustainable development of all the aforementioned aspects.

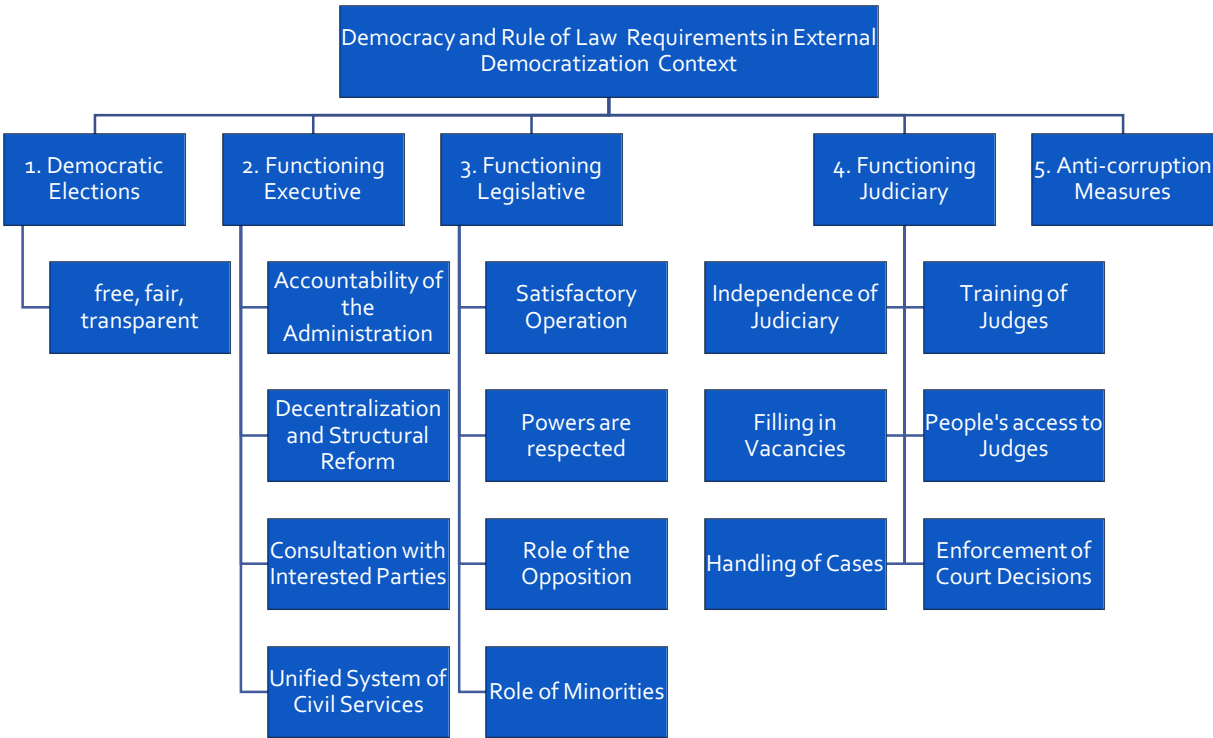


Illustration 1 – The EU’s RoL requirements in external democratization (Arnull: 2002)

The corner stone of democracy are free, fair, and transparent elections. This means that each legible voter has one vote which has the same weights as any other; that everyone has the right to run for office and that the public has right to monitor the elections to make sure that the procedural rules are being respected. Since elections decide who will govern the country, unfree or authoritarian regimes often try to tamper with the process to create results in their favor. The RoL not only provides the rules based on which elections are being held but also

the tools to enforce them. As the RoL's executive organ, the judiciary can investigate allegations of breached rules or inconsistencies, decide upon and sanction electoral fraud, or declare elections void which leads to new elections. Without a functioning RoL, there is no reliable protection of the electoral process as has seen repeatedly in African countries, Communist regimes, or lately in post-Soviet states.

With view to the executive and legislative, there persists a power triangle with the judicative which is regulated through a system of checks and balances. This mechanism ensures that none of the branches becomes too powerful on its own. While legislative and executive formulate, vote upon and implement laws, their 'righteousness' – either formal or with view to its content - can be tested by the people using the court system. If a law is considered defective or in breach of the constitutional provisions, the courts can require the legislature to make respective changes. Without the RoL, particularly the judicative limiting the powers of the other two branches, it would become easier for policy makers to introduce laws that infringe upon the civil liberties and political rights of the citizens. Particularly with view to the legislature, the judicative has a special safeguarding power in terms of protecting the rights of the opposition and minorities.

Last, the RoL and corruption are mutually influential. While corruption can erode RoL structures, a functioning RoL can be potent enough to discourage corrupt behavior and thus sustainably weaken respective structures. Corruption describes both a mindset and an action where a person (in the political context an office holder or civil servant) is acting in his official capacity while personally profiting from it. The most common forms of corruption are bribery and embezzlement. Psychological and economic studies show that corrupt behavior is contagious (Gino, Ayal, & Ariely, 2009; Ponsioen, 2014) which means that once people see others acting in a corrupt way they are also more likely to use similar means to reach their goals. Especially once certain expectations regarding corrupt behavior are established even people who oppose the idea of corruption may be forced into subduing themselves to those practices (e.g. in countries where doctors or civil servants' wages are too low to survive and patients are expected to bribe them for faster or better treatment as for example is currently the case in Ukraine).

Corruption can affect both aspects of the RoL: the rules themselves as well as their implementation. Undue influence in the law-making process can lead to laws without proper,

democratic justification that privilege or discriminate against certain groups of society. Undue influence within the court system can lead to either delays/non-prosecution/weaker sanctioning or to harsher sentencing than would be required by law. Both actions result in the obstruction of justice which demonstrates the arbitrariness of the justice system. As has been explained earlier, trust in the justice system rests on the predictability and adequacy of the ruling elites. Hence, corruption diminishes the people's trust in state institutions and therefore endangers the democratic stability of a state. Luckily, the aforementioned studies also show the reverse: the stronger the state fights corruption, the less acceptance there is not only in public but in civil society, too. If the justice system strictly pursues and sanctions corrupt behavior, it signals independence, strength and reliability of state actors and their policies. As the previous analysis has shown, the RoL not only provides the foundation for all other dimensions of democratic quality (Diamond/Morlino, 2005) but it also reflects the entire democratic quality of a regime (Linz/Stepan, 1996).

As will become evident in the next part, all the introduced elements are particularly relevant for the case studies at hand because both Croatia and Turkey's democratization and EU convergence process has been delayed due to problems that a functioning RoL may address.

II.1.5 The Role of the Rule of Law in Croatia and Turkey's Democratization Process

The COM's Progress Reports focus on the development a candidate has made from one year to the next. In order to evaluate the quality of the progress it is crucial to know where the country started, if and how the judicial reforms pushed by the COM have been implemented what made the countries comply or not comply with them. Therefore, the following sections will focus on the historio-political backgrounds of Croatia and Turkey's RoL, their respective judicial reforms during the accession process (until 2012) and its reform drivers.

These reform drivers, as introduced by Noutcheva and Aydin-Düzgit (2012) include the *credibility of the EU's accession promise*, the *candidate's state capacity* to implement the Accession criteria as well as the impact of the *reform-based differential empowerment* of domestic players versus the ruling political elites. After outlining the individual findings for both case studies, an examination of the overriding themes, shared characteristics and differences will ensue.

Croatia's Historic-Political Background

After the breakdown of the Socialist Federal Republic of Yugoslavia in 1992, two main political parties dominated Croatia's politics: the leftist successor of the former communist party, the Social Democratic Party (SPD) and the anti-communist, nationalist Croatian Democratic Union (HDZ) led by Franjo Tudjman who led Croatia to independence and through the mid-90s post Yugoslavian wars which reinforced Croatia's feeling of national strength. Tudjman opposed Croatia's convergence with the EU because he believed the Union to denounce Croatian independence and try to tie it back to its old enemies. Consequently, while other Western Balkan states started their accession process, he rejected the EU path which resulted in the country's isolation (Jovic, 2006: 86-89). After his death in 1999, a center-left coalition governed Croatia which opened "the way for democratic change and European integration" (Noutcheva/Aydin-Düzgit, 2012: 65). Until 2003, none of the parties specifically supported EU membership, possibly based on the perception that the EU did not support Croatian independence. This changed when HDZ won the elections headed by Ivo Sanader who made EU-accession his priority and consequently paved the way by cooperating with leftist political groups and integrating Serbian minorities into his government. He also catered to the EU's interests by reconciling and cooperating with other Western Balkan states and lobbying Member States for their support in Croatia's accession. However, both the inclusion into NATO as well as the signing of the SAA stood still between 2001 and 2005 based on the non-extradition of two leading war-criminals as requested by the International Criminal Tribunal for the former Yugoslavia (ICTY). This can be attributed to the prevalence of Tudjanist supporters - particularly in the police and judiciary - which considered their extradition non-constitutional and treacherous. This view was shared by a large part of the population (Jovic, 2006: 95-98). Once Croatia complied with the requests, EU accession process continued.

Croatia's Judicial reforms during the Accession Process

Croatia has been applauded for the "complete overhaul" of its judicial system in the key areas judicial independence, impartiality, efficiency and professionalism (Noutcheva/Aydin-Düzgit, 2012: 62). While its RoL performance deteriorated slightly between 2004 and 2006, it picked up since 2007. In September 2005, Croatia adopted the Justice System Reform Strategy as well as an Action Plan for its implementation as required by the COM. Between 2009 and 2011, the Framework of the Accession negotiations was introduced which responded to the

criteria of Chapter 23. Constitutional Amendments were made in 2010 targeting the judicial independence requirement “by boosting the autonomy of the State Judicial Council and the State Prosecutorial Council and reducing the power of the Justice Ministry in judicial appointments” (Ibid: 63). Furthermore, Croatia boosted its efficiency between 2005 and 2010 by substantially reducing its case backlog while maintaining generally high numbers (Reding, 2010). Nevertheless, the development of Croatia’s judicial system is still perceived as inadequate by its citizens who feel that there is a persistently high level of corruption without adequate investigation, prosecution, or conviction of government officials (Doric, 2010).

Croatia’s Reform Drivers

The *EU’s credibility* in making Croatia the next Member States were generally high, except for the period between 2008 and 2009 where Chapter 23 negotiations were put on hold due to Croatia’s lack of cooperation with the ICTY. Still, fluctuations in its RoL performance persisted wherefore the credibility of the accession promise was not Croatia’s major reform driver.

Out of the Western Balkan states, Croatia’s *state capacity* is the most advanced in terms of enforcing the Acquis rules and ensure the citizen’s compliance. Although investments in the judicial capacity and RoL reforms have been made, progress has remained moderate (Noutcheva/Aydin-Düzgit, 2012: 64). While the “state capacity has helped advance the regulatory and legislative agenda” of the COM this can be mostly attributed to the fact that it was within the interest of the ruling elite to carry out the reforms in order to convince the EU of its good will (Ibid.).

Croatia’s RoL progress was deeply impacted by the *differential empowerment* aspect. During the two governments led by Sanader (who was prosecuted for corruption and nepotism in 2010) Croatia’s RoL performance was deteriorating suggesting that the lack of reform progress can be attributed to its interference with domestic interests. In order to set herself apart from the old HDZ leadership and win credentials with the EU, new Prime Minister Jadranka Kosor had Sanader arrested. Thus she signaled her sincerity in fighting corruption in her own ranks (Loza, 2010). Kosor’s strategy – following EU requirements to assert herself vis-à-vis her electorate and Brussels – shows that RoL reform progress rises once domestic and EU interests align and support for the ruling elites is secured.

Turkey's Historic-Political Background

Up until 1924, Turkey was an Islamic state with politics and religion deeply intertwined. The reforms introduced in 1924 abolished the Caliphate, including the “symbols of Islam such as the Arab alphabet, the headscarf, and religious attire” (Müftüler-Baç, 2016: 66) and laid ground for a secular state. The religious traditions were suppressed until the 1980s, when Islamist elites assembled to form the Islamist-oriented Justice and Development Party (AKP). The Party came to power in 2002 and continuously gathered electoral support, not least based on their pro-European stance including political reforms in line with the Acquis (2002-2008).

Turkey's Judicial reforms during the Accession Process

Turkey's reform process can be categorized into three stages: the ‘golden age’ (2002-2005), the ‘slowing down’ (2005-2008), and its ‘reversal’ (since 2011). In the early 2000s, a number of RoL-relevant legislation have been introduced including the new Civic Code (2001), the new Labour Law (2003) and a new Penal Code (2004) including the abolition of the death penalty which aimed at harmonizing Turkish rules with the European Acquis. “From 2001 to 2008, a total of 14 Constitutional Amendments were adopted along with 9 Harmonization packages to the EU Acquis. From 2005 to 2008, the Turkish parliament adopted 148 laws in an attempt to adjust to the EU's political conditionality.” (Ibid.). The slowing down of the reforms in 2008 coincides with the public statements made by influential EU Member States emphasizing accession alternatives such as the “privileged partnership”. During that period, Acquis criteria such as the freedom of religion and personal rights were used by the government to fight secular opposition in order to reintroduce the headscarf and other religious symbols. When secular groups moved the Constitutional Court to prohibit the AKP in 2008 based on its anti-secularist actions, not only did they lose the case but the attempt backfired resulting in a number of cases against opposition and journalists (e.g. Ergenekon 2007 and Balyoz 2010) which led to largely legally unfounded measures including imprisonment without trials. As Yazici (2009: 166ff.) argues, after 2008 conditionality has deteriorated to a catalyst which instead of strengthening democracy and the RoL enabled the AKP to abuse liberal democratic principles to restore its conservative, authoritarian leadership. Since, 2011, opposition groups and dissenting opinions were “effectively blocked and censured” through the closing down of papers, the imprisonment of journalists, the banning of social media sites and similar measures such as the banning of Twitter and Youtube in 2014 (Ibid.). Moreover, secular high-ranking military officials have been

persecuted without due process and legislative measures have been taken to limit the autonomy of the judiciary (Müftüler-Baç, 2016).

Turkey's Reform Drivers

Up until the mid-2000s, the EU's accession promise was credible to Turkey. However, with the start of the accession negotiations a number of factors continuously decreased the EU's *credibility*. France's and the Netherlands' rejection of the Constitutional Treaty signaled that a deeper level of EU integration was not wanted by the Member States which was echoed in the additional accession criteria that were developed specifically for Turkey. These stated that the process was "open-ended" with a possible alternative outcome to accession, included more Chapters to be closed than any other candidate before (35) and introduced the absorption capacity criterion which had not been included previously (Redmond, 2007). In combination with the elections of Merkel (2005) and Sarkozy (2007) who publicly stated the "undesirability of Turkish accession" and the freezing of negotiations in 2006 due to the Cyprus conflict with Greece the perception of Turkey being willfully kept out of the Union increased (Noutcheva/Aydin-Düzgit, 2012: 68).

Turkey has a high *state capacity* depending on the political will behind a policy which can and has been used both to promote and hinder the RoL progress. While Turkey suffered from a large backlog of cases which significantly impedes a functioning RoL (Vald Helden, 2009), another aspect is more crucial: the role of the Kurdish terror organization PKK. During periods where terrorist activities declined (1999-2005), reforms strengthening the RoL were introduced. When the PKK increased their terrorist activities in 2006 and 2007, this led to nationalist reaction which stifled the RoL progress (Noutcheva/Aydin-Düzgit, 2012: 69). Consequently, in times when state capacities were absorbed to fight domestic terror, reforms stalled.

Differential empowerment has a great impact on explaining Turkey's RoL progress dynamic. The AKP accomplished its broad electoral support through making promises to two very different societal groups: the conservative, Islamic population was promised extended religious freedoms beyond the previous secular limitations while moderate voters and liberals were lured through pro-European liberal-democrat policies (Özel, 2003). Once the AKP gathered enough domestic clout to follow its own agenda, it abused democratic principles to suppress oppositional groups. For example, civilian control over the military is in line with

EU standards but was used to silence and persecute secular voices. Constitutional amendments aimed at strengthening the RoL (specifically judicial independence) included the 2009 Judicial Reform Strategy. While it was de jure in line with Acquis provisions it de facto enabled the executive to create a favorable set-up of the Council on the Judiciary. Consequently, reforms were being selectively implemented based on the government's interests rather than on the requirements set out for EU accession.

Croatia's and Turkey's reform drivers in comparison

Both Croatia and Turkey experienced a RoL reform rollercoaster with alternating phases of speedy implementation, stagnation, and even reversal. Noutcheva and Aydin-Düzgit analyzed the reasons behind the latter two by looking at the reform drivers. According to them, the credibility of the accession promise was not explanatory in these fluctuations as Croatia's RoL performance deteriorated even during the opening of accession talks (late 2005/2006) whereas Turkey's implementation track record improved even during times where major EU Member States publicly denounced its accession in favor of a 'privileged partnership' (2008). Hence, EU credibility is not crucial for reform progress. Regarding Croatia and Turkey's state capacity it was both sufficiently constant and high (although slightly higher in Croatia than in Turkey) between 2004 and 2009 to adequately implement the reforms. Therefore, reform reversals cannot be explained through a lack of resources. Differential empowerment turned out to be most influential on reform progress. If EU accession criteria aligned with the ruling elites' domestic interests, they were being implemented. Wherever this was not the case, implementation was fragmented, stalled or reversed. Substantial progress on "EU-advocated institutional and legislative changes" (Ibid.) was observed in Croatia especially after 2009 and in Turkey before 2005. This shift temporally coincided with Croatia's new Prime Minister Jadranka Kosor who was trying to demarcate herself politically from her predecessor Ivo Sanader who already faced corruption charges at that point. It also aligned with the time period where Turkey's AKP party gathered enough domestic electoral support both from its Islamic base as well as the political center to partially emancipate itself from the EU accession question. Consequently, while EU incentives triggered the RoL reforms in both countries, they were only influential as long as the ruling elites' interests did not clash with those of the EU.

II.2 External Democracy and Rule of Law Promotion

In the following section, external democracy promotion will be defined and examined based on its types and implementers, their motives and methods. Special focus will be put on the EU's democratization efforts, particularly its most relevant tool: conditionality.

According to Carothers, "Democracy aid is all aid for which the primary purpose not the secondary or indirect purpose, is to foster democracy in the recipient country..." including economic and social programs (Carothers, 2000: 188). Schmitter and Brouwer (1999: 10ff.) concretize his notion by adding that it consists of "all overt and voluntary activities adopted supported, and (directly and indirectly) implemented by (public or private) foreign actors explicitly designed to contribute to the political liberalization of autocratic regimes, ... or consolidation of democracy in specific recipient countries.". This external influence is targeted at all basic patterns of political decision-making as well as the general order within the democratizing country so as to at least meet the minimal criteria of democracy (Sandschneider, In: Freise, 2008: 23). Respectively, external RoL promotion is aimed at implementing the basic conditions, institutions, norms and values of a democratic RoL as well as instilling a democratic legal culture to create a functioning judiciary. A functioning judiciary in turn is characterized by being able to protect the citizens' interests vis-à-vis each other and the state branches while being protected against undue political influence or abuse by those state branches as well.

II.2.1 History and Actors

Historically, external democracy promotion originated in the United States where it initially served American security and economic interest in the aftermath of WWII. According to the then-popular modernization theory²³, only economically strong countries have democratic potential wherefore financial support was granted to particularly Western-European countries (i.e. Marshall plan²⁴) in order to jump-start their economies which in turn was necessary to create new markets for American goods and services. In the 1980s, Reagan's administration initiated explicit democracy promotion programs in countries where it served its political interests yet still supporting numerous non-democratic regimes in South-America and the

²³ Modernization theory stresses that traditional societies will develop economically and politically when adopting modern practices in terms of production, technology, transport, communication etc. just like the already modernized societies have. In turn, the latter can support the former in achieving said modernization level (Lipset, 1959).

²⁴ The Marshall Plan was US sponsored a broad economic investment program to rebuild Western Europe after the WWII including loans, natural resources, food and other goods (Kimmel, 2005).

Middle East (Giesendorf, 2008: 107/8). Those interests concerned inner-political motives as well. By furthering the agenda of exporting democracy, the Reagan-administration enabled the convergence of neo-liberal and neo-conservative thought which in turn “facilitated a broad political alliance ranging from cold warriors to human rights activists” (Wolf, 2009: 69). This approach resulted in a narrowing-down of the concept of democratization where democracy was defined as a “procedurally limited version of polyarchy” and democracy promotion was equated with the support of a capitalist market democracy. The support of human rights was aimed at implementing civil and political liberties such as free electoral processes, the institutionalization of a democratic RoL, the freedom of expression, press and association (Ibid.).

The 1990s were coined by a transformation of democracy promotion strategies, not least because two other important democratization actors entered the scene: the World Bank and the European Union (then European Communities). The World Bank introduced the leitmotif of ‘good governance’ into the democratization process which included the elements of public accountability, pluralistic political institutions, anti-corruption policies, the RoL as well as freedom of the press and human rights protection. Economic support became conditional on democratic goals which became the hour of birth of the principle of conditionality (Schraeder, 2002: 2018). The EU’s engagement changed the underlying tone of democracy promotion from “pressuring countries into compliance” – as has been done by both the US and the World Bank - to “an open and constructive dialogue” with an “emphasis on persuasion and learning as well as positive conditionality and capacity-building, with negative incentives such as sanctions ...as a last resort” (Börzel/Risse, 2009: 39). Thus, the EU presented an alternative approach to Eastern European countries – not only to its fellow external democratizers – but to authoritarian suppression during Communism which made the offer of democratization more appealing to the respective target countries (Simmons, 2011).

The mentioned democratizers represent the multitude of actors that are involved in the democratization business: individual states (such as the USA), international (Worldbank, UN) and supranational organizations (EU). Other influential players include NGOs such as Human Rights Watch or Amnesty International as well as democracy-promoting non-governmental institutions like churches, unions and foundations that aim to support specific goals in their respective partner countries (Giesendorf, 2008: 122).

II.2.2 Motives

While there are a number of reasons that lead to the actual increase and spread of external democracy promotion, there are only two groups of rationales behind its implementation: democracy promotion as an instrument to further one's own agenda or democracy promotion as an end in itself. The factors that explain the rise of external democracy promotion include the passing of the 'danger' of communism because it enabled particularly the United States to refocus its foreign politics attention to other areas as well as the concurrently arising opportunity in post-communist states to influence their transition to democracy. Thus, a re-legitimization of foreign aid and development policies was instigated which was supported by an academic paradigm shift that portrayed external democracy promotion as something that was "feasible and useful" (Wolf, 2009: 71). Since the mid-1970s – as part of Huntington's third wave of democratization – the rethinking of western democracies regarding foreign policy was guided by the combination of idealism (a deep-felt moral obligation to expand democratization because it provides people with more freedom) and realism (the realization that it could become an effective tool in pursuing self-serving interests) (Freise, 2008: 24). Idealistically, democracy is – at least from the western point of view - the only legitimate governing system and must therefore be supported in other countries as well in order to grant their people with civil rights and liberties. From the instrumental standpoint, external democratization furthers strategic, economic, power-political and security motives of both donors and recipients. Strategically, recipient countries may be able to justify unpopular political decisions under the democratization-umbrella (such as retraction of social security systems). Economically, donor and recipient countries may seek to secure economic cooperation (e.g. Marshall Plan) or the supply of natural resources. Ideologically, the expansion of Western political values also increases the overall sphere of Western influence in other areas. With view to security questions, external democratization may help solve conflicts (e.g. the prevention of problem expansion to one's own territory by supporting local conflict resolution), the fight against terrorism or the support of international stability and security based on the democratic peace theory (Giesendorf, 2008: 120-21).²⁵

²⁵ Democratic peace theory stipulates that democracies are unlikely to go to war with one another because democratic leaders are being held accountable for their actions by the public including the blame for losses, because there is less hostility between countries with similar basic values and because war would endanger the wealth gathered in most democratic societies (Reiter, 2012).

II.2.3 Types and Methods

According to Giesendorf (2008: 124), there are six types of external democracy promotion interventions which can be divided into a group of direct and indirect forms. Forms of indirect democratization include *contagion*, *diffusion*, and *zeitgeist*. Forms of direct democratization include *control*, *consent*, and *conditionality* which are case-specific instruments and measures (see Table 4). *Contagion* describes the unintentional adoption of democratic norms, values and institutions, for example through geographic proximity to other democratic states. *Diffusion* is a somewhat similar process because it leads to the expansion of democratic values and structures through growing international interdependence thus bridging longer distances. It often occurs through economic cooperation when business cultures and practices are being imported from democratic countries. *Zeitgeist* embodies an underlying national or international spirit, a type of short-term constant trend that informs political decision-making. Since those factors cannot actively be controlled by democratizers, there are no specific instruments or measured to be named. The strongest and most aggressive type of external democratization is *control*. It enables the donor to initiate and direct the entire transformation process to a point where she can force the recipient to comply with her wishes. This has been the case for example with Western Germany after WWII. A less-oppressive and more fruitful approach is *conditionality* (see II.2.4 for specifics on EU accession conditionality) which applies the carrot-and-stick-method of ‘luring’ recipient countries with positive incentives (‘carrots’ in terms of financial support or intangible advantages) or punishing their lack of compliance with sanctions (‘sticks’ such as cutting aid or reducing cooperation). The least intrusive and hardest to come by type of direct external democracy promotion is *consent*. It is based on both the donor’s and recipient’s agreement over the ‘righteousness’ of the measures and the “acknowledgement that stable, sustainable democracy needs deliberate support and engagement of inner societal and political groups” (Ibid.). The subsequent table provides an overview of the six types of external democracy promotion:

Type of Democracy Promotion	Instrument	Measures
Contagion	-	Political (diplomacy/pressure)
Control	Sanctions	
Consent	Incentives of cooperation	
Conditionality	Incentives/sanctions	Economic (sanction/incentives)
Diffusion	-	

		Military
Zeitgeist	Positively influencing background variable	

Table 4 – Six types of external democracy promotion (cf. Giesendorf, 2008: 124)

II.2.4 The EU as an external Democracy and Rule of Law Promoter

The European Union has been one of the most active external democracy and RoL promoters both inside and outside its borders. Democracy and RoL promotion has become “a centerpiece of the EU’s foreign policy” that is backed up “by considerable financial and personal resources” (Börzel/Risse, 2009: 36). In fact, the EU is the world’s largest donor who provide for more than 50% of the global official development assistance (Ibid.). The reason behind EU democratization attempts are the Member States long-term economic and geo-political interests (instrumental motivation) as well as the belief that Europeanization will lead to democratization which is an end in itself (idealist motivation) (Simmons, 2011). As this thesis focuses on the RoL assessment of candidate countries, the EU’s enlargement procedure and the role of its monitoring Reports within it will be explained in further detail.

Accession procedure

Since detailed descriptions of the EU enlargement process have been provided elsewhere and are not necessary for the purposes of this thesis, this section will be confined to providing the reader with a general understanding of the process.²⁶ There are three stages to becoming a new member state. First, the potential new member files an application for membership with the Council which - based on the recommendations of the Commission - decides whether or not the country will be granted the status of a candidate country. This decision is dependent on two questions: a) Is the applicant country geographically and socio-culturally ‘European’ and b) Does it possess a sufficient degree of institutional stability in terms of the political Copenhagen Criteria (for details see next section) to fulfill the preconditions in the foreseeable future? (Ludwig, 2011: 74).²⁷ As will be pointed out later on, many countries with which the enlargement process has been conducted have not fulfilled said criteria before moving on to the second stage. In fact, some even still had significant institutional

²⁶ Schimmelfennig/Sedelmeier (2002), Pridham (2002), Laursen (2013)

²⁷ If a country is geographically European or not opened up a whole other debate regarding historical and natural borders. This debate will not be part of this thesis. Countries that have been scrutinized based on that criterion where Morocco (deemed not European) and Turkey, whose territory is mostly located in Asia (deemed European) (BBC, 2000).

shortcomings at the time of their accession (Kochenov, 2004) which is one of the reasons for taking a closer look at the Reports assessing their developments (cf. Chapter III.3.1). Having been granted the candidate status is a necessary but not sufficient precondition for becoming a member state. The second stage of the enlargement process is the opening of the accession negotiations which carry their name wrongfully because it suggests parity. In reality, the Commission mostly advises the candidate country on how and until when which aspects of the remaining Copenhagen Criteria have to be implemented and the candidate countries have to comply. This notion is supported by a German MEP stating that the negotiations are “asymmetrical” and “no real negotiations because the applicant country has to do what the EU says” (EP, Brussels, July 2016). There are 35 Chapters to be “negotiated” and closed before the Council declares the candidate country’s accession maturity. This is not automatically linear process either. In case progress in the eyes of the Commission is stagnating or even reversed, negotiations can be put on hold, as has been done in Turkey between 2010 and 2013 (Euractiv, 2013). Once the candidate country’s accession maturity has been asserted, the Council obtains both the recommendation of the Commission and the application approval from the Parliament and makes its decision based on them in line with Article 49 (1) EUV (Lisbon). The accession treaty is signed between the candidate country and each of the Member States and must be ratified by all parties to the contract before coming into effect. Since all existing Member States need to be satisfied with the results of the negotiation process before signing and ratifying the contract this can become a difficult and lengthy procedure that is prone to political power games. For example, Croatia’s accession had not been ratified by the Danish, Dutch, Slovenian and German parliaments in February 2013 for domestic political reasons whereas the accession was planned for (and eventually happened on) July 1st of the same year (Pavelic, 2013). The following Table 5 summarizes the responsibilities of the COM and the Council within the accession process:

Responsibilities during the Accession Process	
European Council	COM
<ul style="list-style-type: none"> • Principle decision to enlarge 	<ul style="list-style-type: none"> • Prepared documents and Aps acting on the mandate of Council and EP
<ul style="list-style-type: none"> • Drafted CC and formulated conditionality principle 	<ul style="list-style-type: none"> • Prepares common positions for negotiations which are usually followed by Council and EP
<ul style="list-style-type: none"> • Requested framework of pre-accession documents from COM 	

Table 5 - cf. Kochenov (2008: 58/59)

Conditionality

“The idea of conditionality is beautiful in theory: there is a clear set of criteria (the Copenhagen criteria), an impartial institution to conduct assessment compliance (the European Commission), and the prospect of membership for those who pass the test.”

Kochenov (2008: 56/57)

As outlined in II.2.3, conditionality is a strategy based on positive incentives and sanctions used to foster the compliance of recipient countries with the criteria of democratization as outlined by the donor country. It “refers to the logic of consequentiality (“do X get Y”) as opposed to the logic of appropriateness (“good people do X”)” (Simmons, 2011). Although the EU is not the only external democracy promoter using said approach, it has been the “prime focus of attention in the existent research” due to its “more extensive and systematic” implementation of the concept in comparison to other donors (Pridham, 2008).²⁸ According to Kochenov (2008: 53-55), the EU had multiple reasons for introducing conditionality. Officially, it served as a security mechanism to assure that only mature countries could enter the Union. By creating an “impartial assessment of the applicants’ progress towards accession” it was supposed to free the accession process “of any political considerations” and make it purely merit-based (Ibid.). According to the European Council and the COM, “Jeder

²⁸ For studies in the role of IOs in external democracy promotion and their methods see Whitehead (2001), Linden (2002), Kubicek (2003), and Peevehouse (2005).

einzelne Bewerberstaat wird nach denselben Kriterien beurteilt und auf dem Weg zum Beitritt seinem eigenen Tempo entsprechend seinem Vorbereitungsstand voranschreiten“ (Conclusions of the Chair, 15./16.06.1998, In: Ludwig, 2011: 77). While this statement establishes the equality for all candidate countries regarding the basis for accession decision-making, it still enables a customized timeline for each individual case based on the level of preparation, and thus accession maturity. The Council suggests that the compliance with conditionality will lead to accession by stating that “accession will take place as soon as an associated country is able to assume the obligations of membership by satisfying the economic and political conditions required” (European Council 1993, 13, In: Baracani, 2009: 66). Those requirements are codified as the Copenhagen criteria (CC) and explained in detail in the subsequent section. The perception of automaticity has been challenged by some scholars stressing that it neglects the role political preferences play in the accession process. In fact, Kochenov even argues that high benchmarks were deliberately being set to create a justification tool for the Union’s “own unwillingness to accept the CEECs into its ranks”. Given the lack of clarity regarding some of the criteria as well as their differing levels of application for case studies examined by Kochenov, his hypothesis is comprehensible yet requires more in-depth testing, especially for the current candidate countries.

EU conditionality comprises of range of instruments ranging from “political dialogue and persuasion to capacity-building” over ex ante (threat to close off or postpone membership) to ex post conditionality (Börzel/Risse; 2009: 42). The reason behind the success of conditionality in terms of implementation-outcome has been attributed to the governing elites’ rationale in recipient countries. Rather than converging to democratic norms through a socialization process, the ruling elites tend to make “rational, cost-benefit decisions” when accepting democratization support through the EU (“rational bargaining model” according to Schimmerpfennig/Sedelmeier 2005). The ‘carrot’ of financial and institutional support, and ultimately, full membership (i.e. access to markets as well as security and stability) is so powerful that recipient countries are highly motivated to at least superficially comply with EU conditionality requirements. However, this is only true as long as the promise of full membership is actually “credible” and the general domestic conditions within the recipient country are “favourable” (Simmons, 2011). In contrast, if “domestic structures and identities” are in conflict with European ideas and values or the recipient countries “lack the necessary

capacities to introduce required changes”, conditionality becomes significantly less effective (Börzel/Risse; 2009: 10).

Accession requirements - the Copenhagen Criteria

By using the Copenhagen Criteria (CC), the EU creates a link between Europeanization and democratization. It asks of the recipient country to not only internalize democratic principles but also to become a liberal democracy shaped after its own Member States. The idea behind the Copenhagen Criteria is applaudable: making the accession process more feasible and transparent by providing clear requirements a candidate country has to fulfill in order to secure their maturity for full membership. Before critically assessing the quality of the criteria they will first be outlined. The Criteria have been developed during a conference of the Council in Copenhagen in June 1993.²⁹ There are three dimensions to the criteria: *political*, *economic* and *judicial*. The *political* dimension requires the guaranteed institutional stability of the democratic order in accordance with the RoL as well as the respect for human rights. The *economic* dimension demands a functioning, effective market economy that can withstand the competitive pressure within the Union. The *judicial* dimension (*Acquis communautaire*) calls for the adoption of all political and economic goals as well as the implementation of all the legal obligations that arise from the common vested rights of the Union including the articles of agreement, treaties, case law, Opinions, Recommendations as well as binding and non-binding legal acts (Ludwig, 2011: 64).

²⁹ The criteria were introduced in the Conclusions of the conference, DOC/93/3 from 22/06/1993: p.13.

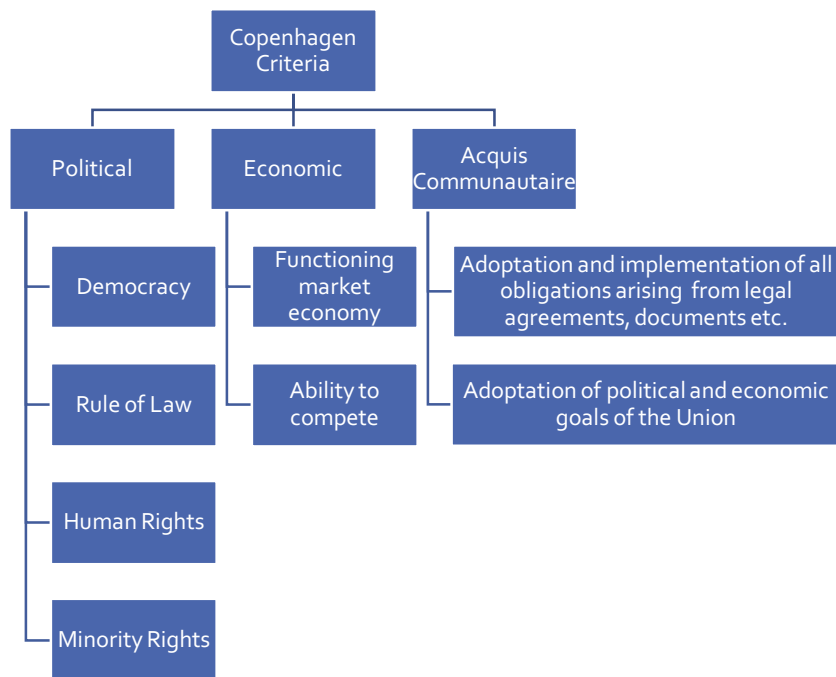


Illustration 2 - Elements of the Copenhagen Criteria (Europa.eu)

With view to democracy and the RoL the political criterion consists of four branches that have to be addressed by the candidate country: *legislative*, *executive*, *judicative*, and *anti-corruption measures*. With view to the *legislature*, elections have to be free and fair and conducted within a multi-party system. The opposition must be able to carry out its tasks and adequate immunity regulations must be established. Regarding the *executive*, political stability must be granted and a functioning civil control must be exercised over military, intelligence services and the police. Furthermore, the public service sector needs to be effective and efficient through establishing transparent human resource policies, adequate training, pay and technological equipment. For the *judicative* it is crucial to grant judicial independence which requires transparent human resource criteria (recruiting, promotion, discipline), adequate pay, continuous training particularly in EU law, reduction of case backlogs and duration of proceedings, easy access to the justice system through legal aid and court appointed defense as well as the general equipment of courts. *Corruption* must be targeted by ratifying relevant international agreements as well as implementing national laws that criminalize corrupt behaviors. This must go hand in hand with establishing efficient anti-corruption institutions and fostering a political culture of zero-tolerance (Ludwig, 2011: 117).

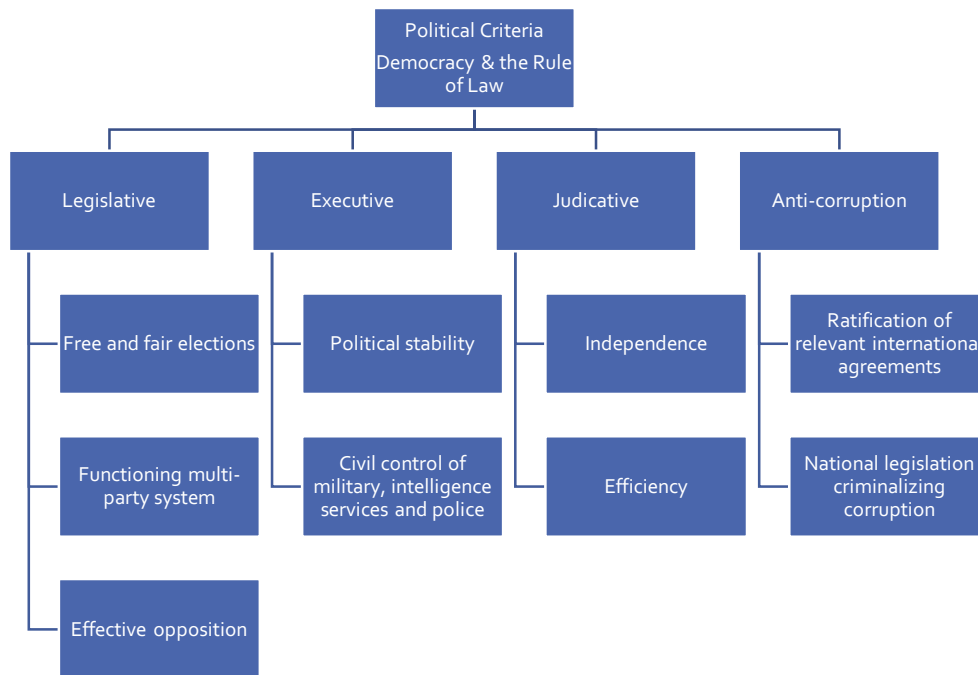


Illustration 3 - Elements of the Political Criteria: Democracy and the Rule of Law (Ludwig, 2011: 117)

Critical acclamation of Copenhagen Criteria

So far, the Copenhagen Criteria appear rather clear and feasible. The concept of free and fair elections is acknowledged and progress can be tested through election monitoring. The same applies to the civil control over powerful institutions, the ratification of specific legislation and even most of the efficiency requirements. However, some other aspects remain rather vague. For example, there is no clarity as to when an opposition can work effectively, what exactly constitutes political stability or when a culture of anti-corruption has been established. This is due to the fact that those values are hard-to-quantify commodities. Most importantly, it remains unclear what exactly is meant by and expected of candidate countries with view to the RoL. While the latter question will be addressed in detail in II.2.6, a brief general critique of the Copenhagen Criteria will be offered at this point.

The lack of clarity regarding some of the criteria is the biggest problem because without it, there can be no telling if and when a country has succeeded in meeting the goals and is ergo ‘ready’ for membership. Also, it cannot be proved if all candidate countries are being held to the same standards if there are no common standards to begin with (Kochenov, 2008: 300). This defeats the whole purpose of the criteria to create a more feasible and transparent accession process. But even if the criteria were specific enough, other aspects endanger their

purpose. First of all, it remains unclear whether the sources of the information used for testing the criteria fulfillment are actually decisive for the final evaluation (Ludwig, 2011: 77). Moreover, it can be expected that the political and economic context of the EU and its respective candidate countries may have a stronger influence on the evaluation than the actual fulfillment of the criteria (Maresceau, 2003: 9). Not only do those aspects make the accession process ‘error-prone’ but they also abet the chance of willful abuse. If the lack of clear criteria coincides with political bias by decision-making agents this can lead to more favorable or more detrimental evaluations – and therefore ‘wrongful’ accession decisions – of the candidate countries in question.

Being aware of those issues, why does the European Commission not concretize the requirements to avoid such critique? There are two reasons for the omission both of which are easily comprehensible. The ‘inconvenient truth’ is that having clear-cut criteria for candidate countries would put pressure on all existing Member States to fulfill them as well in order to avoid the notion of double standards. Although ‘old’ Member States are not being tested as to their abidance of the criteria it becomes easier to point fingers and criticize them if specific violations can be proved.³⁰ The ‘convenient’ reason relates to the diversity of political, economic and judicial structures among the Member States. Although all Member States are democracies based on the RoL with a market economy, there are remarkable differences in the actual set-up of those institutions. The most prominent and relevant example for this thesis is the judicial system. Whereas most Member States have a civil law system, the UK has a common law tradition. Whereas in some countries there are strong hierarchies in the judicial system with court presidents making major managerial decisions, in others the hierarchies are rather flat and the court presidents are merely *primes inter pares*. If there were specific instructions on how to organize the judicial system, this would equal a qualitative judgment regarding what is the ‘better choice’. This in turn would negate the fact that all Member States have their own traditions and peculiarities and have usually shaped their institutions according to their needs. Although European integration aims at streamlining institutions and processes, each member state – due to its sovereignty – is entitled to individualized solutions as long as

³⁰ Some of them would surely not withstand such a test, such as Greece or Spain for economic reasons or Bulgaria and Romania regarding their rule of law and corruption issues. In fact, Member States that are in breach of the criteria can de jure be held responsible for it and punished by suspending their voting rights. However, de facto, this tool has never been implemented before (Eur-Lex, online).

they don't contradict the *Acquis communautaire*.³¹ At this point, focus will be put on what inspired the thesis: claims of political bias within the accession process.

II.2.5 EU Monitoring and Progress Reports

As has been mentioned before, the accession decision is not merely merit-based but influenced by political preferences of the Member States. However, every accession decision is heavily based on the findings and recommendations that are included in the COM's Reports on the candidate country which it prepared for the Council and the European Parliament. Consequently, political interests should be reflected in the Progress Reports evaluations. The COM decides whether a country adequately implements reforms and sufficiently complies with the Copenhagen Criteria to become a full member. The specific requirements behind the CC accrue from the pre-accession strategy of the Council and the COM and comprise of eight types of legal documents which are being prepared by the COM and have been used during the accession procedure of the CEECs. They were first introduced in 1997 as a "system of benchmarking" between the accession candidates (Börzel/Risse, 2009: 41). Those documents belong to either one of two groups: those addressed to the individual candidate country (COM's Regular Reports/Progress Reports, COM's Comprehensive Country Monitoring Reports, Accession Partnerships) and general application documents which are based on the findings of the first group (COM Agenda, Yearly Composite papers, Strategy Papers, Comprehensive Monitoring Reports) (Kochenov, 2008: 67). Initially, the COM prepares an Opinion for every candidate country which is followed by an annual report outlining the country's accession progress throughout the previous year. White Papers further engage with a country's specific issues that have to be addressed. The Comprehensive Country Monitoring report then provides an individual reform plan that has to be implemented. The COM'S Composite (Strategy) Papers and Comprehensive Monitoring Reports summarize the findings of the individual Reports and include them into one, less specific overview. Based on the COM proposals, the Council releases Accession Partnerships which outline clear sets of priority areas for each country. Furthermore, Monitoring Reports on the State of Preparedness are being issued for every country, too (Kochenov, 2008: 76/77). The content of the Reports is the testing of the Copenhagen Criteria. In order to gather information on the status quo and

³¹ In fact, if any violations are suspected the EP can and does send fact finding mission groups to the respective countries as happened in Hungary after critically acclaimed legislative changes in 2012 as well as in Lithuania, Ireland, Italy, the Netherlands, France and Romania when potential fundamental rights breaches were being investigated (European Parliament, 2012)

progress of each CS, the Reports include replies to questionnaires delivered from the COM to the candidate country's governments, the results of bilateral meetings and bilateral agreements, assessments made by the member state, Reports and resolution from the European Parliament, and works of various international organizations (OSCE, Council of Europe), NGOs and other bodies such as think tanks (COM (97) 2000 final).³² The structure of the Regular Reports usually follows that of the Copenhagen Criteria and addresses Democracy and the RoL in summary in 1.1 of Sub-Chapter "B.1 Political Criteria" as well as more detailed under "Chapter 23 Judiciary and fundamental rights". The RoL related categories as well as specific requirements will be outlined in the next section.

II.2.6 The EU's RoL concept within the accession process

European Commissioner José Manuel Barroso acknowledged that "[t]he rule of law is one of the founding pillars of the European Union" and it "...is what [the] Union is built upon." He also emphasizes that "[t]he European Union has a crucial role in upholding the rule of law as the Guardian of the Treaties" (COM, 2014, IP/14/237). The EU's Justice Commissioner, Viviane Reding, asserts that the "Respect for the rule of law is a prerequisite for the protection of all other fundamental values upon which [the] Union is founded" (Ibid.), thus agreeing with the notion that a functioning rule of law is a precondition for liberal democracy. This part will outline the EU's general RoL concept as well as provide a specific checklist of demands the EU is making toward the candidate country during the accession process.

General EU RoL Concept

The general RoL conception of the European Union derives from principles set out in case law of the European Court of Justice and the European Court of Human Rights. It comprises of the following elements as summarized in a COM press release of a framework aimed at safeguarding the RoL within the European Union³³:

- (a) The principle of legality, which in substantial terms includes a transparent, accountable, democratic and pluralistic process for enacting laws;*
- (b) Legal certainty, which requires amongst other things that rules are clear and predictable and cannot be changed retrospectively;*

³² Bulletin of the European Union Supplement 5/97 AGENDA 2000. For a stronger and wider Union.

³³ European Commission, (2014): *European Commission presents a framework to safeguard the rule of law in the European Union*. Strasbourg: IP/14/237

- (c) *Prohibition of arbitrariness of the executive powers. The principle of the rule of law regulates the exercise of public powers and makes sure that every State intervention has a legal basis and is constrained by law;*
- (d) *Independent and effective judicial review, including respect for fundamental rights. The Court reiterated that the EU is a union based on the rule of law in which the acts of its institutions are subject to review of their compatibility with, in particular, the Treaties, the general principles of law and fundamental rights. Citizens are entitled to effective judicial protection.*
- (e) *There is a clear link between the right to a fair trial and the separation of powers. Only a tribunal which is independent from executive powers can guarantee a fair trial to citizens. The Court referred to an operational separation of powers implying an independent and effective judicial review, pointing out that "[...] EU law does not preclude a Member State from simultaneously exercising legislative, administrative and judicial functions, provided that those functions are exercised in compliance with the principle of the separation of powers which characterises the operation of the rule of law";*
- (f) *Equality before the law. The Court has emphasised the role of equal treatment as a general principle of EU law by stating that "it must be recalled that the principle of equal treatment is a general principle of EU law, enshrined in Articles 20 and 21 of the Charter of Fundamental Rights of the European Union".*

With view to the RoL categorizations previously introduced in the Literature Review (cf. II.1.2), the EU RoL concept is thick because it does not only guarantee formal and procedural RoL elements (i.e. legality or equality before the law) but also value-laden components (fundamental rights). Within Tamanaha's categorization, the EU RoL concept is substantive and would correlate with model 2, "Right of Dignity and/or Justice" because not only does it embrace individual rights but places human dignity as the base for any other fundamental right at supreme value. The concept introduced above is to be used when assessing the member state's compliance with the RoL criterion as part of the new framework to safeguard the RoL in the European Union which was introduced in 2014. However, within the accession process, the elements required of candidate countries have a different focus which will be

outlined subsequently. Similarly to the general EU RoL concept, there is no clear definition or checklist for its components for various reasons.³⁴

Specific EU Accession RoL Requirements

The European Union has been critiqued repeatedly for the non-existence of one comprehensive RoL concept that is implemented internally and externally (cf. II.4.1). That is not entirely true. In fact, a guidance note to the DG Enlargement within its 2005 enlargement package does outline the general criteria the Progress Reports' authors use to evaluate the RoL within the candidate countries. These consist of the elements summarized in the table below:

Judicial system
<p>Structure:</p> <ul style="list-style-type: none"> • Possible changes <p>Functioning:</p> <ul style="list-style-type: none"> • Assessment of judicial system • Including judicial capacity <ul style="list-style-type: none"> → Legal basis of judicial system → Independence and impartiality → Transparency and openness → Status and remuneration of judges → Number of judges → Number of judicial vacancies → Length of judicial proceedings → Training-including on good judicial practices → Human rights law, and European law → Equipment
Anti-corruption policy
<p>Indication of level of corruption in the country, linked to surveys and / or public perception.</p> <ul style="list-style-type: none"> • Existence / implementation of anti-corruption policy/programme/legislation • Institutional set-up in fight against corruption, division of tasks between various bodies involved • Existence and implementation of code of conduct for civil service / public administration • Relevant training for staff in public institutions – Cases of corruption in administration; reaction to these by relevant authorities – Ratification / implementation of main international conventions in the field – Involvement in Council of Europe Group of States Against Corruption (GRECO)

³⁴ Kochenov (2008), Zalewski (2004) and Guarnieri/Piana (2011) have speculated on the advantages and possible intentions of leaving the RoL criterion unspecific ranging from need to keep the accession process flexible over not being able (or willing) to agree on one definition in order to not offend any MS who has a different judicial system to intentionally not providing one so as to not having to automatically include CS once they met the targets.

Table 6 - Rule of Law elements included in Guidance Note 2005, Enlargement package to DG Enlargement

When considering the aspects listed above it becomes visible that there is a lack of clear definition and differentiation within the Guidance Note. Additionally, no explanation is provided as to when a satisfactory level of compliance is reached or how to evaluate the criteria in general. In order to provide a clearer picture of the standards that have to be met by candidate countries when seeking to fulfill the RoL criterion, a synopsis of Notes, Opinions, and Recommendations from several EU institutions will be assembled in the following sections. First, the general EU accession RoL concept will be introduced, then its individual elements and meaning will be deducted from previous Progress Reports, and last, specific items to be used as benchmark criteria for the subsequent comparison will be deducted from said synopsis. The results are summarized in Table 7.

Before accession negotiations start, the democracy requirement (political aspect of the Copenhagen Criteria) trumps all other conditions as emanates from the Commission's 1999 Composite Paper (p.30, In: Wennerström, 2007: 164). Without a functioning democracy, a country has no chance of becoming a member state. However, once the negotiations have begun, "democracy [is] a closed chapter" and the RoL element starts to dominate the accession process (Ibid: 216). Although there is a hierarchy of both concepts they are deeply intertwined in protecting civil rights and liberties because "democracy sets the standards for the formal method by which laws are made, whereas the rule of law sets quality standards for the law itself" (Lautenbach, 2013: 63). This means that voting is merely a measure to determine who is governing but it does not affect the contents or the quality of the governing process and the rules being made as a result of it. In contrast, the RoL includes qualitative requirements of the laws being enacted and is able to enforce them vis-à-vis the legislative and executive. The RoL requirement has de facto been introduced and applied since 1993 as part of the Copenhagen Criteria although it has not been legally required until the Amsterdam Treaty took effect in 1999 (Ibid.). According to the Commission, the RoL is a founding principle of the European Union and the "fundamental principle of any democratic system seeking to foster and promote rights" (COM (98) 146, p. 2, In: Wennerström, 2007: 172).

According to the European Commission, the RoL within the accession context consists of five elements: *the supremacy of law, the separation of power, respect for fundamental rights and*

*freedoms, an independent judiciary and the prevention of corruption.*³⁵ *Supremacy of the law* means that in order to protect the civil rights and freedoms within a liberal democracy, there must be specific laws that are binding to and respected by all individual and institutional members of society, including all three state branches. This is crucially important because no matter what the content of the law, there will be infringements by one of the parties affected by it. “Should [such] an infringement occur, legal remedy must be available, for if there is none, the law will lose its proper function” and as a result people will stop abiding by it (To, 2006: 43). Furthermore, a state is governed by the rule of law if its individual organs “are aligned and limited in such a way that the state cannot illegally infringe on citizens’ rights” (Wennerström, 2007: 172). This means that there is a system of checks and balances to protect the citizens’ interests vis-à-vis the disproportionately more powerful state actors. All of them need to respect and protect the *fundamental rights and freedoms* granted by the laws. Since laws are “inanimate creatures” that cannot defend themselves, there needs to be a body capable of enforcing them vis-à-vis the state branches that drafted or implemented the respective policies (Ibid.). The precondition for being able to make decisions challenging the other state branches is that the *judiciary* is independent. Last, corruption hinders norms and institutions from being implemented and functioning properly which is the cornerstone of the RoL.

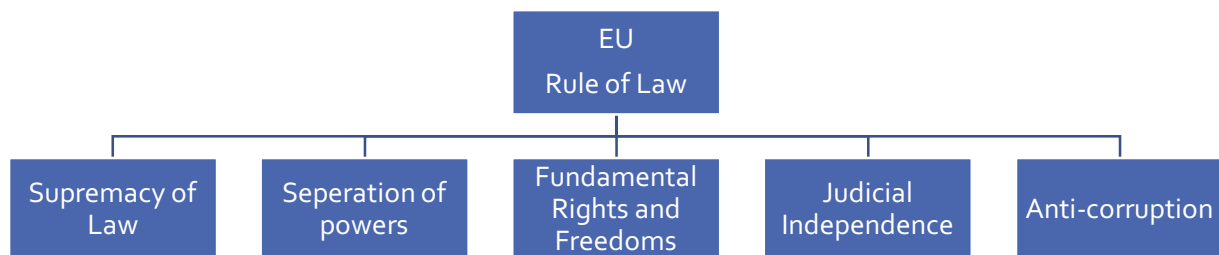


Illustration 4 - Five Elements of the EU accession RoL concept

³⁵ The reason for the lack of a codified concept can be searched at the level of the Member States. The European Council actually attempted to create a legal definition of the rule of law in the area of Justice and Home affairs that could be used for the enlargement context. As Wennerström (2007: 172) polemically asserts, “The efforts were, however, abandoned as Member States realized that they were in fact creating requirements they would have to survive being compared to themselves”. In the meantime, the Commission went ahead using the procedures suggested in Agenda 2000 (including the Copenhagen Criteria) while the Member States “took their time in deliberating on the issue” without ever agreeing on the sought-after legal definition (Ibid.)

In order to understand how the European Commission operationalizes this rather broad concept the thesis will use the criteria that are directly and indirectly tested within the Commission’s yearly Progress Reports for candidate countries. Under Chapter 23 Judiciary and Fundamental Rights, the Commission splits the RoL into six subcategories, which are partially congruent with the elements introduced above, namely judicial independence, anti-corruption, and fundamental rights (although the latter will not be part of this thesis’ analysis). The remaining two aspects (supremacy of law and separation of powers) are preconditions for the opening of negotiation talks and therefore not revisited at this point. Instead, four other subcategories are being evaluated within the Reports: judicial impartiality, judicial accountability, professionalism/competence, and efficiency. Following the summarizing Illustration 6, all of the subcategories will be explained and specific EU requirements highlighted.

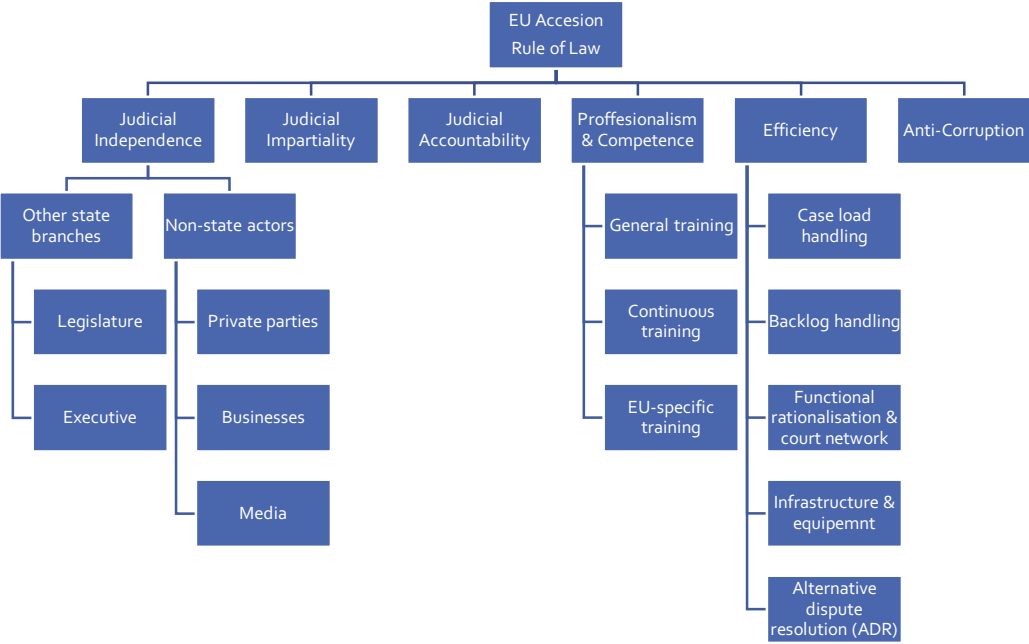


Illustration 5 - EU rule of law requirements gathered from Chapter 23 Judiciary and Fundamental Rights

Each of the six sub-categories of the EU accession RoL Concept consists of a number of elements whose requirements will now be presented. Although there exists no single document outlining them, a synopsis of the requirements the COM used in previous accession rounds (specifically the CEECs 2004 and 2007) generates a clearer picture of the individual aspects.

Judicial Independence and Judicial Impartiality

“The overall objective of guaranteeing judicial independence is to ensure a reasonable perception of impartiality, judicial independence is but a “means” to this “end”.

Chief Justice Lamer (In: Dodek/Sossin, 2010: 3)

Among the RoL elements, judicial independence is considered to be *primes inter pares* (Dodek/Sossin, 2010: 43) because it enables impartiality. Impartiality is the judiciary’s biggest asset because the only reason for people to abide by the laws and seek conflict resolution within the justice system is that they can trust to be judged neutrally and thus, fairly. This is particularly important because the judiciary is the only institution that can alleviate the “sharp power asymmetry” between citizens and public institutions (Stein, 2010: 582). However, as Chief Justice Lamer rightfully asserts, impartiality is merely a perception which means that it is hard to operationalize let alone measure it. Consequently, a proxy is needed that can be both operationalized and measured: judicial independence. Both terms will be defined before examining their relationship further. According to the Supreme Court Canada, *judicial impartiality* “refers to a state of mind ... of the tribunal in relation to the issues and the parties in a particular case. The word ‘impartial’ ... connotes the absence of bias, actual or perceived” (Supreme Court Canada, Valente vs. The Queen, In: To, 2006: 86). This means that the judge must not have preconceived ideas about either the parties to the conflict nor its resolution before being presented with the facts of a case. This state of mind or attitude towards judgments strongly depends on the type of relationship the judiciary has with other actors. The more it is entangled with them the less likely it can be impartial. The specific requirements for judicial impartiality are outlined in a Council Note from 1998.³⁶ Point 2 reads:

“To ensure the impartiality of judges, every conceivable measure should be taken, and in particular it should be made impossible to remove judges from office, care should be taken to ensure the widest separation of courts from the activities of public prosecutors,

³⁶ EUROPEAN UNION THE COUNCIL NOTE Brussels, 3 July 1998 (09.07) (OR. d) 10116/98 LIMITE CK4 32 PECOS 96 from : to : Presidency K.4 Committee Nos. prev. docs.: 10019/97 JUST 6; 7830/98 CK4 20 Subject: Common principles for the rule of law in criminal matters <http://database.statewatch.org/e-library/10116.en98-rule-of-law.pdf>

administrators, legislators and the government, as well as appropriate remuneration, and appointments to the courts should be made on the basis of objective criteria. Decisions concerning the partiality of a judge and disciplinary measures against judges should be a matter for the courts alone.“

Judicial independence is defined as the “status of the judiciary to be free from undue influences from any one” that derails the judges from making their case-decisions based “on the facts ... and their convictions of applicable laws” alone (To, 2006: 43, 75). “Any one” can be anyone from parties to the conflict (individuals, companies, other state branches) over external interest groups (media, lobbies) to internal institutions (hierarchies within the judiciary, training/education/literature that creates the judges understanding of the law) (Ibid: 75). Judicial independence is not an end in itself and it is not “good per se” (Piana, 2010: 24). Instead, it is a tool to foster the impartiality of the adjudication process by enabling the judges to “approach the issues with an open mind, ready to respond to the legal and factual merits of the case” without the undue influences of “local governments, vested interest of any kind public and parliamentary opinion, the media, political parties and pressure groups, their own colleagues, particularly those senior to them” (Bingham, 2010: 92-93). Judicial independence consists of formal guarantees that ensure impartiality which in turn provides the judge's decisions with the necessary legitimacy (Shapiro, 1981). Furthermore, the specific conditions of judicial independence can be used to assess the legitimacy of a democratizing country (Dakolias, 1999). Moreover, experts on democracy promotion (Carothers, Dietrich, Hammergreen, Linz and Stepan) agree that judicial independence can create “favorable conditions for the establishment of democratic institutions” (Piana, 2010: 3). Within a Recommendation from The Council of Europe from 1994, the Committee of Ministers delivers the following definition and requirements of judicial independence:

“In the decision-making process, judges should be independent and be able to act without any restriction, improper influence, inducements, pressures, threats or interferences, direct or indirect, from any quarter or for any reason. The law should provide for sanctions against persons seeking to influence judges in any such manner. Judges should have unfettered freedom to decide cases impartially, in accordance with their conscience and their

*interpretation of the facts, and in pursuance of the prevailing rules of the law. Judges should not be obliged to report on the merits of their cases to anyone outside the judiciary.*³⁷

As becomes clear, judicial independence is a precondition for judicial impartiality which in turn is the cornerstone of the RoL. The European Parliament confirms that notion when stating that “the independence of the judiciary is one of the pillars upholding the rule of law and fundamental to the effective protection of the rights and civil liberties of all” (In: Wennerström, 2007: 164). There has to be the right degree of judicial independence. As Russell (2010: 601) proclaims “there is no way in which judges can be totally disentangled from all power or influences outside the judiciary. Nor would we want such a total separation “. In fact, absolute judicial independence is unrealistic and would also be hurtful for two reasons. Unrealistic, because if the judiciary had to carry out all the organizational and administrative tasks itself, it would have no capacities for actually judging. Hurtful, because the system of checks and balances among the three state branches has been put in place so that no single branch can overstep its constitutionally prescribed boundaries. If the legislative and executive have absolute no influence on the judiciary, it could abuse its power vis-à-vis the other branches and the citizens which is aggravated by the fact that it has not been democratically legitimated to begin with. That is why the concept of judicial accountability has been established as a security mechanism in liberal democracies’ RoL systems.³⁸

Specific EU accession requirements for Judicial Independence and Judicial Impartiality

Since autocratic regimes used the judiciary as a tool to implement their power and interests, judicial independence (JI) is the overriding goal of all EU accession related reforms. The quality of JI can range from full politicization to limited political influence, a hierarchic structure or even self-government, which is the goal of EU accession reforms.³⁹ The three

³⁷ COUNCIL OF EUROPE COMMITTEE OF MINISTERS RECOMMENDATION No. R (94) 12 OF THE COMMITTEE OF MINISTERS TO MEMBER STATES ON THE INDEPENDENCE, EFFICIENCY AND ROLE OF JUDGES (Adopted by the Committee of Ministers on 13 October 1994 at the 518th meeting of the Ministers' Deputies

³⁸ See Piana (2010) for a thorough account of what types of judicial accountabilities exist within the EU enlargement context.

³⁹ This spectrum was introduced by Guarnieri/Piana (2011: 116/7). Full politicization means that “politics can exert direct influence on all significant elements of the status of the judges: recruitment, career, discipline, removal” which is likely to influence judges in their decision-making (Ibid.). Limited political influence is targeted only at the top hierarchy of the judiciary such as appointing Supreme Court judges who can then influence lower-level judges through their decision-making. Within a hierarchic model, individual judges enjoy a higher level of external independence but a high level of internal dependence because their career entrusted to

criteria the EU checks with view to JI are the institutional independence of the judiciaries, the independence of the individual judge as well as the budgetary independence of the judiciary (Kochenov, 2008: PU, FN 177). Institutional independence means that there are legal structures or rules in place that are designed to protect judges from “improper influence or pressure and thus promote independent decision-making” (Jackson, 2012: 21). Institutional independence can be dissected into substantive and personal independence. Substantive institutional independence means that the judge is subject only to the law.

In order to grant his institutional independence from the other state branches, the legislative and executive may not invalidate court decisions except for specific cases (amnesty/pardon), cannot intrude on court administration, and must not interfere with decisions upon judges’ competences.

In order to secure this the judiciary needs to have a self-governing administrative body that grants it independence from other state branches. Usually, this body is called the Judicial Council (JC) and is ideally shaped after the Neo-Latin model. According to Kochenov (2008: PU, FN188), it must function effectively, set its own agenda, and needs to “have a sphere of competence distinct from that of the Ministry of Justice”. Furthermore, the aforementioned CoE Recommendation No. R (94) 12 requires that

“...all decisions concerning the professional career of judges should be based on objective criteria, and the selection and career of judges should be based on merit, having regard to qualifications, integrity, ability and efficiency. The authority taking the decision on the selection and career of judges should be independent of the government and the administration. In order to safeguard its independence, rules should ensure that, for instance, its members are selected by the judiciary and that the authority decides itself on its procedural rules. However, where the constitutional or legal provisions and traditions allow judges to be appointed by the government, there should be guarantees to ensure that the procedures to appoint judges are transparent and independent in practice.”

Consequently, the JC should be in charge of the general administration of the courts and supervise the way the judiciary functions in the respective country. More specifically, the

higher ranks. Self-government is the highest form of JI because it grants a high level of both external and internal independence by severely restricting political influence and the role of higher ranks through transparent processes and clear criteria for career decisions.

Council should hold the power over selection, appointment, promotion, evaluation and disciplinary control of the judges and prosecutors. Its set-up is crucial for its function because only when the majority of its members are judges elected by their colleagues with equal voting rights one can secure the body's independence from other state branches (which are usually also represented in the Council, although with minority seats) (Piana, 2010: 58).⁴⁰ The minority members are usually legal scholars appointed by Parliament but may also be appointed by the Ministry of Justice (MoJ). The Council is also in charge of creating and implementing judicial training programs (e.g. in Judicial Schools or Judicial Academies). Personal institutional independence safeguards the judge's decisions to be made without fearing negative consequences, wherefore "matters of removal, discipline, career, and salary need to be strictly regulated" (Guarnieri/Piana, 2009: 5). The selection of judges must be conducted in a transparent, non-discriminatory, non-biased and independent way. This means that there must be clear selection criteria and specific rules against patronage, incompetence, nepotism and exclusion (Kühn, 2012: 610). The process must be open to anyone independent of race, gender, class, social origins or former professions (Ibid.). Also, the power of selection must de facto be with the presidents of the regional courts with the MoJ playing merely a formal role. Moreover, personal networking may enable applicants to have better chances in the process in order to avoid the cementation of existing power hierarchies in the courts (Ibid.). The appointment can be carried out for a probationary period in order to assess the applicant's qualification as long as it is not abused to "select judges who are most loyal to the authorities" (Nußberger; 2012: 891).

The independence of the individual judge correlates with the notion of personal institutional independence. Here, additional requirements concern tenure and remuneration. The aforementioned CoE Recommendation demands a guaranteed tenure until a mandatory retirement or the expiry of the judge's term in office. The EU does not take an absolute stand on tenure because there are advantages and disadvantages of life-long appointments. A long tenure is usually associated with greater JI because the judge does not have to be concerned with reappointment. However, renewable terms emphasize the quality of the candidate. As long as the re-selection process is carried out impartially and purely merit-based, this can

⁴⁰ The majority principle is based on Recommendation Nr. 32 of the Venice Commission Final Report, Note 36 (In: Nußberger, 2012: 89). The idea behind it is that a body made up of judges only will lead to judicial corporatism where "judges lose sight of the context they operate in and take decisions not accepted by society" while a majority of other state branches' representatives endangers the independence of the organ (Ibid.).

increase the overall quality of the justice system (Jackson, 2012: 24). Naturally, those “ifs” are usually the stumbling block of transforming judiciaries. With view to remuneration, the judges’ pay should be “commensurate with the dignity of the profession and burden of responsibilities” and of a level that is “fixed so as to shield [the judges] from pressure aimed at influencing their decisions and more generally their behavior within their jurisdiction, thereby impairing their independence and impartiality” (CoE CM R 95 + CoE European Charter on the Statute for Judges, §6.1). Additionally, the judges’ incomes should be transparent and their asset declaration be reviewed by the self-governing body (Kochenov, 2008, PU). When paying each judge the same salary irrespective of his speed or quality of work, motivational incentives are limited to promotions. Regarding the level of salaries, they need to be high enough to attract specialists but without distorting the general wage structure. However, Nußberger (2012: 892) warns that merely raising salaries does not diminish corruption as this is a question of ethical culture.

Budgetary Independence means that the JC possesses budgetary competences in order to secure enough founding to fulfill its task properly. This means that the Council must be allowed to draft its own budget without too much influence of the legislative. It then proposes the budget to the Parliament which has to vote upon in. At the time of the CEECs accession, the budget was supposed to equal 2-4% of the country’s GDP (Kochenov, 2008: PU, FN 177). However, the Council must not get overinvolved in budgeting since this is a prerequisite of democratically legitimized politicians and the JC must avoid its own politicization or inadequate financing. The cooperation between judiciary and executive therefore usually persist in the former’s right to prepare a budget and be consulted prior to its finalization and voting by the latter (Kühn, 2012: 609).

Judicial Accountability

The concept of accountability describes an environment (e.g. internal norms or external means) that constrains individuals in exercising their power.⁴¹ In the JI context, the actors

⁴¹ Piana (2010: 28-35) introduces five types of accountabilities: legal, institutional, professional, managerial, and societal. Legal accountability refers to the mechanisms of legal control of law (e.g. judicial review), institutional accountability addresses competence and impartiality issues, professional accountability focuses on peer control of law, managerial accountability stresses efficiency standards, and societal accountability “covers any kind of control of law exercised by private actors” (Ibid., 31) such as civil society organizations or citizens. Overlaps between Piana’s categorization and the EU rule of law requirements can be detected: institutional accountability addresses the elements of impartiality and professionalism; professional accountability covers the areas of professionalism and competence; and managerial accountability highlights the efficiency criterion.

whose powers have to be limited are those of other state branches (particularly the executive) vis-à-vis the judiciary. In the judicial accountability context, the judges' powers must also be scrutinized by the judiciary itself, by parties to a conflict and civil society. JA is a safeguarding mechanism so that the freedom invested in the judiciary through JI is not abused. As Shapiro succinctly summarizes, no regime "is likely to allow significant political power to be wielded by an isolated judicial corps free of political restraints" (1981: 34). Consequently, the courts are expected "to be independent at retail, not wholesale...to make their decisions according to the law...to be...the servants of the lawmakers" (2001: 280, In: Guarnieri/Piana, 2009: 4). The three main features of accountability are transparency (information), justification for decision-making as well as reaction (punishment/compensation/ declaration of adequacy of decision) (Morlino, 2010: 54). Specific elements that help instill JA include judicial reviews, peer reviews and review mechanisms for civil society. Additionally, disciplinary proceedings as well as requirements to disclose income may help to safeguard the judge's responsible use of his power.

Professionalism/Competence

Many countries that transition to democracy either employ judicial professional who have been trained under totalitarianism or newly recruited legal professionals whose training is below EU standards. However, JI only works if the judiciary's staff is professionally and personally competent. Professionally, lawyers and judges must be trained adequately in domestic and EU law and undergo continuous training after. Although there are no specified standards, requirements that are repeatedly included in the CEECs Progress Reports include the government-funded building of a Judicial Academy (theoretical training) as well as in-practice training "in the courts and, where possible, with other authorities and bodies before appointment and during the career" (CoE Recommendation, In: Kochenov, 2008: PU, FN 294) According to the aforementioned Recommendation, "such training should be free of charge to the judge and should in particular concern recent legislation and case-law. Where appropriate, the training should include study visits to European and foreign authorities as well as courts". The JC should also provide a long-term concept for judicial schools and create uniform national curricula for the law students (Ibid.). On the individual level, Piana (2010: 52-55) argues that the higher the degree of ethical integrity of the judges the less likely they will be open to undue influence such as corruption or partisan interests. Consequently,

she emphasizes the importance of education instruments within the vocational training that focus on ethics and integrity.

Efficiency

The reasons for an inefficient justice system are complex: most transitioning judiciaries are generally poorly equipped and organized or simply overburdened with the adaption of new rules and procedures inherent in the democratization process. Additionally, there is an excessive protection of certain individuals which is hindering cases. All mentioned aspects lead to large case backlogs as well as long trial proceedings and a lack of verdict implementation. Ideally, the courts have the proper facilities as well as being sufficiently staffed and equipped to carry out their functions in a time-sensitive manner in order to avoid the backlog of cases. In order to create such an environment, recommendations in the COM Progress Reports for Lithuania, Hungary, the Czech Republic and Slovakia (1999/2000) included the computerization of courts (case-management tools, court information systems, publishing court decisions, transparent automatic case assignment systems), stricter and clearer rules (in absentia regulations, shorten cassation procedures, accelerate minor crimes proceedings) as well as productivity boosters (reduced workload for judges using legal assistants, financial incentives to increase productivity) (Kochenov, 2008: PU). Time can be saved by taking the appropriate measures “to assign non-judicial tasks to other persons, in conformity with Recommendation No. R (86) 12 concerning measures to prevent and reduce the excessive workload in the courts” (Ibid.).

Anti-corruption Measures

Corruption is in many transitioning countries one of the key obstructions to justice. In fact, Rose/Shon (2001: 341) find that it embodies “a more pervasive threat to the rule of law than...political oppression”. Corruption disables the equal treatment of all citizens which undermines the public trust in office holders, and thus, the legitimacy of the political system and civil service. While the EU has no binding Acquis on corruption, it does rely on soft laws (international agreements and measures covered by the Justice and Home Affairs Pillar) to make recommendations to the candidate countries. Those include the creation of a code of conduct, regular reporting on the matter, the disclosure of conflicts of interests as well as refusal of favors and gifts. These measures aim to ensure that a public official does “not allow his or her private interest to conflict with his or her public position. It is his or her responsibility to avoid such conflicts of interest, whether real, potential or apparent.”

Furthermore, he “should never take undue advantage of his or her position for his or her private interest”.⁴² In order to do so, the catalogue of recommendations includes the reporting of any corrupt events or assumptions thereof; the avoidance, disclosure and recusing of oneself in case of a conflict of interest; the refusal of gifts or favors as well as the effective, efficient and economical use of public and official resources (Ibid.). Moreover, the creation of transnational judicial networks and the support of an active civil society should help foster a democratic political culture and respect for the law (Tunkrova/Saradin, 2010: 48). The following table summarizes each EU accession RoL category, it’s criteria, elements and specific requirements.

EU accession RoL Requirements			
Category	Criterion	Element	Requirements
Jl⁴³ JIm⁴⁴	Institutional	Self-governing body <ul style="list-style-type: none"> • Selection • Recommend for Appointment • Evaluation • Promotion • Disciplinary Proceedings 	<ul style="list-style-type: none"> • Majority of judges selected by peers • Minority of legal scholars appointed by Parliament • Independence of MoJ • In charge of admin. Tasks • Objective criteria • Clear rules against nepotism/patronage • Non-discriminatory • Independent of MoJ • Unbiased distribution of cases
	Individual	Pay	• Adequate remuneration
		Tenure	• Impossible to remove
	Budgetary	Budget	<ul style="list-style-type: none"> • Right to prepare • Right to be consulted prior to

⁴² Recommendation No. R (2000) 10 of the Committee of Ministers to Member States on Codes of conduct for public officials

⁴³ Freedom from undue influences from any one, decisions made solely based on law and professional interpretation of it.

⁴⁴ Defines as the bsence of actual or perceived bias in judgment.

			finalization/voting <ul style="list-style-type: none"> • Sufficient to carry out functions (2-4% of GDP)
JA ⁴⁵			<ul style="list-style-type: none"> • Judicial review • Peer review • Review mechanisms for civil society
Prof./ Competence ⁴⁶		Training (initial/continuous)	<ul style="list-style-type: none"> • Adequate initial and continuous training in domestic and EU law • Adequate staffing, facilities, infrastructure (IT, court management)
Efficiency ⁴⁷		Computerization	
		Fighting Backlog	<ul style="list-style-type: none"> • Speedy proceedings to avoid large backlog of cases • Delegation of non-judicial work
Anti-Corruption ⁴⁸		Political culture of respect for law	<ul style="list-style-type: none"> • Code of conduct/ethics • Reporting • Avoidance/disclosure of a conflict of interest • Refusal of gifts/favors • Effective, efficient and economical use of public and official resources

Table 7 - Summary of EU RoL accession requirements for candidate countries

⁴⁵ Defined as constraints on judicial power of individuals (judges) to prevent abuse.

⁴⁶ Defined as the adequate training to fulfill judicial function.

⁴⁷ Defined as the optimal usage of infrastructure and resources.

⁴⁸ Defined as measures taken that prevent the abuse of professional power for personal gain.

II.3 Measuring the Rule of Law

If external democracy promotion, specifically financial aid, is contingent on fulfilling specific conditions. Consequently, democracy promoters must be able to assess a country's democratic progress. That is why democracy measuring indices have been developed as quantitative and/or qualitative empirical analysis tools that help gain an insight into the democratic quality of a given country at a given time (Pickel, 2006: 159; Lauth et.al., 2000: 12). As Coppedge et.al. put it, “without some way of analyzing regime types through time and across countries we have no way to mark progress or regress on this vital matter, to explain it, to reveal its consequences, or to affect its future course” (2011: 247). This part of the thesis will address all the relevant questions surrounding such evaluation tools: Why measure democratic/RoL progress? Who measures it? How should adequate measuring be conducted? Which indices have been selected for the purpose of this thesis and why? Are their methods adequate to provide valid results? In comparison, which of the indices is methodologically stronger or weaker than the other and why? Which implications can this have for the results of the comparison?

II.3.1 Actors and Motives behind Democracy and Rule of Law Evaluation

There are about as many democracy measuring indices as there are concepts of democracy itself.⁴⁹ Although not all of them can be clearly classified, Pickel/Pickel suggest a categorization into four groups (2006: 158). The first group relies on a strong institutional measuring level (Polity, Lijphart, and Bollen). The second one follows an objectively participatory approach such as Vanhanen. The third group examines constitutional realities (Freedom House) whereas the fourth uses qualitative approaches (Democratic Audit, Saward, Elklit). Initially, democracy measuring indices were constructed in a bipolar way in order to decide whether a country is authoritarian or democratic. Since the beginning of what Huntington coined the “Third Wave of Democratization” in the early 1990s, indices focus on the spectrum between the poles. „Es ist die Grundannahme dieser Studien, dass jedes Regime

⁴⁹ E.g. Dahl's Polyarchieindex, Vanhanen's Demokratie-Index, Freedom House, Polity IV, Democracy Audit, World Value Survey¹², Transparency International, Bertelsmann Transformations-Index (BTI), UNDP HDI¹⁵, Vetospieler-Ansätze, Index des institutionellen Pluralismus, Index ‚Institutionelle Schranken der Exekutive‘, Executive-party index (Lijphart), Federal-unity Index (Lijphart), Präsidentialismus-Parlamentarismus Indizes, Index defekter Demokratie (IDD), Bollen's Index of Political Democracy (POLDEM), Neuer Index Demokratie (NDI), Demokratiebarometer, the Economist Intelligence Unit's Democracy Index (Mayer, 2008:43).

entweder gar nicht oder aber in gewissem Maße demokratisch sei, so dass es vor allem darauf ankomme, Grade der Demokratisierung zwischen null und einem bestimmten Maximalwert zu bestimmen” (Lauth et. al. 2000: 9). Consequently, rather than making dichotomous assessments, modern indices measure the level of democracy in intervals.

Democratic progress is measured for multiple reasons: academic curiosity or economic interest in the development of a country, political consultancy, enabling comparisons between different countries, potential analysis for investments or aid programs. “Billions of dollars in foreign aid intended to promote democracy and governance in the developing world is contingent upon judgments about how democratic a polity is at the present time, its recent history, future prospects, and the likely causal effects of giving or withholding assistance” (Coppedge et.al., 2011:248). The Worldbank has implemented democracy relevant criteria as part of their structural adaptation plans (Beetham/Weir, 2000: 75). Within the realms of political conditionality, the criterion of ‘good governance’ focuses on the RoL, stability of the law, accountability and anti-corruption measures (Nuscheler, 2004: 625). Multiple indices are being consulted to assess the level of progress or recess in the target countries. Moreover, indices like the Human Development Index or Freedom House have been used to design conditionally criteria for development aid (Lauth et.al. 2011: 11). In line with its accession process, the European Commission measures the democratic progress of its candidate countries to assess their level of maturity in meeting the Copenhagen Criteria. Moreover, the indices may not only be used to justify policy decisions or decide upon the level of financial aid. They can actually be crucial in shaping our normative views on how a society should be politically and economically oriented and shaped. As Giannone (2010: 70) put it:

“One can consider measuring instruments as genuine methods for understanding social reality; however, they are not politically and ideologically neutral. Therefore, to the extent that they are not neutral, they can be used as tools for acquiring and/or strengthening such a hegemony. [Since some of the indices have] become the global pattern-setter[s] of democracy, i.e. the real creator of models throughout the world for the discourse on democracy, [the] indexes are often considered as a condition (i.e. a sort of natural order of things) rather than a narration (i.e. a product of human action). Consequently, they are often used uncritically as the way to describe the state of democracy in the world, blurring over the fact that their underlying political and scientific framework is questionable.”

Thus, democracy measuring indices can become powerful tools not as much in mirroring as in shaping reality which, as is openly disclaimed by some (Freedom House and BTI for example), may be a specific purpose of their existence. Therefore, a critical acknowledgement of their backgrounds and methodologies will ensue.

II.3.2 Quality Criteria for Adequate Measuring

The currently most distinguished quality control model for democracy and RoL measuring indices has been developed by Munck/Verkuilen (2002). This has been referenced and applied regularly by other scholars in the field and will therefore be used for the purposes of this thesis as well.⁵⁰ According to their model, adequate democracy measuring is contingent on three factors: proper conceptualization, the adequate choice of measurement techniques, and the appropriate selection of aggregation procedures (summary see Table 8). An index is the combination of multiple indicators into a list of measurement readings (Schmidt, 2004: 311). Indicators are observable and measurable constructs (proxies) that serve to indicate not or not directly observable phenomena. In order to create a democracy index, the designer must create a “tree” consisting of a “stem” (definition of democracy), “branches” (indicators) and “leaves” (attributes that make up the indicator) (see Illustration 7). When *conceptualizing* democracy, the index designer must avoid minimalist and maximalist definitions. Minimalist definitions will lead to too undifferentiated evaluations that might be helpful to determine whether a country is a democracy or not but do not serve to assess the actual quality of democracy. Maximalist definitions may create a category of every case observed and risk a conflation of the individual indicators or attributes. In order to avoid both extremes, the index designer must make sure to create a vertical organizational diagram of their democracy concept which becomes more specific in each step. “Attributes at the same of level abstraction should tap into mutually exclusive aspects of the attribute at the immediately superior level of abstraction” (Munck/Verkuilen, 2002:12-14). Attributes must be chosen in a way that avoids both redundancy and conflation. As Giannone (2010: 72) correctly acknowledges, “the measurement framework used influences perception, and as a result determines which rights may be perceived as being demandable and which are not.” Therefore, the choice of indicators is crucial because they are at the “core of the structuring of the measuring instruments” and thus, “affect the perception of democracy by the community of citizens” (Ibid.). Giannone emphasizes, that “favouring some indicators over others also serves to

⁵⁰ Coppedge (2002), Marshall et.al. (2002), Pickel/Pickel (2007)

enable and shape a specific definition and measure of democracy” (Ibid.) and offers a convincing example. If a special focus within the measurement is put on the “acknowledgement of civil and political rights” this “works for the construction of a liberal-democratic framework, in which social rights are considered as ‘secondary rights’, or social services provided only to particular categories of users” (Ibid.). Consequently, he emphasized that the choice of methodology – and the conception it is based on – cannot be politically neutral and therefore must be analyzed respectively.

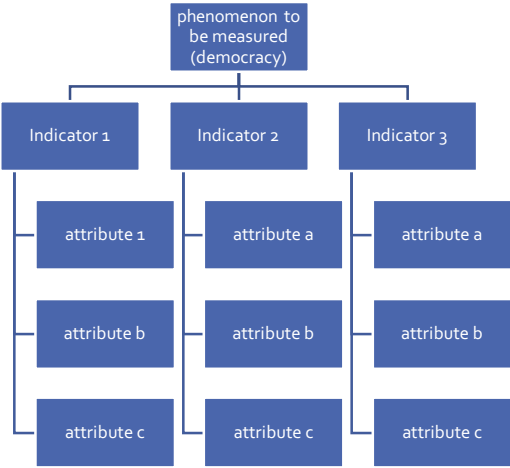


Illustration 6 - measuring tree according to Munck/Verkuilen (2002)

With view to the adequate *measuring*, Munck and Verkuilen require the proper selection of indicators and measurement level as well as the recording and publishing of the coding rules, the coding process and the disaggregate data (2002: 15-22). Those tasks are necessary to ensure the validity, reliability and replicability of the process. When selecting the indicators, the index designer must avoid two pitfalls: choosing indicators that are not selective and/or equivalent and not appreciating the chances of measurement errors. This can be done by acknowledging biases (e.g. that the choice of indicators may be influenced by the availability of data or that increased evidence of an indicator may be due to the fact that it has not been focused on and tested before). According to Bollen (1986: 578) the validity of the indicators can be maximized by removing potentially biased indicators and only use those that “*can be cross-checked through the use of multiple sources*”. Regarding the selection of the measurement level, the index designer can choose between nominal, ordinal, interval, or ratio scales. In order to secure the validity of measures, the measurement level should generate a

high level of homogeneity with as little distinctions as possible. Moreover, theoretical justification and testing are needed wherefore analysts should assess the influence of different assumptions has on their test results (Gifi, 1990; Jacoby, 1991). In order to secure the reliability of measures, the process needs to be arranged in a way that different coders will produce the same codings. Reliability tests are necessary because if they prove weak they may hint at problems in the measurement process. However, they do not specify what part of the process is flawed. Moreover, even strong reliability does not automatically ensure a high quality of the measures because coders may be similarly biased and therefore unconsciously produce similar but faulty codings. Replicability is another element of ensuring the quality of a measuring process. Independent scholars can only replicate the measuring if the coding rules (including a list of all indicators and their respective measurement levels), the coding process (including the sources used), and the generated disaggregated data have been made publicly available.

With view to appropriate *aggregation*, Munck and Verkuilen (2002: 22-25) demand the proper selection of aggregation level and rule as well as their publication. When choosing the level of aggregation, the index designer must be aware that the higher the level of aggregation, the more reduced the level of validity becomes. Therefore, he must make and justify his decision theoretically. When selecting an aggregation rule, the goal is to clearly identify “what attributes are to be aggregated and in what order” (Ibid.). This means that the analyst must prepare a theoretical justification for the linkage of attributes. If both attributes have the same weight, they can be added; if both are necessary features, their scores can be multiplied; if both are sufficient, it suffices to take the higher score (Ibid.). As is the case with the coding rules, the aggregation rules and process must also be published to ensure both the robustness of the aggregated data and its replicability.

Challenge	Task	Standard of Assessment
Conceptualization	Identification of attributes	Concept specification: Avoid maximalist definitions (the inclusion of theoretically irrelevant attributes) or minimalist definitions (the exclusion of theoretically relevant attributes)
	Vertical organization of	Conceptual logic: Isolate the "leaves" of

	attributes by level of abstraction	the concept tree and avoid the problems of redundancy and conflation
Measurement	Selection of indicators	Validity: Use multiple indicators and establish the cross-system equivalence of these indicators; use indicators that minimize measurement error and can be crosschecked through multiple sources Reliability
	Selection of measurement level	Validity: Maximize homogeneity within measurement classes with the minimum number of necessary distinctions Reliability
	Recording and publicizing of coding rules, coding process, and disaggregate data	Replicability
Aggregation	Selection of level of aggregation	Validity: Balance the goal of parsimony with the concern with underlying dimensionality and differentiation
	Selection of aggregation rule	Validity: Ensure the correspondence between the theory of the relationship between attributes and the selected rule of aggregation
	Recording and publicizing of aggregation rules and aggregate data	Robustness of aggregate data Replicability

Table 8 – Quality Criteria for measuring index (Munck/Verkuilen, 2002: 8)

In practice, the differences between indices' set-ups are usually based on their divergent goals. While some are merely testing one or a small number of countries, others compare a multitude of cases. The more countries are being tested and the greater the difference among

the countries, the less exact indices using short or small-range scales tend to become (Lauth et.al., 2000: 15). Often, this also affects the depth of analysis because the amount of indicators will be reduced for temporal or economic reasons (Mayer, 2008: 41). Other limiting factors are the availability of raw data or the ability to generate it by oneself. Especially in long term studies, changes in external data generation may affect the comparability. Also, changes within the democratic quality of a country can affect specific indicators and thus tamper with the results. Munck/Verkuilen (2002) offer the example of corruption which may not rise itself but which becomes more visible once the freedom of the press is strengthened.

II.3.3 Research Gaps: Indices

In contrast to FH and Bertelsmann, the European Commission maintains a low profile regarding its methodological approach towards the progress report creation. In order to analyze the quality, and thus reliability, of the COM's Progress Reports, a number of aspects have to be discovered which have either not been codified or publicized by the COM. First, the process of progress report creation remains largely unclear. So does the accession RoL concept employed by the Union. These aspects will be clarified through in-depth expert interviews with EU DG NEAR ENLARGEMENT officials and the revision of internal documents on the candidate countries' evaluations. Afterwards, Munck and Verkuilen's analytical framework will be used to assess the Reports' academic quality and thus, reliability. Furthermore, the consequences of the choice of methodology of the indices will be reviewed.

III. Democracy and Rule of Law Index Comparison

The indices used throughout this thesis will be compared based on the institutions that developed them and their supporters as well as their goals and methods, thus assessing their strengths and weaknesses and their potential effects on the evaluation outcomes. Apart from primary and secondary sources, expert interviews with COM, FH and BTI personnel will be used to provide a clearer picture of the evaluation Reports' creation processes.

III.1 Index Selection

The democracy evaluations that will be used to compare to the European Commission's Progress Reports have been created by Freedom House (Nations in Transit) and the Bertelsmann Foundation (Bertelsmann Transformation Index). Before explaining their methods and discussing the possible implications of their respective strengths and weaknesses, this part will justify their selection. Coppedge et.al. conclude that in order to measure X appropriately, one needs to "agree on what X is" (2011: 248). Consequently, when comparing democracy indices, their understanding of the concept is crucial. Since the thesis specifically focuses on the European RoL understanding throughout the accession process, indices must be selected that employ a similar methodology including a most-similar conceptualization of the RoL. Moreover, as the COM's Reports claim to do, they need to examine and evaluate de facto changes rather than mere policy developments in order to fulfill O'Donnell's "going beyond the regime" criterion (2001). Additionally, they must cover the time period in question, that is from the start of the negotiation talks with Croatia and Turkey in 2005 until the decision to accept Croatia into the EU in 2012. Most importantly though, if the existence of a political bias in the COM's Reports should be tested, their independence of the European Union must be granted and an own stake in the accession processes of Croatia and Turkey must be ruled out. As will be outlined in detail in the following sections, the general approach of all three indices is very similar at the drafting level. Professionals collect internal and external information on a country and rate its' performance according to their compliance with international RoL standards. Afterwards, vertical and horizontal harmonization procedures are introduced to make Reports more objective and comparable. As has been examined during the first step of the qualitative content analysis, there are a total of nine RoL-elements that are being evaluated in the Reports

from the COM, FH and BTI. Out of these nine, four can be found in at least 75% of all Reports analyzed in the empirical part (judicial independence, efficiency, crime, rights & liberties) which will be compared to one another.

However, a different bias may be inherently applicable: using the same underlying democracy and RoL concepts as well as basing the evaluations on the same sources may be a bias in assessing them itself.⁵¹ This bias could lead to similar evaluation results that are still not adequate in representing the actual quality of democracy or the RoL in the case studies. Coppedge et.al. (Ibid.: 252) highlighted that the high inter-correlation of many indices (Pearson's r around 0,86) can be explained with the coder's very similar general understanding of what constitutes democracy (Western liberal democracy perspective), a *halo effect* and an *extremes-bias*. The *halo effect* results in the projection of the general evaluation of a country onto the individual attributes. If a coder considers a country to be overall less democratic, chances are that each individual indicator and its respective attributes will be scored lower. If an analyst has a positive general perception of a country's democratic quality, he will be more inclined to rate other attributes higher. The *extremes-bias* shows that once perfect scores democracies like the USA, Canada or Sweden are being withdrawn from the calculation, Pearson's r drops significantly to about 0,63. Since policy makers are mostly interested in a precise level of evaluation of countries that are somewhere in-between the extremes of democracy and authoritarian regime, this becomes critical. Indeed, this similar-concept-bias is helpful in testing the key hypothesis. If it persists, the indices should produce similar evaluation results to the COM's Reports. In turn, should they differ in their findings, the conceptual bias can be ruled out which in turn hints to other explanatory factors and may strengthen the key hypothesis.

III.2 Index Backgrounds

“The characteristics of the judges can affect subjective ratings of political liberties and democratic rule. [...] The political orientation of a judge, the relation of the rated country to the judge's home country, the interests of the agency that is funding the ratings, or other

⁵¹ Bollen (1986) and Munck/Verkuilen (2002: 29) argue that these correlation tests can be disregarded because they merely test their reliability of the tests, not their actual reliability. However, if everyone operates with similar tools and material, consistency in results may be similar without reflecting actual developments. Also, since correlation tests have been carried out using highly aggregated data they “leave unresolved the critical issue of the potential multidimensional of the data” (Munck/Verkuilen, 2002: 29).

political, social, economic, and personal factors could affect a judge's ratings. [...] Even if judges were totally objective, a second factor could create method factors in their ratings. This is the incomplete information that is available to a judge."

(Bollen, 1993:1212)

Since democracy measuring indices are the judges of the level of maturity within a given country at a given time, there are two aspects that provide information on the quality of an index: its theoretical foundation and its methodology. Its theoretical foundation merely culminates in its conceptualization but is truly based on the institution creating it. The following background questions will shed light on potential index-inherent biases which have to be critically acknowledged when using them for this thesis: What kind of political actor is issuing it?⁵² How is it financed? What are its purposes? How influential is it within the international debate? Answering these questions may uncover potential biases. Secondly, the indices' methods will be described and evaluated using Munck and Verkuilen's requirements for a methodologically sound index. Their strengths and weaknesses will be highlighted and potential effects on the results of their comparison will be examined.

The European Commission

The European Commission creates yearly Progress Reports on all its candidate countries which address the Copenhagen Criteria in 33-35 Chapters which have to be negotiated and closed before accession. Chapter 23 analyzes the candidate country's progress with view to the Judiciary and Fundamental Rights. As has been explained in II.2.6, Chapter 23, addresses the following RoL elements:

- Judicial independence
- Judicial impartiality
- Judicial accountability
- Professionalism and competence
- Efficiency
- Anti-corruption

⁵² The questions were borrowed from Giannone (2010) who answered them for Freedom House and will be used as source forthwith in conjunction with other scholars who critically assessed said index.

The Reports are based on expert surveys and their goal is to mirror the de facto developments (rather than mere revisions of statutes) in a country in the given time frame within the respective analytical field. Ideologically, the European Union strives for a liberal democracy concept that embraces a market-based economy. The COM's Progress Reports are referenced by other countries' governments, international organizations such as the UN and the World Bank as well as think tanks (including FH and BTI) and NGOs.

Freedom House

Freedom House was founded in 1941 and is one of the oldest democracy measuring indices⁵³. The NGO has published its yearly report 'Freedom in the World' since 1973 and is also acknowledged for its 'Freedom in the Press' evaluations. It is founded by donations and receives partial endorsements from the US Agency for International development and the US Department of State. Since 80% of their funding comes from US state institutions, Giannone (2010) questions its status of a non-governmental organization. This impression is aggravated by the fact that its supporters not include private companies such as Google or Facebook which follow clearly profit-driven agendas. However, Freedom House is also supported by the Open Society Foundation and the Bradley Foundation. Freedom House is a value-driven organization that embraces the US Western liberal democracy concept. It was founded based on the perception that "American leadership in international affairs is essential to the causes of human rights and democracy" and that "the spread of democracy would be the best weapon against totalitarian ideologies"; a belief conceived as a genuine 'mission [...] to expand freedom around the world [and to] struggle against the other twentieth century totalitarian threat, Communism" (Giannone, 2010: 73). Chomsky and Herman concur that Freedom House "has long served as a virtual propaganda arm of the government and the international right wing interlocked with US government bodies such as Radio Free Europe and the CIA" (Ibid.). Although it claims to be bipartisan, its Board of Trustees includes a majority of members of the US neo-conservative spectrum including numerous people who have served under the G.W. Bush administration. The reason for its salience in the international debate are its "(presumed) neutrality, its capacity for political conditionality, and the international reach of its actions" (Ibid.). The Index is used

by national governments, the UN, the World Bank and multiple international aid organization as well as regularly quoted in mass media and even by scholars (although its methodological deficiencies are well understood among academics).

Bertelsmann Transformation Index

The BTI is a universal measuring tool of development and management progress within the transformation processes of 129 countries. The Bertelsmann Foundation publishes the individual country Reports every two years which serve as strategy papers for political consulting or the communication between state actors and other organizations including governments (Germany, UK, US), international organizations (UN, World Bank) and even other indices such as Transparency International and the Ibrahim Index of African Governance. The foundation itself is a pro-European, neo-liberal, market focused think tank that is financially independent of government actors because it receives its founding from the Bertelsmann Foundation. However, when looking at the Board of Trustees of the Bertelsmann Foundation, several conservative and liberal high-ranking European liberal politicians can be found such as the late Guido Westerwelle or Wolfgang Schäuble. Just like Freedom House, it uses expert-based evaluation methods to create its Reports. Table 9 below subsumes the background information on all three indices:

Criteria	European Commission yearly Progress Reports	Freedom House	Bertelsmann Foundation's Bertelsmann Transformation Index
Type of political actor	Supranational executive organ of the European Union	Conservative-led NGO ⁵⁴	German think tank
Financial set-up	Financed through EC funds	80% funding from US government agencies	Bertelsmann Foundation
Other supporters	No	Foundations (OSF, Bradley Foundation)	International think tanks financed and published

⁵⁴ Giannone (2010: 75) argues that due to the majority of its funding stemming from the US government as well as the involvement of high-level Republican politicians on its Board of Trustees, Freedom House is de facto not a non-governmental organization.

		Companies (Google, Facebook)	translation: Gulf Research Center, Russian Center for Post-Industrial Studies, Argentinian Centro para la Apertura y el Desarrollo de América Latina
Purpose	Assess fulfillment of Copenhagen Criteria	Spread democracy, focus on civil and political rights as well as negative freedoms from state intervention	Supporting reform process and principles of entrepreneurship ⁵⁵
Ideological background	Pro-EU Free market-based liberal democracy	Pro-US leadership Western Liberal (Human Rights, Democracy, Freedom) Neo-liberal orientation	Pro-EU Neo-liberal orientation (limited state interference)
Mode of operation	Expert surveys	Expert surveys	Expert surveys
Referenced by	UN, World Bank Governments Mass Media Other indices	UN, World Bank USAID, Millennium Challenge Account Mass Media (Washington Post)	World Bank, European Commission Governments (Germany, UK, US) Other indices

Table 9 - Overview of background information on democracy indices

⁵⁵

III.3 Adequacy of Indices' Methods and Implications for Across-Index Comparison

This part will outline each of the indices' methodology and test it using Munck and Verkuilen's catalogue of requirements. Thus, each index' strengths and weaknesses will be analyzed and its potential impact on the comparative analysis throughout the thesis assessed.

III.3.1 The European Commission's Progress Reports

While FH and BTI make their methodological evaluation approaches available to the public, neither the European Union nor the Commission has publicized any details of their *modus operandi*. Consequently, expert interviews were carried out with EU officials from the DG NEAR (Enlargement) in Brussels in February and March 2016 to shed light on the COM's evaluative process which will now be described. A number of internal evaluation guidelines were mentioned by the interviewees which were subsequently accessed. Their thesis-relevant findings are included into the **VII. Results** and **VIII. Discussion** Chapters while the complete documents are included into the **Annex**. All quotes in the subsequent section stem from said officials who asked to remain anonymous.⁵⁶

The Aim

As opposed to FH and BTI who assess the general political and economic development of their case studies, the COMs Progress Reports are "Acquis-focused". This means that they have been specifically designed to evaluate a country's progress in fulfilling the Acquis/the Copenhagen Criteria in order to make a recommendation to other EU institutions and the Member States regarding the country's accession process. This purpose affects the general tone of the Reports where "positive steps are mentioned first to encourage and congratulate the candidate countries" for their progress before emphasizing what remains to be done. This handling is in line with the EU's carrot and stick approach explained before.

The Process

The COM creates the Progress Reports in four stages before submitting their draft to the chain of command. During the first stage, the candidate country's Delegation creates a two-three page first draft based on their annual findings. Since the Delegation spends all year in the candidate

⁵⁶ Transcripts of the interviews can be accessed upon request and are part of the Annex. One interviewee has been in charge of Chapter 4 for Turkey and responsible for the horizontal comparability of structure and terminology among candidate countries. Another interviewee has been the author for the report on Serbia and is currently part of Directorate A as a Team Leader and Package Coordinator. He is responsible for creating the current methodology as well as harmonizing Reports among each other.

country and disposes of close contacts to its authorities as well as NGOs and civil society, it has the broadest and deepest understanding of the candidate's status quo and progress since the previous year. Mostly, information provided by the state's authorities correspond with those provided by independent sources but sometimes details or the depth of information differ (Interviewee 1). For example, in order to emphasize its progress, a candidate may focus include draft laws which cannot be included into the Reports or provide less detailed information on topics it does not want to be highlighted (Ibid.). By cross-checking with independent sources, the Delegation and Country Unit try to balance such information.

During the second stage, their draft is submitted to the Headquarters which use additional sources for fact-checking. Directorate A officials (Strategy and Policy Unit, ca. 10 people) collect the information and give further instruction on how to interpret them. Each official is responsible for different Chapters. While the Country Units have special insight only into one country they remain ignorant regarding the other candidate countries. Therefore, the Chapter Desk has been created to avoid evaluations being "too generous or too rigorous". By carrying out a horizontal comparison, the Chapter Desk aims to ensure that the same level of detail as well as language is used in all Reports. For each Chapter, one DG NEAR official checks for the same terminology, structure and assessment of progress from country to country.

During the third stage, an inter-service consultation is carried out which consists of an informal process where other Commission Services exchange the Draft and add their respective comments.

During the fourth stage, additional external sources including the Council of Europe as well as international and financial institutions (Worldbank/IMF), think tanks (including FH and BTI) and NGOs (e.g. Amnesty International, Transparency International) as well as information from candidate country's officials are being used to check whether the Commission's assessment correlates with those of other sources. However, the information is meant to be used as a "verification of the own results instead of their incorporation". Also, special meetings are held in Brussels or Strasbourg for additional input. Depending on the subject and level of advancement of the candidate country in question cooperative institutions or expert commissions are consulted. They are either provided with the first draft for comments or they provide their own, independent information. At this stage, additional information from the candidate country's authorities themselves – usually based on the structure of the Progress Reports – may be added.

The Decisions Chain

Once the draft is complete, it consists of 70-80 pages per candidate country that pass through the hierarchical chain from the Head of Unit to the Directorate-General to the Cabinet of Commissioners (each Commissioner is responsible for a specific topic/Chapter) to the College of Commissioners which together decide upon the final version of the Progress Report. Due to the length of the draft, the higher it goes up the chain of command the less it is being read and amended in detail. Consequently, the Head of Unit which knows his country most intimately is most influential in the final assessments being made and language to be used within the report. His corrections are based on the aim to create a “coherent general picture” of the progress made by his candidate. The Directorate General is politically responsible for harmonizing the assessments among all countries and may therefore submit further comments. Linguistic changes made by the Directorate General are aimed at presenting a realistic view of the candidates’ progress as well as their comparativeness. Further comments can be made by the Commissioners regarding their specific topic which may result in slight linguistic changes but usually do not affect the overall assessment. Last, the College of Commissioners decides upon the final versions of the Reports. Both interviewees emphasized that the final Reports are very similar to their drafts and usually only encounter slight linguistic changes after going through the harmonization process. Shortly before publishing them, a meeting takes place between the Director and the candidate country’s Ambassador to prepare the candidate for the main results. However, the candidates are not presented with the full report, may only comment on the general outcome and have no power to request any changes.

Methodology

The creation of the Progress Reports is outlined in the yearly guidance note in the Enlargement package to the DG Enlargement (now DG NEAR). In principle the process has been consistent during the time of analysis (2005-2012) up until 2014. Since 2015, a new methodology has been introduced in order to “continue enhancing credibility” and introduce incentives to countries “to continue their reform efforts” in lieu of “no accession in sight”. The increased comparability ensured through the new process is aimed at “increasing the appeal of the package and stimulate “competition between the countries” (European Commission, Enlargement package 2015, Guidance Note). This section will first introduce the process and methodology of the progress report creation process applicable to the time of

analysis and subsequently shortly outline the changes made since 2015. The latter is included as a base for the policy recommendations introduced in the **IX. Summary** Chapter.

Progress Report Creation 2005-2014

Responsible for the drafting of the Reports are the DG ELARG country teams in cooperation with the respective Chapter Desks and DGs. The overall co-ordination of the drafting and organizational procedures is carried out by the DG ELARG's Strategy Co-ordination Unit (unit A.1). Each country team appoints a contact person at working level for the entire process as well as authors for each part of the report.

The actual creation of the process occurs in six stages. First, using all available relevant sources (contributions from the candidate countries, the Delegations, the various Commission services, Reports from peer reviews, technical consultations and expert assessments, Reports from EP and Member States, IMF, Worldbank, OECD, EBRD, EIB etc.) the country teams prepare the first draft of the Reports while consulting informally with the Chapter Desks. Secondly, the Chapter Desks comment on the relevant parts during cross-sectional readings, thus ensuring coherent assessments between the countries. Third, during the horizontal readings the coordination unit helps the country teams complementing the received information as well as conducting sectorial meetings for any potential outstanding issues. Fourth, the coordination unit and the Chapter Desks go through the results of the inter-service consultation and complement the findings with relevant sectorial and horizontal indications. Fifth, the texts are revised by the country teams and new information may be added up until the cut-off date of September 30. Finally, another horizontal reading is carried out by the coordination unit before transmitting the draft. Afterwards, the draft goes through a chain of reviews and corrections on higher (political) levels as described during the expert interviews in VI.1.1 and in the Annex of the 2006 Guidance Note (p. 17-20, cf. Appendix).

Roles of EU and non-EU actors

The DG Enlargement cooperates closely with the other DGs who comment on their first drafts by including input from all services whose competence is helpful as well as Eurostat for the preparation of statistical data and the translation service. The Commission's Delegations play a key role in the assessment of progress made with view to the political criteria as well as administrative capacity of a candidate country. After the finalization of the the first drafts the Heads of the Delegations are invited to Brussels for reading and commenting by the coordination unit. While the Member States play no role in the preparation of the Reports,

they may proactively and informally offer input. The European Parliament are not privy to the creation, however their Reports and meetings of Joint Parliamentary Committees may be drawn upon as a source and in that case explicitly mentioned. While candidate countries are invited to contribute to the Reports their input must be double-checked to verify information. During a meeting in Brussels prior to Report finalization, they are presented with the preliminary findings without accessing the actual report in order to give them the opportunity to provide final updates if applicable. The Guidance Note clearly states that “There is no question of negotiating with the candidate countries the findings of the Reports, which are Commission documents.” (European Commission, Enlargement package 2005, Guidance Note). Other organizations involved include the Council of Europe, the OSCE, relevant NGOs and financial institutions which are asked specific questions previously prepared by the country teams. As the Guidance Note emphasizes, the use of “different **independent sources** and **cross-checking** of information are of great importance”. While no specific other sources are mentioned, the expert interviews confirmed that other indices are consulted, including Freedom House and possibly Bertelsmann. In the Enlargement Package Guidance Note from 2015, it is specifically included that “Indices and assessments by other organizations...can be used to feed into our assessment” and will be added “in an overview of key indices for the enlargement countries in a new annex to the Strategy paper” (European Commission, Enlargement package 2015, Guidance Note).

Content of assessment – Rule of Law

An informal working document of the Commission, the Guide to the Main Administrative Structures required for implementing the Acquis (2005, included in the Appendix), addresses the goals for Chapter 23, specifically the Judiciary and Anti-corruption policies (as well as fundamental and citizen’s rights). For the Judiciary “The establishment of an independent, reliable and efficient judiciary is of paramount importance.”

With view to anti-corruption policy, “The JHA Council’s approval (14 April 2005) of the Commission’s Communication on a Comprehensive EU Policy against Corruption (Brussels, 28.5.2003 COM(2003) 317 final) encompassed 10 Principles for Improving the Fight against Corruption in Acceding, Candidate and Other Third Countries, set out in the Annex to that Communication.” The principles infringed in said document are summarized in Table 10 below:

Judiciary
<ul style="list-style-type: none"> • Sufficient human resources and qualified staff
<ul style="list-style-type: none"> • Adequate and modern equipment, acceleration of court proceedings
<ul style="list-style-type: none"> • Reduction of the number of pending cases so as to avoid unreasonable delays
<ul style="list-style-type: none"> • Measures to ensure the enforcement of judgements
<ul style="list-style-type: none"> • Procedures to ensure ethical conduct by the judiciary and the effective access to justice
Anti-corruption principles
<ul style="list-style-type: none"> • The implementation of national anti-corruption strategies or programmes, covering both preventive and repressive measures
<ul style="list-style-type: none"> • The ratification of „all main international anti-corruption instruments they are party to (UN, Council of Europe and OECD Conventions)
<ul style="list-style-type: none"> • The implementation of competent and visible anti-corruption bodies (i.e. well trained and specialised services such as anti-corruption prosecutors)
<ul style="list-style-type: none"> • Objective and merit-based criteria for access to public office
<ul style="list-style-type: none"> • Quality management tools as well as auditing and monitoring standards to grant integrity, accountability and transparency in public administration (judiciary; police, customs, tax administration, health sector, public procurement)
<ul style="list-style-type: none"> • Codes of conduct in the public sector
<ul style="list-style-type: none"> • Whistle blowing rules
<ul style="list-style-type: none"> • Public awareness-raising campaigns in the media and training
<ul style="list-style-type: none"> • Clear and transparent rules on party financing as well as control <i>mechanisms</i>
<ul style="list-style-type: none"> • “White lists” for integer companies <i>as incentives</i> against corruption

Table 10 - Goals of Judiciary Reform (Guide to the Main Administrative Structures required for implementing the Acquis, 2005)

In the annex of the Guidance Note, the country teams are presented with a skeleton structure of the Reports outlining what aspects to include into which section of the 33 Chapters. In the 2005 Guidance Note, the sections relevant to this analysis refer to the Judicial system and Anti-corruption policies and include the following aspects to be evaluated by the analysts:

Judicial system
<p>Structure:</p> <ul style="list-style-type: none"> • Possible changes <p>Functioning:</p> <ul style="list-style-type: none"> • Assessment of judicial system • Including judicial capacity <ul style="list-style-type: none"> → Legal basis of judicial system → Independence and impartiality → Transparency and openness → Status and remuneration of judges → Number of judges → Number of judicial vacancies → Length of judicial proceedings → Training-including on good judicial practices → Human rights law, and European law → Equipment
Anti-corruption policy
<p>Indication of level of corruption in the country, linked to surveys and / or public perception.</p> <ul style="list-style-type: none"> • Existence / implementation of anti-corruption policy/programme/legislation • Institutional set-up in fight against corruption, division of tasks between various bodies involved • Existence and implementation of code of conduct for civil service / public administration • Relevant training for staff in public institutions – Cases of corruption in administration; reaction to these by relevant authorities – Ratification / implementation of main international conventions in the field – Involvement in Council of Europe Group of States Against Corruption (GRECO)

Table 11 - European Commission (Enlargement package 2005, Guidance Note, Annex)

In the Guidance Notes of 2006 and 2007 no specifications for RoL-relevant sections have been made and between 2008 and 2012, Chapter 23 on the Judiciary and Fundamental Rights merely mentions the four sub-fields that constitute the RoL evaluation: judiciary anti-corruption policy, fundamental rights, and EU citizens' rights. Thus, the critique offered by academics regarding the lack of clarity of criteria can partially be retraced. Although initially a more comprehensive list of criteria has been introduced, no concrete instructions are provided as to how to evaluate them.

Linguistic Description and Assessment Rules

Within the Guidance Notes from 2005-2012 a number of linguistic requirements are made of the authors to ensure neutrality and comparability of the Reports. As these rules distinguish them from other Reports and may help explain some of the critique targeted at the Reports, they will be introduced selectively.

First, aspects to be measured and evaluated are limited to “**decisions actually taken**, legislation actually adopted, measures actually implemented, and structures actually in place and functioning. No reference should therefore be made in the Reports to draft legislation which is under preparation, or measures which are planned to be taken.” (Ibid., all emphasis in original). Secondly, authors are asked to “Make sure that **issues which were criticised** in the previous Report are explicitly followed up” in the following report. Whereas instructions or “pre-judging the future” are prohibited, under the changed methodology since 2015 “guidance on reforms” in key policy areas is explicitly recommended. Regarding linguistic intricacies, special attention is given to cases where no progress could be found. In that case, “it will be appropriate to indicate that ‘No new (or particular) developments are to be reported in the area of’ whereas “the phrase ‘no progress had been made over the past year’ should be reserved for cases where no progress whatsoever has been registered, and where no action is planned for the future either, as it has a negative connotation.” (Ibid.). If “**nothing much has changed**” the note specifies to “not repeat *verbatim* the same assessment as in the previous Report” but to instead “use wording such as ‘continues to’ (all emphasis in original, Ibid.). Slight changes and additions have been included in the following Guidance Notes. For example, the 2006 Guidance Note includes the stipulation that in each area each paragraph has to begin with the positive developments (“progress made”) before outlining the “negative developments and/or shortcomings” and finishing off with a “sentence describing the overall level of preparedness”. Not knowing this procedure may lead to misinterpretation of the results as the former part is focused on progress made since the last years while the latter evaluates the level of compliance with the Acquis. Further inclusions into the 2006 Guidance Note are the prescribed omission of words to start sentences with such as “*Despite, even though, although*” as well as expressions like “*seem, appear, expect, should, must, important*” (European Commission, Enlargement package 2006, Guidance Note, emphasis in original). The 2007 Guidance Note stresses that the Commission needs “to be able to defend and explain every part of the Reports. So all information should be checked/ retrieved, and

sentences such as ‘according to source X’ are not allowed” (European Commission, Enlargement package 2007, Guidance Note). Furthermore, no individual cases should be mentioned nor the names of persons. Also, “all elements of comparison with other candidate countries” should be avoided. The same is true for “subjective assessment language (such as ‘surprisingly’, ‘remarkable’)” (Ibid.) in order to not affect the neutral tone of the Reports. Moreover, special advice is given for the evaluation of Croatia and Turkey. “Particular attention should be given to areas where benchmarks for opening or closing chapters have been defined: describe developments which are relevant for meeting the benchmark, adding ‘which is a key for the accession negotiations in this chapter’” (Ibid., emphasis in original). While no additional rules were introduced in the Guidance Notes from 2008-2010, the 2011 one stresses the RoL-relevant prescription that “Generally, no individual judicial cases should be mentioned” and that names of persons should be avoided, too (European Commission, Enlargement package 2011, Guidance Note). This is of particular importance since Croatia was repeatedly evaluated based on its dealing with war criminals while Turkey has been assessed based on the way it treated members of opposition and media. Both instances have been expressed exemplarily in the other indices. Thus, the perception could arise that the COM’s Reports ignore these instances whereas they might just be “linguistically hidden” based on the prescriptions introduced in the Guidance Note. However, the omission of such details may simply be attributed to the limited space granted for every single chapter reviewed in the Reports. The Guidance Note furthermore states that special focus should be put on “the relation between the political criteria and Chapter 23, where the more factual and technical elements should be put in Chapter 23.” Whereas, “as usual, the analysis and assessment of the developments covered by the political criteria should be dealt with in the political chapter” (Ibid., emphasis in original). As the Guidance Note of 2012 states, the Reports should not exceed 55 pages in total except for Turkey which is granted 75 pages “which allows for a more detailed assessment of the progress with regard to political criteria” (European Commission, Enlargement package 2012, Guidance Note). This special treatment could in itself give rise to political bias perceptions. The more detail a report entails, the more attention will be drawn to deficits whereas a less comprehensive report may apply fewer space and scrutiny so the same aspects.

Progress Report Creation since 2015

In 2015 a “strengthened reporting methodology” was introduced as part of the Enlargement package. According to the expert interviews, the change of the methodology was the response to the external critique of lacking transparency and comparability. It also took into account the new enlargement environment which explicitly stated non-accession within the next years. The Guidance Note outlines that it also serves as a tool to incentivize the candidate countries. “With no accession in sight, enlargement countries need incentives to continue their reform efforts. More **comparability** between countries could increase the appeal of the package and stimulate ‘competition’ between the countries” (European Commission, Enlargement package 2015, Guidance Note, emphasis in original). Instead of focusing only on progress as has been done the previous Reports, there was a shift to measuring ‘readiness’ for accession. Specifically, performance started to be monitored by measuring the effectiveness of COM’ policies including IPA. In line with the shifted focus, key political priority areas were defined which are evaluated through specific checklists which in turn are measured through concrete assessment scales. Those areas include the RoL “(judiciary, corruption and organised crime, both in the political criteria and Chapter 23), freedom of expression, economic criteria and public administration. This will be complemented in the Acquis section by application to a limited number of pilot chapters (public procurement, statistics and financial control)” (Ibid.).

The checklist for the RoL includes the following criteria (indiscriminate selection of indicators from the internal reporting guidelines for the Commission on the Justice reform, based on the EU Justice Scoreboard, for more detailed indicators see Appendix):

STRATEGIC DOCUMENTS AND BUDGET
Has the country a comprehensive judicial reform strategy/action plan resulting from a wide consultation in place?
Is it effectively addressing the main shortcomings? Has it realistic timelines? Has it a logical sequencing of actions? Has it responsible bodies/persons that are clearly defined? Is there a budget for its implementation (ideally individual actions are budgeted)? Is there a monitoring mechanism in place allowing for a mid-term review and leading to appropriate corrective actions when needed?
MANAGEMENT BODIES/ JUDICIAL/PROSECUTORIAL COUNCIL
Is there a Judicial/prosecutorial Council, independent from government and administration, responsible for managing the justice system incl the appointment, promotion and career of Judges/prosecutors?
INDEPENDENCE AND IMPARTIALITY
Are the principles of judicial independence are set out in the Constitution and reflected in internal law?

Do judges enjoy both external and internal independence when deciding an individual case? Do prosecutors enjoy a certain level of autonomy when working on an individual case?
ACCOUNTABILITY
Is there a code of ethics for judges, prosecutors and lawyers or defined standards of conduct? May a breach lead to the removal from office or disciplinary sanctions? Is there a mechanism in place to monitor compliance with the Codes?
Is integrity training part of the curriculum for initial training? Is integrity a criteria in the initial selection and nomination process, including for senior positions?
PROFESSIONALISM AND COMPETENCE
Is the system for the recruitment, selection, appointment, transfer and dismissal of judges, and prosecutors independent of political influence? Is the entry in the judiciary based on transparent, merit-based and objective criteria, fair in selection procedures, open to all suitably qualified candidates (national competition) and transparent in terms of public scrutiny.
QUALITY OF JUSTICE
i.e. Training: What is the actual percentage of judges participating in <ul style="list-style-type: none"> • Continuous training in EU law • In-service training for the use of computer facilities in courts • In-service training for management functions in the courts • In-service training for specialized judicial functions • General in-service training • Initial training
i. e. Budget and Resources <ul style="list-style-type: none"> • Budget for courts (in EUR per inhabitant) • Budget for prosecution offices (in EUR per inhabitant) • General government expenditure on law courts as a percentage of GDP • Number of judges (per 100.000 inhabitants) – only full time judges • Number of Prosecutors (per 100.000 inhabitants)
EFFICIENCY
<ul style="list-style-type: none"> • Length of proceedings (=the time (in days) needed to resolve a case in court, that is the time taken by the court to reach a decision at first instance) in civil/commercial cases, administrative cases and penal cases. • Clearance rate (=the ratio of the number of resolved cases over the number of incoming cases in a year)

Table 12 - Selection of Justice Reform Reporting Guideline Questions (Internal COM Document, 2015)

Similar reporting guidelines including specific indicators have been issued by the Commission in 2015 on Chapters 5, 18, 32 as well as on corruption, freedom of expression,

organized crime and public administration reform. Since the latter four documents are relevant to the RoL, they will be attached in the Appendix. Given that they are not relevant for the time of analysis within the thesis, they will not be examined any further.

The assessment scales used to evaluate the indicators listed above comprise of an “Early stage’, ‘some level of preparation’, ‘moderately prepared’, ‘good level of preparation’ and ‘well advanced’”. While a detailed methodology of translating a certain development into one of these stages is currently developed by the Headquarters the Delegations are asked in the meanwhile to provide their own assessments. Developments in the remaining areas will be evaluated based on their progress with a similar scale comprising of “‘Backsliding’, ‘no progress’, ‘some progress’, ‘good progress’ and ‘very good progress’”. While “limited”, “some” or “good progress” have been used in previous Reports to qualify the developments, the explicitly negatively connoted newly introduced linguistic evaluations “backsliding” and “no progress” can be viewed as a response to the accusations by Member States and the MEPs that the COM’s Reports are too positivistic and not critical enough candidate’s shortcomings. Furthermore, as opposed to the previous Reports, reform guidance is explicitly advised for the key policy areas asking the authors to include a maximum of three key issues or reforms that should be prioritized by the candidate within the following year. While the previous Reports included facts and closed proceedings only, the new methodology allows for the inclusion of interim steps (“e.g. adoption of a draft law by the government, but law not yet adopted by the parliament”) (Ibid.) if they are of particular relevance. Also, the consultation of indices and assessment by other organizations is explicitly allowed giving the example of Transparency International’s Corruption Perception Index. Moreover, an overview of key indices will become part of the Annex to the Strategy paper “which will allow for further comparisons between the countries, without the Commission approving or taking over other organisations’ methodologies.” (Ibid.)

Strengths and weaknesses

Based on Munck & Verkuilen’s index evaluation system, the COM Progress Reports should be completely unsuitable to provide an academically sustainable (as in: reliable and reproducible) account of the candidate’s development. First of all, there is no clear definition what the RoL means to the COM and using the Acquis with the large amount of information creates a concept so stretched that it is unfit to make clear evaluations. Particularly the political criteria checked in the Reports (which the RoL is part of) are considered “more sensitive” wherefore

they are more closely scrutinized by members of the chain of command who have not drafted the Reports but carry political responsibility. Not only are they “more difficult to assess” because sometimes there exist “no written rules on how to assess” issues such as corruption (Ibid.). Judgment on political criteria is considered “harder than on technical questions” and may not be “as factual because of political consideration behind” them (Ibid.) Moreover, the pick-and-choose method employed until 2015 to take account of the individuality of the candidates is in itself the reason for the peerlessness of the Reports. Also, regarding the implementation of laws, the combination of multiple information as well as the evaluation of ‘track records’ entails judgmental aspects. For example, if within the anti-corruption field more arrests have been made this could be attributed to more documentation on the topic, a more efficient police work or, on the contrary, to a worsening of the problem. Which facts are combined and what conclusions are drawn thereof leaves significant room for interpretation which makes the COM’s assessment method difficult to reproduce.

Furthermore, the lack of a specific RoL concept or checklist opens the door for any claims of politically biased Progress Reports. However, according to a DG NEAR staffer, the bigger methodological problem causing perceptions of unjust evaluations is the strong focus on progress and remaining deficiencies. How much progress a country can make depends significantly on where it stands within the accession process. At the beginning, it is much easier to show much progress by adopting laws and creating institutions. The further the accession process continues, the more challenging it becomes to fulfill the Copenhagen Criteria because laws then have to be implemented and respected and institutions have to be made functional and sustainable. This requires more attention, conviction and resources than the initial steps. This leads to a distorted picture of ‘weaker states’ that are apparently being evaluated better than ‘stronger states’ and vice versa. Moreover, the final sentence of the report which is not progress-focused but supposed to represent the country’s “state-of-play” is not substantiated by factual information which decreases its transparency and credibility and adds to the perception of random and/or politically biased assessments (DG NEAR official).

While the RoL aspects within the Reports are repetitive throughout the years and among the case studies at hand, there is no clear or codified organization of the attributes let alone a specific level of abstraction. The same is true for the selection of indicators which is not justified anywhere. While measurement levels do exist (language regime), their lack of

specific anchoring examples as well as the fact that those codification guidelines are not publicly available make the replicability of the process and/or the results impossible. In fact, all that is publicized about the creation of the Reports are the general sources which include replies to questionnaires, bilateral meetings and Agreements with the candidate countries, assessments made by Member States and candidate countries themselves, European Parliament's Reports and resolutions as well as the work of various international organizations, non-governmental organizations and other bodies (European Commission, 2000).⁵⁷ Since there is no quantitative assessment of the qualitative analysis, Munck and Verkuilen's aggregation criteria do not apply to the COM Progress Reports.

Nevertheless, two aspects speak for the quality of the Reports. First, no other index possesses even remotely the range of resources with view to time, money, people and close networks to officials of the country to be evaluated as does the Commission. Having so many valuable information at one's disposal highly increases the chance of adequately understanding and depicting a country's status quo and development. Additionally, although the COM does not publish its specific method, it does have internal guidelines and a clear-cut structure of coordination among its officials who create the Reports as well as a system which – at least since 2015 – increasingly aims at creating reproducible and transparent codifying mechanisms.

Moreover, it is necessary to emphasize that the COM has never attempted nor claimed that the Progress Reports can hold up to academic standards because they are ultimately a tool for political decision-making.

III.3.2 Freedom House's Freedom in the World Reports

Although Freedom House does not claim to measure democracy but freedom, the elements it uses are congruent with the key elements tested in other democracy indices. According to Freedom House, freedom is considered to be “an individual requirement (‘freedom as experienced by individuals’) and is defined as ‘the opportunity to act spontaneously in a variety

⁵⁷ EU Commission, COM-Document, Bulletin of the European Union Supplement 5/97 AGENDA 2000 For a stronger and wider Union Document drawn up on the basis of COM(97) 2000 final 15 July

of fields outside the control of the government and other centers of potential domination” (Giannone, 2010: 76). The checklist that is based on this concept include the central values of liberty, civil and political rights, freedom from government interventions, and individual protection mainly in the market sphere. According to Giannone, this happens to the “detriment of socio-economic rights” and leads to the “gradual vanishing or lack of relevance of the value of quality” (Ibid.: 78) and therefore constitutes a neo-liberal worldview. Much to the dismay of scholars of methodology, the index’ founder Gastil asserted that the indexes’ “categories are developed not so much out of any theoretical understanding of democracy as from the experience that these were headings under which information relevant to the rating system has most often been available” (Ibid.: 77). However, since policy makers use the index to evaluate the level of democracy in their target states and scholars have examined it in comparison with other democracy indices, it will be treated as such within this thesis as well.

The Aim

The goal of the Reports is to measure the degree of freedom in terms of *political rights* and *civil liberties* of as many countries of the world as possible.

The Method and Rule of Law Conceptualization

In its ‘Freedom in the World’ report, states are being scored on a scale from 1 (most free) to 7 (least free) by country experts. “The report’s methodology is derived in large measure from the Universal Declaration of Human Rights, adopted by the UN General Assembly in 1948. Freedom in the World is based on the premise that these standards apply to all countries and territories, irrespective of geographical location, ethnic or religious composition, or level of economic development” (Ibid.). States with a calculated score between 1,0 and 2,5 are considered “free”, states ranging between 3,0-5,0 as “partially free” and states scoring 5,5-7 points as “unfree”. As part of the *political rights*, there are three categories (A. electoral process, B. parliamentary pluralism and participation, C. functioning of government, plus additional discretionary questions) with 10 corresponding questions which have to be scored a certain number of points by the country experts. Each score corresponds to a rating between 1 and 7. With view to *civil liberties*, the categories D. Freedom of expression and belief, E. Association and organizational rights, F. the Rule of law, and G. Personal autonomy and individual rights are examined by answering 15 individual questions (Freedom House, 2014).

The RoL categories consists of four sub-categorical questions as outlined in Table 13 below and stem from FH’s 2014 Freedom in the World Methodology section. The highlighted questions correlate with the RoL criteria addressed within the COM’s Progress Reports.

Criterion & Sub-questions
Is there an independent judiciary?
Is the judiciary subject to interference from the executive branch of government or from other political, economic, or religious influences?
Are judges appointed and dismissed in a fair and unbiased manner?
Do judges rule fairly and impartially, or do they commonly render verdicts that favor the government or particular interests, whether in return for bribes or other reasons?
Do executive, legislative, and other governmental authorities comply with judicial decisions, and are these decisions effectively enforced?
Do powerful private concerns comply with judicial decisions, and are decisions that run counter to the interests of powerful actors effectively enforced?
Does the rule of law prevail in civil and criminal matters? Are police under direct civilian control?
Are defendants’ rights, including the presumption of innocence until proven guilty, protected?
Are detainees provided access to independent, competent legal counsel?
Are defendants given a fair, public, and timely hearing by a competent, independent, and impartial tribunal?
Are prosecutors independent of political control and influence?
Are prosecutors independent of powerful private interests, whether legal or illegal?
Is there effective and democratic civilian state control of law enforcement officials through the judicial, legislative, and executive branches?
Are law enforcement officials free from the influence of non-state actors, including organized crime, powerful commercial interests, or other groups
Is there protection from political terror, unjustified imprisonment, exile, or torture, whether by groups that support or oppose the system? Is there freedom from war and insurgencies?
Do law enforcement officials make arbitrary arrests and detentions without warrants or fabricate or plant evidence on suspects?
Do law enforcement officials beat detainees during arrest and interrogation or use excessive force

or torture to extract confessions?
Are conditions in pretrial facilities and prisons humane and respectful of the human dignity of inmates?
Do citizens have the means of effective petition and redress when their rights are violated by state authorities?
Is violent crime either against specific groups or within the general population widespread?
Is the population subjected to physical harm, forced removal, or other acts of violence or terror due to civil conflict or war?
Do laws, policies, and practices guarantee equal treatment of various segments of the population?
Are members of various distinct groups—including ethnic and religious minorities, LGBT people, and the disabled—able to exercise effectively their human rights with full equality before the law?
Is violence against such groups widespread, and if so, are perpetrators brought to justice?
Do members of such groups face legal and/or de facto discrimination in areas including employment, education, and housing because of their identification with a particular group?
Do women enjoy full equality in law and in practice as compared to men?
Do noncitizens—including migrant workers and noncitizen immigrants—enjoy basic internationally recognized human rights, including the right not to be subjected to torture or other forms of ill-treatment, the right to due process of law, and the rights of freedom of association, expression, and religion?
Do the country’s laws provide for the granting of asylum or refugee status in accordance with the 1951 UN Convention Relating to the Status of Refugees, its 1967 Protocol, and other regional treaties regarding refugees? Has the government established a system for providing protection to refugees, including against refoulement (the return of persons to a country where there is reason to believe they fear persecution)?

Table 13 – FH’s ‘Freedom in the World’ Rule of Law Questions

In order to grant continuity and reliability, a country’s score is checked against its previous year’s rating. Only specific events that significantly alter the quality of freedom in either of the categories lead to a score change. However, trend arrows can signify whether a country has a positive or negative development compared to the previous evaluation.

In 2011, 59 analysts and 20 senior-level academic advisors were engaged in the process. The sources the analysts used to make their assessments are foreign and domestic news Reports,

academic analyses, nongovernmental organizations, think tanks, individual professional contacts, and visits to the region (Ibid.).

Strengths and weaknesses

Freedom House's methodology has undergone substantial scrutiny by numerous scholars which led to a common evaluation: the methodological inadequacy of the index.⁵⁸ Before examining the specifics, the index' advantages will be highlighted. The strengths of Freedom House are the yearly evaluations which grant a high currentness of data, the wide range of countries assessed as well as its transcendence of institutions and focus on de facto developments within a country (Pickel, 2006: 2019). Furthermore, the usage of country experts as well as the inclusion of secondary sources (e.g. NGO Reports) is advantageous (Müller/Pickel, 2007: 532). However, critique is offered regarding every aspect of Munck/Verkuilen's catalogue.⁵⁹ By being a tool "for measuring democracy, good governance, and human rights" the concept is overstretched and therefore maximalist (Landman/Häusermann, 2001). Moreover, the concepts that are provided are clearly biased toward a free-market liberal democracy (Scoble/Wiseberg, 1981). According to Giannone (2010), "Bollen underlined their Cold War and pro-market biases, while Bollen and Paxton emphasized that the index of FH systematically favors Christian and Western countries, and tends to adversely codify Muslim and Marxist-Leninist countries". Furthermore, "Mainwaring et al. pointed out that FH measurements 'contain two systematic biases: scores for leftist governments were tainted by political considerations, and changes in scores are sometimes driven by changes in their criteria rather than changes in real conditions'" (Ibid.). With view to the adequacy of measurement and aggregation numerous aspect are being criticized. Each item of the indicators is assigned a score while it remains unclear in which case what score is justified thus impairing a reproducibility of the process. The same applies to the threshold between "free", "partly free", and "unfree" which according to Diamond (1999) leads to arbitrariness and a lack of accuracy for very different cases. There is no clear set of coding rules, the different indicators are simply given the same weight and therefore added without proper theoretical justification. While the selected aggregation rule is clear and explicit it is still deficient. "Scores for the two attributes— political rights and civil rights—are generated by adding up the scores assigned to each of its respective components" (Munck/Verkuilen,

⁵⁸ Giannone (2010), Landman/Häusermann (2003), Diamond (1999), Scoble/Wiseberg (1981), Bollen/Paxton (2000), Mainwaring (2001), Munck/Verkuilen (2002)

⁵⁹ For a detailed account see Munck/Verkuilen, 2002, pp. 20-25, 70-76

2002: 25). Its deficiency lies in the lack of a theoretical foundation for its long list of components (Ryan, 1994: 10), the unaccounted for equal weighing of each attribute as well as the lack of reliability based on the fact that none of the crucial information for reproducibility have been made public. As summarized by Munck/Verkuilen, “because no set of coding rules is provided, and the sources of information are not identified with enough precision, and because disaggregated data have not been made available to independent scholars, ‘the aggregate data offered by Freedom House has to be accepted largely on faith’” (In: Giannone, 2010: 69). Nevertheless, in practice, faith is being put into Freedom House’s results by multiple notable policy makers and aid donors, including the European Commission who benchmarks its results against FH’s.

III.3.3 Bertelsmann Transformation Index

The Bertelsmann Foundation is a think tank that aims to promote reform processes in line with the principles of entrepreneurial thinking. It created the Bertelsmann Transformation Index (BTI) to generate key learnings and best practices from transformation processes in countries all over the world to be applied to newly developing nations. The BTI is published every two years. Just like Freedom House, the BTI has an underlying neo-liberal concept of democracy although it includes the notion of social justice (“democracy under the rule of law and a market economy anchored in principles of social justice represent goals”) and considers transformation to be a “comprehensive and politically driven change in which an authoritarian system and a state-dominated or clientelist economic order evolve in the direction of democracy and a market-based economy” (BTI Manual: 123). The BTI openly disclaims its normative positioning but argues that its ideological bias does not steer it towards a “particular existing institutional model such as the German model of social market economy or specifically European models of constitutional democracy” (Ibid.: 127).

The Aim

The BTI’s goal is to measure the level of democracy and market economy in countries that are not yet consolidated democracies.⁶⁰ Thus, the BTI has a smaller range of case studies than FH’s Freedom in the World (129 vs. 195 countries), yet a much bigger one than the COM

⁶⁰ The BTI makers acknowledge the absence of a clear threshold of consolidation wherefore they exclude all countries that were members of the OECD in 1989 from their analysis as well as most states with fewer than two million citizens.

which only focuses on accession candidates (currently seven). However, since all of the EU's candidate countries are transitional states, they are included in both in FH's and the BTI evaluation.

The Process

Bertelsmann divides the world into 7 regions. Each region consists of 18-22 countries and is headed by a regional coordinator who chooses each country's two country experts who author each respective report. The BTI employs approximately 250 experts from "leading universities and research institutions around the world" (Ibid.: 128) whose list is published online. Those include experts from civil society organizations which have been chosen at the recommendation of local coordinators based on their perceived "level of professional expertise, considerations of independence and impartiality" (Ibid.). The country experts (one foreign and one local expert, usually one politically scientist and one economist) use a standardized codebook to assess altogether 17 criteria (indicators) through 49 individual questions for their country. Country expert 1 provides a numerical score for each question together with a written explanation including empirical evidence. As a benchmark for the experts' evaluation, the codebook includes exemplary answers which are linked to specific ranges of scores. The regional coordinator checks for completeness and forward the draft to the second country expert who – independently from the first – makes his numerical assessment and then factually checks and comments on the first draft. The regional coordinator then harmonizes both scores and texts for each country and among all countries of his region taking into account the intra-regional variance and scope. The preliminary scores are then taken to the regional coordinators' meeting which lasts three days and is aimed at harmonizing the scores across the regions as well as noting which texts have to be amended based on the corrected scores. The new scores are then submitted to the BTI-Board, a steering committee consisting of scientists with different country expertise who – in dialogue with the regional coordinators – determine the final scores and texts. The aggregated data are eventually comprised into two indices: the Status Index (focusing on the state of political transformation) and the Management Index (focusing on the state of economic transformation).

Since 2016, an additional monitoring tool has been introduced. Students at the University of Heidelberg check the individual scores across all seven regions in order to attenuate the structural weakness that one regional coordinator is ignorant of the others' scoring practice.

Furthermore, the codebook changes slightly every two years based on issues that have been identified in the previous version.

The Method and Rule of Law Conceptualization

The BTI rests on three pillars or “branches”: Political transformation, Economic Transformation, and Transformation management. The political transformation is measured within five criteria (indicators: Stateness, Political participation, Rule of law, Stability of democratic institutions, Political and social integration) under which 18 questions are subsumed. The state of economic transformation comprises seven criteria (Level of socioeconomic development, Organization of the market and competition, Currency and price stability, Private property, Welfare regime, Economic performance, Sustainability). Last, the government’s management performance is comprised into five criteria (Level of difficulty, Steering capability, Resource efficiency, Consensus-building, International cooperation).

Regarding index aggregation, “the Status Index is formed by calculating the average of the total scores given for the dimensions of political (democracy status) and economic (market economy status) transformation. The state of transformation in each analytic dimension is equivalent to the average of the scores of the associated criteria. Criterion scores are in turn based on the average scores of the equally weighted indicators that comprise the criterion.” Ibid.: 130). The BTI combines both analytical dimensions according to its “normative premise...under which transformation is always conceived as a comprehensive transition toward democracy and a market-economic system”.

The Management Index is formed by calculating the average of scores given for the management criteria, which is then offset against the assigned level of difficulty. Relevant thresholds are assigned by the BTI that – even if falling short in one relevant threshold – result in the country being classified as an autocracy (Ibid.: 131).⁶¹ Out of all other indices, the BTI is the only one that highlights governance performances as an individual feature. The BTI creators justify their calculation of the political and economic transformation scores with the high correlation of the two based on their interdependence (Ibid.: 127), although no further analytical explanation is provided. The written assessments of the country expert surveys – based on de facto implementation rather than de jure manifestations) are translated

⁶¹ Due to its normative set-up, the BTI team admits that authoritarian states are automatically disadvantaged in the assessment of all criteria that are based on democratic justification of government acting.

into numerical ratings on a scale of 1 to 10 (lowest to highest level of democracy) and then examined in a “multi-stage review process so as to make them comparable both within and across regions”. “In order to ensure the validity, reliability and comparability of the assessment” initial reviews of the country experts’ assessments are conducted by the regional coordinators “who examine the content to ensure it is both complete and consistent”, followed by intraregional and interregional calibration across all 129 countries. The BTI-Board discusses all scores before adopting them. (Ibid.: 129). How this process is executed in detail has not been explained in the methodology handbook. The BTI team acknowledges that its high level of aggregation reduces the complexity of any inspected country wherefore they publish all non-aggregated individual scores online and for free.

The specific questions the experts have to answer with view to the RoL are those comprised in Table 14 below and fully correlate with aspects covered in both the COM’s Progress Reports and FH’s Freedom in the World analysis:

The BTI’s RoL Questions
To what extent is there a working separation of powers (checks and balances)?
To what extent does an independent judiciary exist?
To what extent are public office holders who abuse their positions prosecuted and penalized?
To what extent are civil rights guaranteed and protected and to what extent can citizens seek redress for violations of these rights

Table 14 – The BTI’s Rule of Law questions

Each question is complemented with a short descriptive paragraph and a range of answers that correlate with specific scores, as illustrated below:

3	Criterion: Rule of Law State powers check and balance one another and ensure civil rights.								
3.1	Separation of powers								
	<p>To what extent is there a working separation of powers (checks and balances)?</p> <p>This question refers to the basic configuration and operation of the separation of powers (institutional differentiation, division of labor according to functions and, most significantly, checks and balances). However, it does not refer to the tendency toward convergence and a fusion of powers that can be observed in parliamentary systems. It does include the subjection of state power to the law.</p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="border-top: 1px solid black; border-bottom: 1px solid black;">There are no constraints on the basic functions involved in the separation of powers, especially mutual checks and balances.</td> <td style="text-align: right; border-top: 1px solid black; border-bottom: 1px solid black;">10 <input type="checkbox"/> 9 <input type="checkbox"/></td> </tr> <tr> <td style="border-top: 1px solid black; border-bottom: 1px solid black;">The separation of powers is restricted partially and temporarily (e.g., to ensure governability). Fundamentally, though, a restoration of balance is sought, especially by the other branches.</td> <td style="text-align: right; border-top: 1px solid black; border-bottom: 1px solid black;">8 <input type="checkbox"/> 7 <input type="checkbox"/> 6 <input type="checkbox"/></td> </tr> <tr> <td style="border-top: 1px solid black; border-bottom: 1px solid black;">One branch, generally the executive, has an ongoing and either informally or formally confirmed monopoly on power, which may include the colonization of other powers even though they are institutionally differentiated.</td> <td style="text-align: right; border-top: 1px solid black; border-bottom: 1px solid black;">5 <input type="checkbox"/> 4 <input type="checkbox"/> 3 <input type="checkbox"/></td> </tr> <tr> <td style="border-bottom: 1px solid black;">The separation of powers is nonexistent or exists only on paper.</td> <td style="text-align: right; border-bottom: 1px solid black;">2 <input type="checkbox"/> 1 <input type="checkbox"/></td> </tr> </table>	There are no constraints on the basic functions involved in the separation of powers, especially mutual checks and balances.	10 <input type="checkbox"/> 9 <input type="checkbox"/>	The separation of powers is restricted partially and temporarily (e.g., to ensure governability). Fundamentally, though, a restoration of balance is sought, especially by the other branches.	8 <input type="checkbox"/> 7 <input type="checkbox"/> 6 <input type="checkbox"/>	One branch, generally the executive, has an ongoing and either informally or formally confirmed monopoly on power, which may include the colonization of other powers even though they are institutionally differentiated.	5 <input type="checkbox"/> 4 <input type="checkbox"/> 3 <input type="checkbox"/>	The separation of powers is nonexistent or exists only on paper.	2 <input type="checkbox"/> 1 <input type="checkbox"/>
There are no constraints on the basic functions involved in the separation of powers, especially mutual checks and balances.	10 <input type="checkbox"/> 9 <input type="checkbox"/>								
The separation of powers is restricted partially and temporarily (e.g., to ensure governability). Fundamentally, though, a restoration of balance is sought, especially by the other branches.	8 <input type="checkbox"/> 7 <input type="checkbox"/> 6 <input type="checkbox"/>								
One branch, generally the executive, has an ongoing and either informally or formally confirmed monopoly on power, which may include the colonization of other powers even though they are institutionally differentiated.	5 <input type="checkbox"/> 4 <input type="checkbox"/> 3 <input type="checkbox"/>								
The separation of powers is nonexistent or exists only on paper.	2 <input type="checkbox"/> 1 <input type="checkbox"/>								

Illustration 7 - BTI RoL Question specification

Strengths and Weaknesses

When comparing the BTI to Freedom House, similar strengths and weaknesses can be identified. The BTI has a wide range of countries which it assesses regularly (every two years). Also, it focuses on de facto changes rather than mere revisions of statutes. In contrast to Freedom House, it does publicize its disaggregated data. However, all other points of critique are similar. The BTI bases its evaluations on a biased concept of democracy, although in contrast to Freedom House it does include socio-economic factors. It does not theoretically justify its selections of indicators nor its sub-questions. It provides no clear anchoring examples on how to answer and later score each question on the scale of 1 to 10. It does not sufficiently deduce the interconnectedness of democracy and its economic variables wherefore the aggregation of its scores remain unfounded. Since - during the time period analyzed - none of the measuring, coding or aggregation rules were made public, the evaluations could not be reproduced by independent scholars.

Table 15 below summarizes the findings from above using the criteria offered by Munck/Verkuilen.

Criterion	COM Progress Reports	Freedom House	BTI
1. Rule of Law Conceptualization			
1.1 Minimalist/ maximalist definition?	Maximalist, yet not conclusively defined	maximalist	maximalist
1.2 Vertical organization of attributes by level of abstraction?	no	yes	yes
2. Measurement			
2.1 Valid + reliable selection of indicators?	no	Indicators conflated	Lack of theoretical justification
2.2 Valid selection of measurement level?	no	Arbitrary → Individual scores (0-4) → Transformation of total point on 1-7 scale	Arbitrary → Individual scores 1-10
2.3 Replicability through publicized coding material and disaggregate data?	no	Not sufficient	Not sufficient
3. Aggregation			
3.1 Valid level of aggregation?	Not applicable ⁶²	No, not theoretically justified	No, not theoretically justified
3.2 Valid selection of aggregation rule?	Not applicable	No, combination and setting off against each other of pol. + ec. scores not theoretically sufficiently justified	No, combination and setting off against each other of pol. + ec. scores not theoretically sufficiently justified
3.3 Replicability though publicized aggregation rules and data?	Not applicable	no	Disaggregated individual scores online, aggregation rule described
Scores			
Nr. of fulfilled criteria	0	1	2
Nr. of unfulfilled criteria	4	7	6

⁶² 3.1-3.3 are not applicable for the COM Progress Reports because they do not include a quantitative assessment

Nr. of not applicable criteria	3	0	0
---------------------------------------	---	---	---

Table 15 - Fulfillment of Munck/Verkuilen's index quality test

III.4 Implications of Indices' Methodological Set-up for Empirical Analysis

All three indices show sufficient similarities to compare their evaluations to one another. All indices are based on a similar understanding of democracy and its sub-component, the RoL, as is portrayed in similar questions/aspects covered in each institution's report. Furthermore, all texts have undergone similar, comprehensive calibration and harmonization processes within the case study and across the countries covered which is likely to counterbalance evaluations made by individual authors.

However, the indices differ in their aim, range, methodological approach and the depth of resources upon which their reports are based. The COM's Reports are evaluating individual, specific progress from the candidate's last year's status quo and seek to incentivize it to continue its reform efforts in preparation for accession. In contrast, FH and BTI assess the general level of democracy with a country in comparison to all other case studies. Consequently, deviating results can be expected given the evaluation of factually similar developments. Additionally, the more countries are compared to one another, the less detailed each country's description and evaluation will become. For example, FH assesses almost all countries in the world, non-dependent of their democracy-status. In contrast, BTI only compares countries in transition, which all face similar challenges that cannot be compared to fully established democracies. The COM covers the smallest range of countries (seven) which – due to their candidate status – are also in transition to becoming more democratic countries. Naturally, they receive a bigger scrutiny than the 150+-countries analyzed by FH. This is also reflected in the range of Democracy and RoL- elements tested within their conceptualizations and methodological frameworks: the COM has the most comprehensive set of RoL-elements whereas BTI and FH just issue a small section of their analysis to the RoL. However, they all tend to highlight the same RoL-elements as particularly important. Finally, the COM can allocate a much bigger range of resources (staff, connections, time, financial assets) to the evaluation of each candidate. Not only is the European Union financially more apt than FH and BTI but its scrutiny also equals an investment into its own security. Including a 'badly

vetted' country into the Union would be disproportionately more expensive than carrying out a comprehensive testing process.

When tested against Munck and Verkuilen's quality criteria, none of the three indices appears to conform to academic standards of reliability and reproducibility due to their (partially) intransparent and (partially) discretionary decision-making and aggregation mechanisms. However, there is no such thing as a perfect index and since all three employ similar reviewing and harmonization mechanisms, their results may differ slightly but should generally produce rather similar results unless other interfering factors can be substantiated.

IV. Political Bias Claims

Since the European Union is a political institution politically biased accession decisions may be expected intuitively. Political bias in this respect has been defined as the intentional adaptation of evaluation results according to the EU's political preferences regarding the (non-) accession of particular candidate countries. The interviewed experts from the COM agreed that in general political considerations do influence the accession process. However, the Copenhagen Criteria were invented to make it more feasible and transparent once a country has been signaled that it was on the accession track. Hence, once the criteria have been introduced, using them adequately and applying them to each candidate equally is a precondition to the functioning of said criteria in 'getting the right ones in and keeping the wrong ones out'. Therefore, the employment of the criteria should be as little arbitrary as possible. The following paragraphs will indicate that the EU has been accused of employing the CC arbitrarily. Moreover, the importance of valid and reliable progress evaluations from the European Commission in keeping the Union stable and credible will be advocated.

Suspicious about possible political bias regarding the EU accession and particularly the evaluation process have most prominently been uttered by academia (IV.1), Member States (IV.2), the media, NGOs, and candidate countries (IV.3). One could argue that such critique is in itself biased. Media outlets profit economically from 'EU-bashing' in terms of increased runs. NGOs receive more funding when they fulfill their 'watchdog function' by exposing and addressing potential inconsistencies. Ostracizing the EU has also become a popular tool for national politicians in order to deflect from their own responsibilities.⁶³ Especially complaints from candidate countries can be more easily dismissed since they could be related to their disappointment regarding the individual accession progress. However, all of those aspects do not automatically discredit the suspicion. Plus, a limited but growing field of academic research has been dedicated to examining such claims, this thesis being part of them. In order to substantiate the scientific interest of the research project, examples of this critique will be examined in the following paragraphs.

After introducing and categorizing the range of political bias accusations the COM is facing, the second part of the chapter examines the negative consequences the political bias claim has

⁶³ In August 2014, UK's Farage (UKIP), France's LePen (Front National) and Dutch Wilders (Partij voor de Vrijheid) blame the EU for starting a new Cold War on Russia (Escritt/Deutsch, 2014).

on the Union's inner and outer stability (IV.4), the possible repercussions of political bias for Member States and candidate countries (IV.5), the importance of reliable Progress reports (IV.6), and the challenges surrounding the eradication of political bias suspicions (IV.7).

IV.1 Academia

Compared to Academia, all other actors' critiques are more likely to be biased due to underlying agendas. Academia, in contrast, is preoccupied with 'truth finding' because it does not cater to any specific interest group, at least none privy to the EU accession process. Consequently, critique that is being voiced by scholars will be given more attention within this summary. Although only a few scholars have dedicated their research to examining a bias within the accession process, there are five lines of critique that are being represented in their works. The first category claims that accession is in fact only a *political question* and that the fulfillment of the Copenhagen Criteria is secondary (Ludwig, Maresceau, Zalewski, Saatcioglu). The second one questions the *aptitude and usability* of the criteria, i.e. their set-up and lacking definitions (Kochenov, Ludwig, Wennerström, Zalewski). The third category criticizes their *application* and unveiled examples where countries were accepted into the Union although criteria had de facto not been met (Kochenov, Ludwig, Wennerström). The fourth category objects to *double standards between EU members and non-members*, stating that candidates have to fulfill conditions existing members do not comply with themselves (Zalewski, Moravscik/Vaduchova, Spreeuw, Saatcioglu). The fifth one addressed *double standards among candidate countries*, that is, if individual states are being treated differently although they show similar performances within their reform processes (Ludwig, Spreeuw, Zalewski). The following paragraphs will provide a synopsis of the scholars' arguments and findings.

IV.1.1 Category 1 – Accession as a Political Decision

The scholars of the first category agree that the accession process is predominantly based on political decision-making rather than the fulfillment of the specific criteria. Particularly Ludwig (2011: 77) doubts the objectivity of the criteria and indicates that sources used by the Commission to write the evaluations may not be as decisive in the actual Reports as underlying political motives. The most prominent example to substantiate the claim is Turkey. Ludwig (2011: 76) states that the COM only dedicated two lines in its Agenda 2000 program to Turkey which was targeted at describing the reform plans for future Member States. Moreover, a simultaneously published document that was specifically designed to outline EU

relations to Turkey lacked detailed information on future integration plans as well. Ludwig and Zalewski (2004: 34) assume political motives behind the “discriminatory” treatment. Saatcioglu (2013) agrees that “regardless of the degree of Turkey’s compliance membership is ultimately the EU’s decision and it is a political decision since it has to be approved by all existing EU member-states unanimously” and that “can always be blocked within the Union based upon member-states’ political considerations” Zalewski (2004: 39) amends this notion by stating that Austrian Chancellor Wolfgang Schüssel and French President Jacques Chirac have both called for national referendums on the issue of Turkish membership based on their disapproval of Turkish accession. Such a vote has been unprecedented in previous enlargement rounds and therefore also constitutes an example of double standards among candidate countries (category 5). In order to show that Turkey is not the only country where political motives may influence the accession process, Zalewski points at Poland whose acceptance to the Union he accredits to Germany’s interest in the symbolism of reconciliation after WWII (2004: 34).

IV.1.2 Category 2 – Lack of Aptitude and Usability of Criteria

The scholars of the second category address three questions: Are the criteria sufficiently clear to be used as a benchmark? If not, why? What are the consequences of the ambiguity? Kochenov (2009: 300) issues the harshest critique of all scholars, describing the Copenhagen Criteria as a “resounding failure”, the Commission as a “myth-maker” for using them and related documents as “lacking depth and consistency”. First, he asserts that there is “no single document to clarify the meaning of the criteria” wherefore they remain “vague and general” (2004: 7). Only a synopsis of all accession related documents provides an elusive idea of what the Union is asking of their candidate countries. This critique is backed up by Wennerström who considers it a “paradox” to make the CC “a pre-requisite for membership” given that “the lack of a uniform conception was not alleviated by the elevation of the rule of law status of legal founding principle for all EU Member States through the TEU revision” (2009: 69). Kochenov then stresses that due to the ensuing “over-inclusiveness” of the criteria, the accession process degenerates to a “game of guesses” because it remains unclear when a country is being compliant or not. He continues asserting that although the Criteria were initially developed to simplify, improve and depoliticize the accession process, they have in fact done the opposite (2004: 2). According to Kochenov, particularly the fifth enlargement round was not “merit-based” and the relevant Reports were neither complete nor impartial.

Ludwig (2011: 77) follows up on Kochenov's critique but provides a more deliberative account. She agrees that the criteria lack differentiation, particularly with view to the joined concepts of democracy and rule of law, but also within the individual concepts (2011: 122). Ludwig criticizes that there is no consolidated, clear, European RoL concept let alone specific accession criteria (2011: 214) and that the aptitude of the documents is limited to being a general guide for carving out such a concept through induction (2011: 77). She concedes that the general setting of priorities is understandable but that no legal minimum standards are being introduced to be used as benchmarks. However, Ludwig evaluates those shortcomings as comprehensible based on the fact that reforming the judiciary is a troublesome and lengthy process and suggests that the COM used the accession as a leap of faith to incentivize the candidates for future reforms (2011: 119). Guarnieri and Piana (2009: 2) point out that the lack of clarity may be based on different traditions and that in "Western Europe the relationship between courts' independence and the rule of law is not always clear". Not using specific criteria could accommodate those different traditions and thus omit having to pick 'the right one' and upsetting states that chose a different path for their judiciary. Zalewski (2004: 33) interprets the lack of clarity as a method from the Union to increase its flexibility "both to use its leverage to respond to new problems in the candidate countries, and to weigh factors other than the fulfillment of accession conditions" which in turn corresponds to an increased likelihood of political decision-making (category one). Ludwig (2011: 120) reasons that the lack of clear criteria may be dangerous because without clear minimum standards the decision regarding the opening of negotiations or the accession maturity remain arbitrary which is likely to cause unequal treatment among candidate countries (category five).

IV.1.3 Category 3 – Inconsistent and Wrongful Application

In category three, Kochenov, Wennerström and Ludwig further highlight that the implementation of the criteria is problematic. Wennerström observes that especially during the major enlargements from 1997 to 2004, the EU accession RoL approach was "definition by action – no declaration was made as to what vision of the rule of law the Candidate countries were to achieve in order to satisfy the Commission and the EU's requirements" (2009: 69). Kochenov criticizes that in general "little space is dedicated to the analysis of the rule of law in the documents, compared to other criteria, and no reasons behind picking this or that topic for analysis are given." (2004: 30/31). Moreover, he blames the Copenhagen European Council for having established criteria without clarifying "the principles for the

assessment of the progress towards meeting them or the actual means to measure the conformity with them” (2004: 5). Wennerström supports his critique of inconsistency, even within the criteria with certain components being “considered more important than others” and some being “measured in a binary on/off way whereas other components were seen from a progress/no progress perspective” (2009: 72). He also complains about the lack of importance in meeting specific targets, stressing that “the most important factor appears to be a demonstrated willingness to move in the right direction” (2009: 70). Kochenov agrees with this notion and condemns that the criteria have been considered as met even in situations “where judicial output is sub-standard or even “unacceptable”, and where civil service is unregulated and subject to widespread bribery” (in Ludwig, 2011: 212/213). Polemizing his point, he adds: “...as far as the judiciary is concerned, a heavily understaffed and insufficiently trained body of judges without any technical assistance, slowly passing ‘unsatisfactory’ decisions, which are often not executed at all ... seems to be enough [for accession]” (2004: 21). As a consequence of missing to link “the fulfilment of the Copenhagen Criteria with the real progress in the candidate countries” the Commission “deprived itself of needed room for maneuver” (2004: 24). Although Ludwig attempts to explain some of the apparent deficiencies she acknowledges that the criteria are not clear enough to be implemented uniformly. Instead, she finds that in practice they were being stretched to a point of complete randomness (2011: 120).

All the categories introduced above may result in either deliberate or incidental double standards regarding the accession process. The following two categories outline double standards between members and candidates as well as among candidate countries.

IV.1.4 Category 4 – Double Standards between EU Member States and Non-members

Category four exposes how the EU asks its future Member States to fulfill criteria that some of the Member States themselves are not compliant with. Spreeuw observed that certain values and principles that are required of the Western Balkan countries with view to democracy and the RoL have not been sufficiently implemented by existing Member States. Moravscik and Vachudova concretize this notion by adding that particularly in the fields of “internal democracy, state admin, and detailed regulatory protection...the EU-15 have had half a century to accommodate” the relevant standards whereas candidate countries are given a mere fraction of the time while expected to deliver above-average results. Moreover, in some areas such as the protection of ethnic minority rights, “candidates are asked to meet

standards that the EU-15 have never set for themselves” (2003: 46). Saatcioglu delivers an example to illustrate the problem in practice. Regarding Turkey’s accession, the Cyprus dispute regularly stalls the negotiations. However, instead of requiring both parties to the conflict to compromise, Greek Cyprus and Greece can rest on their position. While she acknowledges the Union’s right to expect a settlement of the conflict in order to fulfill the “good neighborly relations” criterion [...] what’s wrong and unfair is the fact that the EU only raised this condition to Turkey, and not to the other sides to the Cyprus dispute, namely, Greek Cyprus and Greece as a guarantor state” (2013). Schimmelfennig et.al. (In: Zalewski, 2004: 29) emphasize why double standards are generally derogatory to EU integration. The “compliance pull” of the criteria is only strong enough if criteria are “clearly defined, shared among the Member States, and coherently applied by the Union as a whole”. “Alternately, if ‘double standards’ become perceptible in the actor state-target state relationship, conditions will fail to exert the same leverage” (Ibid.).

IV.1.5 Category 5 – Double Standards among Candidate Countries

Category five describes another double standard, namely the differential treatment of candidate countries by the EU. The poster child of discriminatory treatment is Turkey. Although the Union clearly stated in its “Helsinki affirmation that Turkey is a candidate State destined to join the Union on the basis of the same criteria as applied to the other candidate States” Zalewski finds that such remarks “smack of hypocrisy” (2004: 105) given that numerous EU Member States called either for national referendums on Turkey’s accession or alternative cooperation models. Ludwig provides more examples, stating that although the Czech Republic in 1997, Romania in 2007 and Croatia in 2004 struggled with similar democracy and rule of law issues they were being evaluated and treated differently leading to different acceptance dates into the Union (2011: 188/119). Acknowledging that some of their new Member States had not reached maturity in some areas including the RoL, the EU established a post-accession verification mechanism for Romania and Bulgaria (Spreeuw, Unknown: 2). On the one hand, this illustrated that those two countries were given more leeway than other candidates states. On the other hand, the lack of effectiveness of such tools in strengthening the judiciary and fighting corruption and organized crime becomes evident when countries do not share the same values as the existing Member States. The findings of this analysis are summarized in Table 16 below:

Critique Category	Points of critique	Scholars
Accession = mostly political question, CC secondary	<ul style="list-style-type: none"> • Political motives more decisive than Progress Reports • Discriminatory treatment of CS 	Ludwig (2011) Maresceau (2003) Zalewski (2004) Saatcioglu (2013)
Aptitude and usability of CC	<ul style="list-style-type: none"> • Criteria insufficiently clear to be used as benchmark (over-inclusive, lacking depth and consistency) • Lack of legal minimum standards • Lack of clear concept may be intentional to grant flexibility 	Kochenov (2004, 2009) Wennerström (2009) Ludwig (2011) Zalewski (2004) Guarnieri/Piana (2009)
Application of CC	<ul style="list-style-type: none"> • Inconsistent application deprives EU of room for maneuver 	Kochenov (2004) Wennerström (2009) Ludwig (2011)
Double standards between MS and CS	<ul style="list-style-type: none"> • Requirements made of CS have either not been set for or not been fulfilled in some MS themselves • Double standards endanger functioning of conditionality principle 	Zalewski (2004) Marvscik/Vachudova (2003) Spreeuw (Unknown) Saatcioglu (2013)
Double Standards between CS	<ul style="list-style-type: none"> • Hypocritical evaluation and treatment of CS • Post-accession conditionality as example for preferential treatment 	Ludwig (2011) Spreeuw (Unknown) Zalewski (2004)

Table 16 - Areas of Critique regarding Copenhagen Criteria from Academia

Although this synopsis sufficiently proves the need to examine the Reports' validity, for the sake of completeness, examples from other critical sources (media, NGOs, politicians) will be introduced briefly.

IV.2 Member States' Positions towards Accession

Traditionally, Member States' had differing positions on EU enlargement ranging. A divide persists between those seeking quick accessions versus those trying to postpone or limit it. Some Member States are interested in controlled accession and others base their decisions on what they considered to be legally and historically just (Veebel, 2011: 9). Ever since the last accession rounds of 2004 and 2007, Member States grew more cautious in their support of candidate countries. This can be attributed to two aspects: first, upon their inclusion in 2007, Romania and Bulgaria did not fulfill the Copenhagen Criteria and have since been reprimanded for significant breaches of the Acquis and possible returns to authoritarian leadership. In order to avoid such developments, Member States react by increasing the range of accession criteria while simultaneously leaving them linguistically vague enough for interpretation, thus raising the threshold for candidates to become new members (Redmond, 2007). Secondly, Member States started to see enlargement as a zero-sum game where both economic means and political power (forfeit of sovereignty) have to be shared with even more countries, thus limiting one's own piece of the cake (Moravcsik, 1999). It is against this background that one has to examine the Member States' opinions on and reactions to Croatia's and Turkey's accession process.

IV.2.1 Croatia's Membership

In contrast to Turkey, Croatia was more successful in lobbying Europe-wide support for its accession, particularly among the more important (bigger and economically stronger) Member States. Germany and France were among the first supporters of Croatia's independence and eventually, EU accession during the Zagreb meeting of 2001 which was held under the French Presidency. The support remained high through 2005, possibly because in contrast to other states of the region, Croatia refused to ally with the US and Britain in the war against Iraq (Jovic, 2006: 101). Once Croatia in its post-Tudjman-era accepted that it was no regional power but part of a small group of states with which it sought closer cooperation, most Central European states (Austria, Slovakia, Czech Republic) and the three Baltic states Lithuania, Estonia and Latvia also supported Croatian accession. The best strategic move was lobbying with the Vatican as it convinced the three catholic countries Poland and Ireland to

support Croatian membership as well. Countries that were more critical towards Croatia's accession included Britain (due to Croatia's opposition to the Iraqi war) as well as Italy and Slovenia. Italy claimed compensation interests for ethnic Italians who were displaced after the WWII while Slovenia raised maritime border issues as well as compensation concerns over frozen assets during the breakup of Yugoslavia (Ibid.: 103).

IV.2.2 Turkey's Membership

Officially, European Opinions on Turkey's accession are mixed. The Union's biggest and most important players (in terms of decision power and financial contributions), Germany, France and Austria, oppose Turkey's membership and aim at the long-term perspective of a privileged partnership. Their governments' opinions are largely backed up by public opinions within their countries. While they acknowledge that Turkey does contribute significantly to European security through NATO (which it already is a member of), doubts spiral around economic weakness, political shortcoming especially with view to rights and liberties, RoL and the freedom of the media. Repeatedly, cultural and religious differences have been criticized as well. A meta-analysis of German, French, British and Turkish press on Turkey's accession showed that 46% of all concerns voiced in German and French media related to those cultural-religious differences (Negrine et.al., 2008: 59). French President Giscard d'Estaing publicly stated in 2002 that "admitting Turkey to the EU would be the end of the European Union because its capital is not in Europe and 95% of its population live outside Europe" stating that therefore "It is not a European Country" (Ibid.: 49). While citing geographic reasons for Turkey's rejection it is more likely that France's difficult experience with its large Muslim population as well as its Armenian minority (Christian but non-European) impacted its perspective (Negrine et.al., 2008: 50). Greece and Cyprus tend to block Turkish accession as well based on regional historical and territorial conflicts but developed a more pragmatic approach in the last years. In contrast, Britain echoes the US policy favoring Turkey's membership based on security interests. Other countries favoring Turkey for strategic reasons include Poland, Romania, Italy, Spain, and Portugal. The former two are the biggest countries to join the Union since 2010. While they publicly support Turkish membership, they remain fearing that EU subsidies could be re-attributed to Turkey. The latter three support Turkey's accession hoping to restore the balance between the powerful North and the underrepresented South of Europe. Yet, they share Germany and France's general concerns regarding Turkey's democratic deficiency and cultural-religious

differences (Redmond, 2007: 305-209). Countries whose governments are divided or marginally in favor of Turkish accession include the Czech Republic, Slovakia, Hungary and some other Baltic countries.

Unofficially, a German MEP described the Turkish EU-accession as a “dishonest process fooling the respective populations” calling it “a zombie that walks the streets of Brussels”. He stressed that there was consensus in Brussels that Turkey will never join, that the negotiations are “dead” and “will never lead to conclusion” (EP, Brussels, July 2016). Especially with the Brexit decision – with the UK being Turkey’s biggest supporter in joining the Union – Turkish accession became even less likely. He added that the UK’s motivation behind supporting Turkey in its bid for accession was founded in the calculation that the more members the Union had the less likely (unanimous) decisions became which in turn was in the British interest. In light of Brexit and recent anti-democratic political developments in Turkey, it is improbable that British support of Turkish EU accession continues.

IV.3 Media Coverage of Political Bias expressed by Political Actors and NGOs

The majority of political bias allegations regarding the EU accession process can be found in mainstream media and mostly stem from either parties to the accession process or NGOs who call the public’s attention to possible intransparencies or inconsistencies. The following paragraph will offer a non-exhaustive synopsis of examples that illustrate the general range, body and tone of critique. Throughout the EU enlargement process, individual politicians have repeatedly criticized the execution and results of the negotiations for being carried out in an unequal manner. This critique has been particularly intense with view to the case studies at hand, Croatia and Turkey. Pertaining to Croatia, the start of the negotiation talks had been made dependent on the capture of a Croatian war criminal, Ante Gotovina, suggesting that once negotiations were conducted efforts to track down former Yugoslavian war criminals would be reduced and could “set a bad precedent for efforts to track down...Bosnian Serb indicted war criminals” (Financial Times, 2005). Doris Pack (MEP), chairwoman of the South-East-Europe Delegation of the European Parliament, claimed this premise to merely be pretext based on the “obviously premeditated intention to postpone the agreed date for the admission of Croatia”. Pack argued that the “Western world” had no genuine interest in the successful capture because it would mean that accession negotiations had to be started which

was disapproved of by several Member States (EPP Group, 2005). At the same time, Austria's chancellor Wolfgang Schüssel denounced the EU's double standard regarding the treatment of Croatia in contrast to Turkey, whose negotiation talks had begun one day earlier, on October 3rd, 2005. Schüssel said if Turkey was trusted to "make further progress we should trust Croatia too..." instead of insisting on capturing Gotovina first. He added that "it is not fair to leave Croatia in the eternal waiting room" (Ibid.).

Claims of unfairness have also been uttered by Bulgarian and Romanian officials based on the process of post-accession monitoring through the European Commission. Since both countries were seen as "insufficiently prepared on the eve of accession" the EU-25 decided to scrutinize their level of corruption while disregarding similar problems in other Member States (EU Observer, 2007). At a Brussels round table discussion, Bulgarian Interior Minister Roumen Petkov demanded common standards to guarantee an "objective evaluation of every member state". Although Bulgarian diplomats "admitted [Petkov's] initiative was primarily meant to counter the image of their country as particularly sleazy" they argued that only Bulgaria and Romania were being tested whereas other countries such as Italy had far more pressing corruption issues (Ibid.). The issue of double standards, both between new and old members as well as among candidate states, has also been addressed repeatedly by NGOs. Prior to the 2004 accession round of ten Central Eastern European Countries (CEECs), Minority Rights Group International has exposed and denounced the fact that although the ratification of the European Framework Convention on National Minorities (FCNM) was a prerequisite for the accession of the CEECs a number of EU Member States including Netherlands, Belgium, Luxembourg, France and Greece had themselves not ratified it (Minelres, 2005). This was particularly relevant because the FCNM constituted the "only legally binding treaty specifically to protect minorities" (Ibid.). Three years later, the Bulgarian NGO Centre for the Study of Democracy criticized the European Commission's post-accession monitoring of anti-corruption progress. Boyko Todorov asserted that if the EU wanted to remain credible it had "to develop better standards [because] identifying progress on corruption now remains largely arbitrary" (Ibid.). Todorov based his notion on the fact that the European Commission used Reports by Transparency International to determine the level of corruption which merely measured the level of perceived corruption and is therefore prone to error. While EU officials agreed that the monitoring process was difficult because "different Member States record

crimes differently”, he pointed out that Bulgaria and Romania had readily accepted the process in order to avoid the alternative: a one-year delay of accession (Ibid.).

IV.4 Negative Consequences of Political Bias Claim: Inner and Outer Credibility of the EU

If politically biased COM evaluations can be proved (or just not ruled out) this is a significant finding because ‘wrongful’ accession decisions can be detrimental for all parties involved: the inner and outer constitution and reputation of the European Union, the individual member and candidate countries as well as potential future Member States. Allegations of double standards always carry the connotation of illegitimacy, or at least unreasonableness. Hence, the addressees of the accusation will feel the need to counter these accusations either by showing that there is actually no differential treatment or by seeking justification. The EU’s unique selling point is full membership because it promises democratic progress, political stability and economic prosperity. These commodities are sought after by the candidate countries.⁶⁴ However, they come at a price, namely compliance with the Copenhagen Criteria which usually requires extensive reforms on the candidate’s part. Those reforms are usually very hard to implement because they require a restructuring of the political and economic set-up of a country which regularly upsets whoever has profited from the previously established system. It is particularly difficult for a government to communicate and implement changes (especially recesses in the social sector or tax raises) to both their citizens and established power elites if the respective reforms do not visibly catalyze the accession process.

IV.4.1 Inner Credibility of the Union

Behavioral psychologists have established that group dynamics are heavily influenced by rules and individual behaviors (Rogers/Senturia, 2013). This is also true for the EU which is essentially a group of individual states. If rules within a group are created but not enforced this will cause the group to disobey not only the respective rules but others as well, based on the projection that if one breach of rule will remain unsanctioned so will another (Cialdini et.al., 2014). Also, group members are very adaptive to the behaviors of their fellow group members. This means that including an “immature” state into the Union that either cannot or chooses not to comply with the Copenhagen Criteria is likely to cause existing Member States to adapt their behaviors to his detrimental one. This would endanger the current integration

⁶⁴ Although they are better described as values most candidate countries are likely to conduct a cost-effect analysis regarding their membership, therefore I will use this economic term

status and could economically, politically and judicially endanger the stability of the Union. For example, anti-democratic policies as have been established in Hungary can cause other Member States to implement similar regulations, thus reducing the quality of liberal democracy within the Union.⁶⁵ Accepting an ‘immature’ country into its ranks also increases the financial risks Member States are taking. A candidate country receives its subsidies from different funds than a member state. The funds granted to Member States are limited. This creates a zero sum game – the more Member States qualify for said funds, the less support each recipient can receive. Since an ‘immature’ country may require more support to fulfill its duties, this money will be drawn from the funds all other Member States have access to, thus reducing the amount they can claim for themselves. This is likely to cause envy, and eventually, political conflict.⁶⁶ As has been pointed out by Trauner (2009) even post-accession conditionality is ineffective in solving these problems. Once a country becomes a full member, the Union has de facto very little leverage in securing its compliance. This became evident when considering the latest EU financial crisis and the (initial) reluctance of individual states to support bailout programs for its peers (BBC, 2010). All the challenges described above can lead to a decline in support by the EU-citizens which translates into reduced voting participation and may results in limited democratic legitimization.

IV.4.2 Outer Credibility of the Union

In international relations there exists the power dichotomy of hard and soft power (Nye, Campbell/O’Hanion). The term hard power refers to military and economic means used to create compliance with one’s interests against someone else’s goals. Soft power, in contrast, emphasizes diplomatic means and focuses on the usage of cultural and historic aspects to win support for one’s goals. The EU has long suffered from the image of a “toothless tiger” (Pauly/Schult, 2012) due to its lack of joined military forces and a generally limited involvement in international armed conflicts (Youngs, 2014). In contrast, any of the super powers of the past and present were/are able to back up their policies and protect their interests using military force if need be. Instead, the EU earned the reputation of a soft power heavyweight based on inner economic and political stability, the successful integration of numerous new Member States as well as being a reliable partner in international negotiations

⁶⁵ For more information on Hungary’s democratic decline see Shirrefs (2014).

⁶⁶ In fact, this is true for any new member state that is likely to receive subsidies. However, the bigger the problems in a given state, the more financial assistance is needed to address them in order to not risk the destabilization of the state and the Union.

(Gordon et.al., 2008). This reconciles the lack of coordinated or joined military power because the EU is acknowledged to be a peaceful actor that would not attack another country to pursue its own interests and in turn would be an ill-advised target due to its many political and economic allies. However, all soft power aspects that turn the EU into a strong international actor will be challenged through biased accession decisions.

Economic and political stability are deeply intertwined. Only within a system of reliable political and judicial structures investments will be made that create infrastructure, employment opportunities, and enable consumerism which in turn generates tax revenues for the state to be used for further development. Since all EU Member States are economically and politically interdependent, failing systems automatically affect functioning systems and weaken them. In a best case scenario, the functioning systems can absorb the losses of the failing systems (similarly to the bail-out policies during the financial crisis). In a worst case scenario, the failing systems pull the functioning ones down which will eventually fail themselves. Either way, this causes negative effects on the overall stability of the Union. This aspect ties in with another focal point, the successful integration of new members. The EU is highly respected for exporting its democratic rules and norms as well as economic prosperity to previously undemocratic and less developed countries while growing stronger as a union. If the EU loses control over this process because new Member States are unable or unwilling to fulfill the required criteria, this will reflect worse on the Union than on the respective Member States because it is supposed to be in the superior power position to either convince or force its members into compliance. This, in turn, ties in with the third aspect: the EU as an internationally respected moderator and negotiator. If the EU cannot convince its own Member States to follow suit with view to the development of democratic and RoL structures, economic regulations or peaceful conflict solutions, it is unlikely to remain a desired moderator in international crises. Combining the factors outlined above, aspects that decrease the EU's inner stability automatically infringe upon the outer perception of the Union and can have a lasting negative effect on its reputation causing a decline in the Union's leverage in international relations. Since the EU is a union of individual states, their clout may individually be affected by biased accession decisions.

IV.5 Repercussions of Political Bias on Member States and (potentially new) Candidate Countries

Accepting a new member state into the Union always results in a shift of geographical borders to the outside and the maceration of borders between neighboring old and new Member States. Consequently, mutual exchange is possible and in fact, sought after, especially with view to trade relations. However, with ceasing border controls, there is no filter any more for what is being transported or exchanged. Consequently, negative repercussions including crime (drug and human trafficking), corruption, poverty, and diseases are more likely to be imported into the European Union. Dealing with the repercussions of those unintended effects is economically and politically challenging because individual governments will struggle to explain to their citizens (and eventually be held accountable for it) why they agreed to the accession in the first place and why tax money will have to be used to remedy the resulting problems. This not only reduces the citizens' trust in their governments, the European Union and future accession rounds, but can also cause racist tendencies as can currently be observed in an upswing of right-wing parties in the European Parliament, particularly in countries that have high migration counts such as Great Britain, France, or Italy (Grabow/Hartleb, eds., 2013). Member States joined the Union with the notion that they would profit more from it than they had to invest in the community; that they would prosper through cooperation and trade rather than having to take care of others while protecting themselves along the way. That is why the Copenhagen Criteria were invented in the first place. If countries are accepted into the Union based on biased Reports, that is, if their maturity level proves too low to fulfill the Criteria, the risk of causing all the effects described above is high. Looking at the developments in Romania and Bulgaria, both countries that did not fully comply with the Criteria upon accession, reinforces these fears, particularly with view to the spread of organized crime and corruption (Gateva, 2010). However, this process is not only being critically followed by the current Member States but also by the attentive eyes of potential future Member States.

Candidate countries often look at previous accession rounds in order to understand what is being asked of them. They serve as orientation tools to design their own accession process. Although most of the negotiations are more or less 'dictates' of the European Commission regarding what regulations to change and what reforms to carry out, the quality of said changes highly depends on how eager the candidate country is in complying with EU

requirements. Countries that join the Union which do not or insufficiently comply with the regulations will likely have negative effects on candidates' reforms, too. Following the bad example of Romania and Bulgaria, other potential Member States could become less serious about fighting crime and corruption hoping to be treated as lenient as Romania and Bulgaria. While it should be in the inherent interest of a country to address those issues for its own development, it is politically difficult and costly to invest in reforms, especially if those reforms limit the influence of previous power elites or require the reduction of social spending. In turn, countries that do implement the required reforms to meet the criteria and are yet declined in their attempts to become full members while seeing other, less mature countries become members, are likely to abandon the membership option and look elsewhere for allies. Since the next closest possible ally for most countries bordering the EU is Russia, disappointed candidate countries could turn to her for support. This has been exemplified by Ukraine's non-signing of the Association Agreement with the EU under Janokowitsch in 2013 in order to secure economic support from Russia. Considering the Russian government's defiance of most of the basic European values including democracy, civil liberties and the RoL, such attachment can neither be sought after by the candidate countries nor the European Union. Similar effects can be expected with view to potential candidate countries which have not yet entered negotiations with the Union. They are more likely to either not address the necessary reforms at all or focus on the political game of gathering support rather than making substantial and necessary changes within their societies. Naturally, the risk of their re-orientation towards other allies is even higher than for candidate countries because they have not yet committed to and invested as much into the accession process as countries that already obtained the candidate status.

IV.6 Importance of reliability and Validity of the Progress Reports

In academia, reliability and validity are clearly defined concepts that define the quality of a research design. The more reliable and therefore valid it is, the more likely the findings will be accepted within the scientific community. Reliability requires that the repetition of research methods by other researchers delivers the same results, ergo a scientific truth. Human judgment methods – such as the progress evaluations employed by the European Commission – are inherently less reliable because “judgment can vary wildly between observers, and [even] the same individual may rate things differently depending upon time of day and current mood” (Shuttleworth, 2008). However, reliability is a necessary condition for validity,

therefore it must be ensured. Validity, in turn, is a superordinate concept that guarantees that the generated research results meet all the requirements of the scientific research method. Internal validity refers to the quality of each step of the research design whereas external validity scrutinizes the results for possible other causal relationships (as described earlier as *unknown unknowns*). In contrast to the other indices to be tested, the European Union does not claim academic reliability and validity for its Reports. However, the fact that both has been challenged through political bias claims from diverse sources (including academia) makes it necessary to test the results. Should they stand the test and political bias can be denounced, this increases the credibility of the European Union and its accession process while discrediting claims that negatively affect its reputation. Should other factors become identifiable that make the evaluations less reliable this would also add to the credibility of the Union because aspect of intentional tempering with the results could be refuted and the true reasons for the political bias perception could be addressed. In turn, proving a lack of reliability and validity would disable the Union from claiming that its critics are merely biased themselves. It would also likely instigate a process that increases the reliability and validity of the evaluations which would also add to the Union's credibility and its long-term stability by avoiding the inclusion of immature countries.

IV.7 Eradicating Suspicion of Political Bias

After examining the negative repercussions of politically biased accession decisions and stressing the importance of reliable and trustworthy Progress Reports this part will address the question whether suspicion of bias can be eradicated at all. From a psychological point of view, facts almost never change preconceived opinions, especially if those are emotionally vested (Lakoff, 2004).⁶⁷ Ergo, even scientific proof of a non-existent political bias is unlikely to change feelings of mistreatment by parties to the accession process. What is true to the actors becomes a question of trust. In order to academically operationalize the concepts of 'truth' and 'trust' this thesis replaces them with the 'facts' and 'reliability'. If the Reports turn out to be reliable this will not automatically install trust in the EU or the accession process but it may decrease the reproduction of old or the creation of new suspicions. This can help

⁶⁷ The studies presented by Lakoff concern political elections and vaccinations. Regular voters tend to vote for the party they identify with even if that party acts against their own interests and even so other parties may be available that represent the voter's actual preferences. People who believed MMR vaccinations to cause autism did not change their opinions after being presented with valid scientific proof that refutes this theory.

alleviate the negative consequence the EU is currently suffering from based on the political bias perception.

Regarding the other groups that are currently criticizing the validity of the Reports (academia, NGOs, media) remaining skeptical is their privilege. However, while it may be beneficial for NGOs and the media to insinuate political bias to create suspicion in order to generate income, academia is profiting from explaining and eventually solving problems. Therefore, it is the role of academia to examine the political bias suspicion and develop recommendations as to how to optimize the progress report tool in order to create more transparency and reliability for the accession process. The research presented in this thesis will confront the notions of political bias as outlined above with facts and depending on its outcome either strengthen or lessen the allegations. Apart from the results, it will draw attention to the issue because even if political bias can be disproved the fact that it is said to exist impedes the accession process and its participants. Most importantly though, it will address the question of reliability, that is the trustworthiness of the Reports.

The thesis challenges the EU to generate more specific RoL requirements and to become more careful with the implementation of the reporting tool in order to prevent arbitrary treatments of its' candidates. Eventually, it emphasizes that trust of all parties to the accession process is created through repeated predictable and comprehensible actions based on clear rules that apply to all EU members equally. Although the suspicion cannot be eradicated altogether, the results of this thesis can help to understand and improve the set-up, implementation and perception of the accession process and its relevant tools.

V. Conceptual Framework

This chapter will outline all relevant variables, justify their selection, define them and clarify how they will be tested throughout the analytical process in order to respond to the main hypotheses.

The thesis examines whether the COM's Progress Reports are politically biased (Dependent Variable I) by analyzing their level of consistency with view to the RoL (indicator of the Dependent Variable I) with Reports on the same countries during the same time period from Bertelsmann and Freedom House. If the COM's Progress Reports are politically biased - as suggested by academia and media (cf. IV.) - their scores should be inconsistent (that is: significantly better/higher or worse/lower) than those of the other indices. After testing the Dependent Variable Political Bias with the Independent Variables Reputation (V.2.1) and Incentivization (V.2.2), the COM's Political Bias claim cannot be substantiated. Consequently, the initial indicator for Political Bias – the Consistency of the Indices' Results – is elevated to become the new, second Dependent Variable (DV II). DV II will be tested through the Independent Variables Sources (V.4.1) and Path Dependence (V.4.2). Drawing from the subsequent conceptual considerations, the thesis' Hypotheses will be outlined in V.5.

V.1 Dependent Variable I – Political Bias

The dependent variable is the COM's potential political bias as indicated by the level of consistency of the RoL evaluations in the respective case studies across the selected indices (European Commission, Freedom House, Bertelsmann).

In preparation of EU accession, candidate countries have to fulfill the Copenhagen Criteria as outlined in more detail in the Literature Review (cf. Chapter II.2.4). Their progress is being monitored and evaluated annually by the European Commission through Progress Reports. All criteria are subsumed and addressed in respective chapters. The Commission decides upon their opening, negotiation and closing. All chapters have to be closed in order for a candidate country to become viable for accession. However, the European Commission has been repeatedly criticized for the inadequacy of said Reports in correctly evaluating de facto developments. Accusations of tempering with results for political reasons have not only been voiced by stakeholders within the accession process - usually candidate countries that feel as if they have been treated unfairly - but increasingly in academia as well (cf. Chapter II.4.1). If

the COM's Reports should differ significantly in their findings from FH and BTI Reports this could be explained through political bias on the Commission's part vis-à-vis the respective candidate country.⁶⁸

Political bias generally refers to a “partial perspective that is often accompanied by the refusal to consider the possible merits of alternative points of view” (Comer/Gould, 2010). For the purpose of this thesis, political bias will be defined as the extent to which the differential evaluation of candidate countries conveys a sense of sympathy or lack thereof by the evaluating institution. Differential evaluation refers to the difference in scores attested to the respective case study's RoL development by the COM versus Freedom House and Bertelsmann. A sense of sympathy or lack thereof is understood as a better or respectively worse evaluation of a specific candidate country by the COM when compared to those of Freedom House or Bertelsmann which corresponds to the general attitude of the European Union concerning said candidate country's accession.

Given the broadness of the definition and the lack of a theoretical background within political science, other fields of study were examined for correlating concepts. In sociology, cultural bias refers to the interpretation and judgment of phenomena in terms particular to one's own culture which result in more preferable judgments of phenomena closer to one's own culture and vice versa (Reynolds, 2000). In empirical research (natural sciences), a reporting bias explains the tendency to under-report experimental results that in relation to the researcher's expected findings were undesired (Fischer/Verrecchia, 2000). In psychology, the concept of political ideology was introduced which examines how a normative set of beliefs regarding the “proper order of society” (Erikson/Tedin, 2003) and its achievability is created. The latter concept will be used for the analysis because of its conceptual closeness to the research question and its thick theoretical backing which enables specific operationalization and thus, testing of the variable at hand.

Political bias can origin from and manifest itself in different aspects. First, it can be rooted in the evaluator's interests and/or value/belief system. The EU may have particular strategic interests in including a country into the Union or keeping it out. Those are usually related to

⁶⁸ At the same time, very similar evaluations of all three indices could also be caused by a similar underlying political bias but this is more unlikely given that all three indices have different aims and that FH and BTI are neither direct nor indirect stakeholders in the EU accession process and economically independent from EU funds. However, similar findings could be attributed to similar underlying values (e.g. Western liberal democracy concept) which – as will be explained above – do not constitute political bias.

security aspects (conflict resolution, peacebuilding, state building, crime prevention etc.), economic interests (expanding markets) or natural resources (gas, coal, earth metals). A political bias can also stem from socialization. Countries whose norms, customs and ideologies are more similar to the ones of the EU (or the analyst's home country) may get more favorable evaluations than those who are more different. In the same manner the cultural and historical ties (or the lack thereof) between the EU (or the analyst's home country) and the accession state may influence his perceptions and assessments and can lead to positive or negative predispositions. Finally, media coverage can influence his opinion, e.g. through public opinion polls or otherwise published opinions.

Secondly, it can appear on the institutional and/or personal level. Since a group of analysts is made up of individuals whose assessments are influenced (sub-) consciously by their own ideas, experiences and beliefs are likely to be inserted into the product of the institution's evaluation. Institutional and individual rationales can either overlap, thus strengthening the outcome, or diverge, in which case it needs to be assessed which one was decisive in shaping the result. Consequently, one needs to take into account the ways the institution's and individuals' thinking could be affected. Four aspects come to mind: a) specific strategic interests regarding (non-) accession, b) socialization towards the accession state, c) cultural/historical ties with it, and d) media coverage thereof. The functioning of the aspects just outlined has been explained in depth in the fields of psychology and sociology⁶⁹ and shall therefore not be reiterated at this point.

V.2 Independent Variables for Political Bias

If the COM's Progress Reports' results differ significantly from those of BTI and FH the independent variables reputation and incentivization have to be tested.

V.2.1 Reputation

BTI and FH also have to take into consideration their reputation as credible, independent think tanks that produce academically valid Reports given that they are being used as guides for international development aid and policies. However, the COM's reputation has a much stronger political relevance which influences its position and power within international politics. If and how new Member States are being accepted into the European Union does

⁶⁹ For example by Doucouliagos (1995), Abreu (2001), and O'Brien (2009).

affect gravely how other international actors react towards the Union politically and economically.

Reputation, according to Boulding is the “product of messages received in the past” which is based on the “total cognitive, affective, and evaluative structure of the behavior unit, or its internal view of itself and its universe” (Boulding, 1985: 120). It ultimately creates a “shadow of the future” (Axelrod, 1984), that is, expectations regarding the future behavior of an actor. Thus, an actor has an image of himself while further images of him persist in the minds of others. Herrmann, focusing on the latter, defines this perception as a “foreign policy maker’s construction of reality” (1997: 404), addressing two issues. First, image and reputation can become reality in terms of interaction between states. This means that states will draw inferences from other states’ behaviors in the past to make predictions for their future behavior. Those predictions, or “strategic judgments“ (Ibid.: 422), will be made regarding the actor’s relative capability, culture, and chance of being either threat or opportunity to oneself. Secondly, it remains unclear what part of the decision maker’s world view is salient in creating his image of another actor, thus hinting at the behavioral aspects behind decision making and the construction of reality. Although in the past Machiavelli⁷⁰ considered a bad reputation – e.g. the violent resolution of conflicts - more desirable and powerful than a good one, today’s successful politics are based on the reputation for honest diplomacy because it enhances a state’s chance of solving future conflicts peace- rather than forcefully (Anne Sartori, 2005). A more tangible definition is provided by Gibler following Sartori’s understanding. According to him “a leader’s behavior during opportunities to honor an alliance commitment creates reputation for that leader regarding his or her likelihood of honoring future alliance commitments” (2008: 433). Downs and Jones add another aspect, namely “the extent to which a state is considered to be an honorable member of the international community” (2002: 96). Keohane⁷¹ points to the fact that although reputation “operates outside the boundaries of formal treaty law” and is sometimes deemed “cheap talk”, it can considerably affect the effectiveness of international bargaining and diplomatic

⁷⁰ Niccolò Machiavelli (The Prince, 1532) wrote: “*Upon this a question arises: whether it be better to be loved than feared or feared than loved? It may be answered that one should wish to be both, but, because it is difficult to unite them in one person, it is much safer to be feared than loved, when, of the two, either must be dispensed with.*”

exchanges (Crescenzi et.al., 2012: 264). Keohane also presented the most sophisticated theoretical framework on how reputation ties into international relations. The ideas actors form about one another depend on the information and potential profits they have from cooperating with others. This information in combination with the beliefs regarding cooperation create a reputation/image in the mind of the actor seeking cooperation. Knowing this, actors actively cultivate their reputations in order to become desirable partners. Based on shared interests, actors become members of (international) institutions which shape norms and ultimately - thus closing the circle –affect the member’s interests, perceptions of and behavior towards others (Keohane, 1999: 378). Regarding international institutions, Boulding emphasizes that their behavior is determined by decisions which in theory should be made based on rational considerations of facts. However, in practice, Boulding stresses that “people whose decisions determine the policies and actions of nations do not respond to “objective” facts ...but to their “image” of the situation. It is what we think the world is like, that determines our behavior” (1985: 120).

With view to cooperation between actors, reliability is key. States choose partners carefully, preferring those who have honored their commitments in the past, thus creating positive expectations for the future (Crescenzi et. al., 2012: 260). Because of that, states care about their reputation to be reliable partners wherefore they usually “comply with international obligations”. In fact, the fear of gaining a ‘bad’ reputation upon defection which results in future cooperation is considered the main reason behind compliance in international politics (Downs/Jones, 2002: 96-100).⁷² In order to establish the credibility of their (potential) partner, states are very attentive to the signals other actors send to their respective alliance partners. As explained earlier, a ‘good’ reputation help finding partners where as ‘bad’ one deters cooperation. However, it remains to be asked who profits or suffers from reputation when and in what areas. Boulding (1985) emphasizes that powerful actors’ images and reputations matter more than those of powerless ones. This is particularly true when it comes to cooperation within international institutions, such as to “regulate the exchange of private or

⁷² However, states de facto have different levels of reliability depending on the respective agreement⁷². Once a state learns that a partner may be reliable in trade agreements while defecting in environmental questions, it reasons that reliability is weakly correlated among treaties. Therefore, even after defection, the image of one another must not necessarily be readjusted. Instead, it follows that states can and do possess “multiple or segmented reputations” (Downs/Jones, 2002: 113).

club goods where exclusion is possible such as trade agreements or alliances” (Downs/Jones, 2002: 112).⁷³

The EU is a powerful actor whose influence in the region and the world depends on perceptions of both its own people and outside actors. Within its borders, the EU needs to justify the accession process and the resources it dedicates to it. Accession is costly both economically and politically - particularly in the aftermath of the economic crisis. Consequently, the EU has two options. Either it produces success stories, thus assuring the people of the adequacy of the high investments, or it can promote itself as a responsible actor by preventing the acceptance of new members and thus save money and avoid an ‘overstretching’ of the Union’s capacities. Internationally, the EU’s draws its clout from its size, economic performance and political power. Just like on the inside it needs ‘wins’ in order remain a respected international player, that is a desirable partner for cooperation as well as an institution that too strong to be attacked.

Testing

To what extent reputation may influence the COM’s Reports will be examined through the expert interviews with DG NEAR officials by scrutinizing to what extent the Member States as represented by the Council seek the Commission to be more open to or critical of enlargement in general during the time period analyzed.

V.2.2 Incentivization

The rationale behind EU accession is conditionality which promises rewards upon the fulfillment of certain criteria. However, this form of incentivization could lead to assessments

⁷³ Both scholars have also proved that reputation matters most in international law, trade, and security and least in environmental issues and human rights (Id.). Particularly with view to economic questions “investors rely on pre-existing knowledge of other countries when evaluating countries about whom they have less information” (Gray/Hicks, 2014: 330). However, there have to persists specific circumstances in order for a “bad” reputation to have consequences. According to Keohane (1999: 376), the defection causing the “bad” reputation has to be known generally and in particular by a party “whose reactions to the violation are important to the violator”. Furthermore, the “expected costs to the violator must exceed the benefits of giving in to the conflicting temptation” (Id.). Last, a party other than the violator must have the deeds and the interest in addressing the violation. Also, thus agreeing with Machiavelli, Keohane finds that a tough reputation trumps that of compliance because both are associated with strength, respectively weakness. The question remains how the credibility and reliability of an actor (as the essence of the image/reputation) can be calculated. Crescenzi et. al. suggests two ways in which alliance seekers can calculate their potential partner’s reliability. First, by assessing their performance in upholding commitments to other states within their institution, and secondly, by evaluating the relative significance of such information “based on the similarity between the alliance seeker and the potential ally’s previous partners (2012: 263).

that are not reflective of the de facto developments in an accession state. By exaggerating the level of compliance and development the COM could give the state praise in advance in order to signal its satisfaction with the development trend and motivate it for further reforms. By understating the progress made or being very critical in its assessment the EC can voice its discontent with the development and mentally pressure the state into compliance without having to use negative conditionality. Incentivization can only be applicable to the EU Reports because FH and BTI have no direct power in influencing the case studies' accession processes.

Testing

Incentivization can be examined linguistically by checking for the use of suggestive language within the COM's Reports. Additionally, interviews with EU representatives will be used to cross-check the findings of the qualitative analysis.

V.3 Dependent Variable II – Consistency of the Indices' Reports

The initial analysis tested for political bias and produced a surprising finding: a high level of consistency among the three institutions' Reports. The origin of this consistency will be examined by testing the independent variables Resources and Path Dependence which will be defined and conceptualized below.

V.4 Independent Variables for Consistency among the Reports

V.4.1 Sources

The more diverse (re-)sources have been used in producing the Reports the more likely its results are objective. Those resources include the number of analysts, the time and money spent on the analysis, the number and range of sources that have been consulted and integrated into the analyst's Reports etc. Using similar amounts of resources or the same sources – given that the methods are similar enough - should provide similar evaluation results and vice versa. Examining the resources can also uncover potential cross references (“copying”) among the indices which in turn would explain similar evaluation results. In contrast, if cross referencing can be detected yet leading to differing evaluation results this

could strengthen the case of some of the other independent variables, particularly political bias, path dependence, reputation or incentivization through conditionality.

Testing such allegations is important for two reasons. First, accession decisions are supposedly based on the Progress Reports' evaluations.⁷⁴ Hence, if they misrepresent the actual democratic maturity level of a country, the EU either risks accepting "immature" states into its ranks or wrongfully denying access to "mature" ones. Secondly, this will have implications for the Union's credibility. The trustworthiness of the EU's accession promise is the key to the functioning of the principle of conditionality which in turn is the driving force behind democratic reforms. False Reports signal to other prospective members that one can either 'fake' one's way into the union by implementing superficial reforms or implementing reforms 'in vain' without being awarded the promised carrot of full membership. Proving claims of political bias could incentivize the European Commission to make their evaluation process more transparent. Disproving claims of political bias could help improve the credibility of the EU and thus strengthen ongoing reforms in candidate countries.

Consequently, it has to be analyzed whether political bias or other factors influence the COM's Progress Reports. As argued before, the political criteria are most relevant for accession, and negotiations are not opened unless a country has fulfilled the basic democratic requirements. Ergo, negotiations regarding the political criteria focus mainly on the RoL situation which is why it will be tested throughout this thesis.

This will be done by comparing the COM's findings to those of other institutions' measuring indices that are not stakeholders in the accession process. FH and the BTI have been selected because they both have measured the RoL developments of the case studies during the same time period and produced worded as well as numerical evaluations thereof. There are three possible outcomes: the comparison could generate equal/very, mixed or significantly deviating results. Should the comparison render very similar results, there are several possible explanations. Potentially, they are based on the same or a very similar RoL concept, use the same/very similar method or the same/similar resources. Also, they could consciously borrow from or copy each other's results or subconsciously be influenced through anchoring

⁷⁴ Supposedly because there if there are *realpolitische* reasons to include (i.e. accession for peace or controlling reasons as has been done during the 2004 enlargement) or not include a country (i.e. unresolved territorial conflicts with Member States as has been the case with Turkey over Macedonia) this is likely to trump the fulfillment of the Acquis thus enabling the political bias perception in the first place.

processes when consulting other sources' materials. Should the comparison generate (significant) deviations, there are also multiple reasons. As suspected by previous scholars on the topic, a political bias could cause them. Employing different RoL concepts and using other methods and resources could also be causal. Anchoring processes with regard to own previous evaluations, the wish to uphold certain images and the goal to use the Reports as incentivization tools may also explain differences in evaluations. If the analysis delivers mixed results, all of the variables introduced above can be examined for explanations.

V.4.2 Path Dependence

The COM's Progress Reports (just like the benchmark Reports of FH and BTI) are not produced in a mental vacuum. In order to establish continuity with view to both methods and results, previous Reports are being taken into consideration when making the current assessment. This includes own and other Reports which can cause an "anchoring effect". This effect has been described in cognitive science and proves that people are being unconsciously influenced by any available environmental information when choosing numerical values – whether it is related to the topic at hand or not. The evaluations that have been assigned to a certain status quo or development are therefore likely to be reproduced in the following years without necessarily reflecting on their adequacy. In fact, even reading the weather report in the newspaper before conducting evaluations can alter results in case of numerical attributions (Joslyn et.al., 2011). This effect does not only concern numbers which would be particularly relevant for FH and BTI Reports given their scaled evaluations. In fact, it applies to a large range of information that are falsely (and subconsciously) being included in a decision-making process.

Anchoring Effect

The anchoring concept originates from the cognitive sciences and describes the general process of how previous decisions inform present and future ones - independent of their actual relevance to the situation at hand. The anchoring effect is considered to be a cognitive bias which describes the human tendency to rely too heavily on the first pieces of information offered which then set an anchor to his decisions. A similar effect has been described in

sociology (“imprinting”) and economics (“path dependence”).⁷⁵ The latter concept has been advanced to become applicable to political science which is why – out of the three suggested, related theories – it will be used to operationalize and test for the anchoring effect.

Path Dependence

The concept of path dependence has been introduced by the economists W. Brian Arthur (1989, 1994) and Paul A. David (1985, 2000). It describes a situation where actors stick to a decision that has been made in the past although other, better (more efficient) options are available. Arthur explained the choice of less efficient options through “increasing returns” which means that the usage of the technology in question (both theorists focused on the development of technology) increases the profits in a self-energizing manner.⁷⁶

Douglas North (1990) advanced the path dependence theory from an economic principle to a theory explaining institutional change which is how it became relevant to social sciences. North acknowledged increasing returns to be a necessary but not sufficient condition for path dependence. He then introduced three sufficient conditions: the imperfection of the market, the existence of transaction costs as well as the bounded rationality of actors. Using these criteria, North’s analysis regarding the concept of path dependence differs from that of Arthur and David: he believes that a lock-out of alternative decisions is not given but that the

⁷⁵ The concept of imprinting „postulates that either initial cognitive schemes, competences, and so forth – of a founding entrepreneur or team, for instance – or specific contextual conditions (...) at the time of the founding imprint organizational processes at later stages, and eventually amount to a replicated pattern“ (Sydow et.al., 2009: 696). The difference between path dependence and imprinting is timing: imprinting postulates a pattern that is clear from the beginning whereas path dependence assumes the pattern to emerge during the process.

⁷⁶ He introduced four reasons for this process: high set-up costs, learning effects (improvement of the product and process will lower the production costs), coordination effects (cooperation of different actors will lower production costs), and adaptive expectations (if one actor expects another to act a certain way, he tends to follow suit). Furthermore, he stated that in the beginning it is relative unclear which option will become prevalent but once a certain equilibrium has been reached, this becomes reasonably stable and eventually lead to a “lock-in”. Small events or contingencies can thus have a big effect on what technology becomes market-leading. David adds to Arthur’s concept that decisions may be uphold even long after the initial reason for the decision has vanished, thus emphasizing the inefficiency criteria of path dependence. To Arthur’s increasing returns-causality David adds three more factors that stabilize path dependence: system scale economies, technical interrelatedness, and quasi-irreversibility of investment (1986: 41ff). The first criterion de facto equals the increasing return concept. Technical interrelatedness was coined based on the QWERTY-example⁷⁶ used for his analysis and states that the benefit of using a certain technology increases if the keyboard layout and the learning background of the user are consistent. Quasi-irreversibility of investment addresses the potential costs of changing from one decision to another which at a certain point become too high to make the transition. This criterion is based on sequences: if little investment has been made before, transition is more likely than as if high investments have been made already.

transition costs to a better alternative are merely (much) higher which is likely to hinder a path dissolution.

The introduced conceptual evolution of path dependence always concentrated on the aspect of efficiency thus explaining the long-term stability of different successful institutions as well as the dominance of their products in spite of their sub-optimality (Beyer, 2005: 10). However, according to Beckert (1996) the efficiency criteria is usually no benchmark for (political) actors because the impact of power, norms, values, traditions, bounded rationality, and function logic is much higher (Beyer, 2005: 10). Consequently, the concept had to be adapted to become functional for social/political sciences.

Path Dependence in Political Science

Paul Pierson (2000) picked up the concept and applied it to institutions, agreeing with Arthur's criteria: once an institution is created there will be actors who are interested in its persistence. High set-up costs, learning and coordination effects set in which increase the likelihood of path dependence. Moreover, since there are no traditional market mechanism at work for (political) institutions there is less pressure to change which in turn reinforces path dependence. The relevance of path dependence to politics becomes obvious when taking into account the importance of collective actions, the high number of institutions including many asymmetrical power relations as well as a high level of complexity that persists in political institutions (Pierson, 2000: 257). According to Pierson, lock-in can be omitted and paths can end when there are either sufficient backlashes or the amplification of the currently taken path is sufficiently reduced. James Mahoney (2000) builds on Pierson by suggesting four reasons for institutional reproduction: utilitarian, functional, power-based, and legitimating reasons. Utilitarian reasons exist when coordination effects and adaptive expectations cause otherwise rational actors to continue inefficient institutions. Functional reasons refer to the reproduction of an institution because of its function, meaning and importance within the greater institutional system. Power-based institutional reproduction persists if powerful actors can forcefully assert their interests against the interests of inferior actors. Legitimating reasons entail that actors feel morally obliged to continue an institution.

Within this categorization, functional and power-based reasons could become relevant for the analysis of possible deviations of the Reports' findings. Should the COM's Reports prove to be inconsistent with the findings of the other indices, this could have functional reasons, such

as adapting the Reports to getting the desired countries in while keeping the undesired ones out of the Union. Thus, the function of the COM as a gatekeeper within the accession process is fulfilled, non-dependent on the possible inefficiency of its evaluations and respective recommendation to the Council. The same logic applies to the power-based explanation. Since there persists a stark power hierarchy between the EU (COM) as the provider of a highly desired good (membership) and a candidate country as a solicitor, it would be easy for the COM to get away with flawed Reports. This mechanism becomes even stronger when certain underlying ideologies are in place. According to Pierson (2000: 260), “once established, basic outlooks on politics, ranging from ideologies to understanding of particular aspects of governments or orientations toward political groups or parties, are generally tenacious”. Therefore, if the COM – or the people creating a specific report – have formed an opinion on a country, they are more likely to abide by it and collect facts substantiating that opinion (a status quo bias, as Pierson calls it).

Testing

Specific tests for path dependence have been introduced by Pierson (2000: 263 ff.) and Sydow et.al.’s Performance and Formation Test (2009: 691ff.). Since Sydow’s test is more focused on organizational path dependence on economic institutions, Pierson’s approach has been chosen since he introduced criteria that help analyze and decide whether path dependent processes are taking place in a political environment. Those criteria are multiple equilibria, contingency, a critical role of timing and sequencing, and inertia.

Multiple equilibria signify that, initially, numerous different options are possible (e.g. evaluation of the status quo in a given country). Contingency refers to the significance of even very small events in causing “large and enduring consequences” (Ibid.) (e.g. the informal reaction of a candidate country to its evaluation). Furthermore, the timing and sequencing of events matters: whereas early events may be crucial in determining a path, later events may have little or no effect because of the level of stabilization that path has gained. Last, inertia means that “once an increasing returns process is established, positive feedback may lead to a single equilibrium” which will then be resistant to change (Ibid.).

Since path dependence requires the back reference to and perpetuation of previous results or decisions, it will also be tested to what extent previous Reports, scores and evaluations are being used for future Reports by each index.

The path dependence test can be applied to analyze both results: striking similarities and significant differences between the indices' Reports. In the former case, one could test whether one or two of the indices followed suit after an initial evaluation of the other by particularly scrutinizing timing and sequencing as well as self-reinforcing feedback mechanisms. However, if that were the case one would more likely attest the similarities to mere copying rather than path dependence. The latter can be examined by looking at the consistency among each of the indices using the criteria outlined above. Although path dependence is an independent Variable introduced alternatively to political bias, the latter can enhance the chance of creating the former as has been suggested by Pierson (2000). Consequently, both variables need to be reviewed in relation to one another.

The following diagram summarizes the research

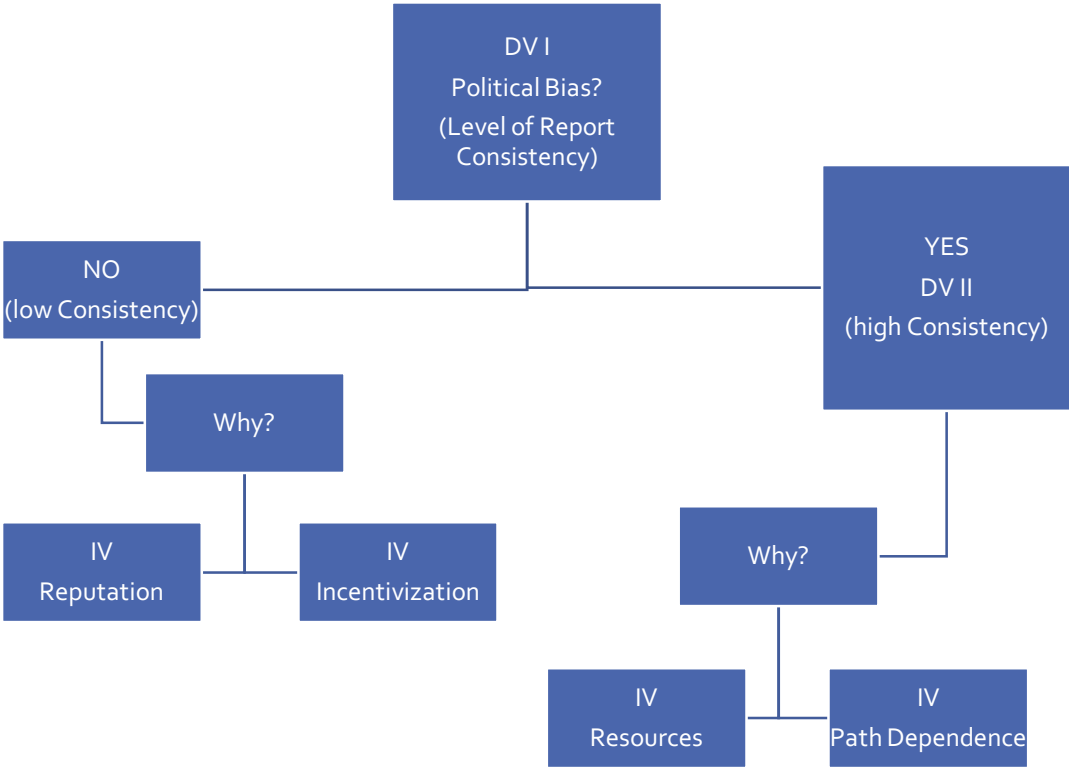


Illustration 8 - Diagram of Conceptual Framework

V.5 Hypotheses

From the Conceptual Framework introduced above, the following hypotheses can be generated:

Political Bias

Thesis: The COM's Progress Reports are politically biased towards the candidate countries as indicated by significantly inconsistent scores when compared to those of Bertelsmann and Freedom House.

Hypothesis 1: The European Commission issues politically biased Reports in order to uphold a specific reputation.

Hypothesis 2: The European Commission issues politically biased Reports to incentivize its candidates.

Level of Consistency among the Reports

Hypothesis 3: The level of report consistency depends on the usage of (re-)sources by the indices. The more overlapping resources are used, the more consistent the scores should be and vice versa.

Hypothesis 4: The level of report consistency depends on the level of inclusion of their respective earlier findings. The more the indices include their old evaluations, the more prone they are to the anchoring effect resulting in path dependence which causes a convergence of their results within their own Reports over an extended period of time. The less they do so, the more independent they remain in their yearly evaluations.

VI. Methodology

The thesis aims to answer one question: Are the COM's Progress Reports politically biased or not? The indicator for political bias is the inconsistency of the COM's evaluation results with those of FH and BTI. If political bias can be ruled out based on consistent results, three other follow-up questions arise: First, are there better indicators to measure political bias? Secondly, why can political bias not be substantiated? Third, why are the indices' results consistent? The latter two questions will be examined as part of the thesis. The research design consists of three interlocked methodological steps.

First, the COM's political bias (Dependent Variable I, as indicated by its' inconsistent results with FH and BTI) will be tested using the qualitative content analysis (VI.1). Secondly, the results of the analysis will be compared to and amended by information gathered in semi-guided, explorative expert interviews with authors and coordinators of the Reports (VI.2). The experts' opinions on the political bias claim as well as their explanations for the high level of consistency among the reports (Dependent Variable II) will also be examined. Third, the interviewed COM personnel referred to internal COM documents pertinent to the Progress Report creation process (VI.3). These have been accessed and reviewed in order to examine and explain both the political bias claim as well as the high level of Report consistency among the indices.

VI.1 Qualitative Content Analysis

Methodologically, the Reports of the COM, FH and BTI are too different to be compared in their original versions. The COM's Reports are qualitative, worded accounts of a country's developments whereas FH and BTI present qualitative and quantitative evaluations (numerical scales) of their case studies. Additionally, they examine different RoL elements. Consequently, in order to conduct a comparison, they have to be made comparable to one another.

Three methodological approaches were considered before choosing the one described in the following paragraphs: phenomenology, grounded theory and qualitative content analysis. Phenomenology is a philosophical method aimed at examining a phenomenon (in this case: the political bias claim) by systematically reflecting on and studying "the structures of consciousness and the phenomena that appear in acts of consciousness" (Menon et.al., 2014).

Grounded Theory is a method from the social sciences that aims to systematically examine qualitative data in order to generate theories on a phenomenon (Strübing, 2004). Qualitative content analysis is a tool from the empirical social sciences whose goal is to organize and structure “manifested and latent content” (Rainer, 1992: 402) through the systematic classification process of coding and identifying themes or patterns” (Hsieh, 2005:1278). The basic material used for the qualitative content analysis are transcribed interviews, articles, Reports, pictures or video footage. Although political bias may be categorized as a phenomenon that appears in the conscious mind of those authoring and/or consuming the reports, the phenomenological approach is not suited to analyze the consistency among the Reports adequately because the “acts of consciousness” of the reports’ authors cannot be measured reliably. Grounded Theory is not a suitable choice either because theories on the phenomenon (reasons for consistent or inconsistent Reports, including political bias) have already been generated before entering the analysis and will be tested after the comparison’s outcome. Given the material that is to be tested (written, linguistic Reports) as well as the goal of testing it (creating a common structure – scale – on which all Reports can be evaluated to make them comparable) the qualitative content analysis is the best methodological fit for the research goal.

As Weber (1990) has described and as has been recited by Hsieh (Ibid.), the qualitative content analysis “goes beyond merely counting words to examining language intensely for the purpose of classifying large amounts of text into an efficient number of categories that represent similar meanings”. The qualitative content analysis is sub-divided into three approaches: conventional, directed, and summative. The conventional approach gathers coding categories directly from the textual data. The directed approach derives initial codes from a theory before entering the analysis. The summative approach counts and compares the appearance of keywords or content units and then interprets their underlying context (Hsieh, 2005). For this thesis, both the conventional approach has been chosen. The conventional approach looks for the RoL elements that a) exist and are b) shared by the three indices because it only makes sense to compare the elements that are included in all or at least the majority of the Reports. This approach has led to Mayring’s scaled, structuring qualitative content analysis (SSQCA) which does exactly that: create one, unified quantitative scale on which qualitative data from different sources can be measured.

Mayring's Scaled, Structuring, Qualitative Content Analysis (SSQCA)

Aim

The SSQCA is used to understand which elements constitute the RoL in each of the three indices and filter out the RoL elements they all share. Based on the shared RoL elements the SSQCA will be used to analyze whether similar or different scores have been generated by each index for each element for the respective case studies Croatia and Turkey from 2005 until 2012. Thus, the consistency of the COM's Progress Reports with those of FH and BTI is assessed and the political bias question can be answered.

The SSQCA generates data for the following working questions:

1. Which RoL elements are represented and tested in the indices?
2. Which RoL elements are shared by all indices?
3. How have the COM, FH and the BTI evaluated Croatia's overall RoL from 2005 to 2012?
 1. How have they evaluated Croatia's individual RoL elements for each year?
4. How have the COM, FH and the BTI evaluated Turkey's overall RoL from 2005 to 2012?
 1. How have they evaluated Turkey's individual RoL elements for each year?
5. How consistent are the three indices' results with view to the overall RoL evaluation for each case study?
6. How consistent are the three indices' results with view to each RoL element's evaluation?

Research Design and Challenges

Mayring's scaled, structuring qualitative content analysis method will be employed to answer the questions raised above. First, the inductive qualitative content analysis will be used to examine which elements constitute each index's RoL concept and which RoL elements all three indices have in common. Secondly, based on the shared elements, the SSQCA will be conducted a second time based on the Coding Manual. This way each index's RoL element and the overall RoL score can be graded and calculated on one, common scale which makes all three indices comparable to one another. Using these scores, statements can be made about the consistency of the COM's Reports and thus, the political bias claim.

The process employed to answer the previously introduced working questions is summarized in the Table 17 below and elaborated on in the following sections:

Questions	SSQCA Approach	Specific steps
1. Which elements constitute the RoL concept of each index?	Inductive qualitative content analysis	<ol style="list-style-type: none"> 1. Identify all elements 2. Create and subsume categories
2. Which RoL elements do all three indices have in common?	Inductive qualitative content analysis	<ol style="list-style-type: none"> 1. Calculate % of RoL elements' appearance in all Reports 2. Choose those all 3 have in common 3. Use those for further analysis/comparison
<p>3. How have the COM, FH and the BTI evaluated Croatia's RoL from 2005 to 2012?</p> <p>How have they evaluated Croatia's individual RoL elements for each year?</p>	Using inductively formed RoL elements, executed based on coding manual	<ol style="list-style-type: none"> 1. Calculate each index's RoL element's score for each year. 2. Calculate each index's aggregated RoL score for every year by adding all element's scores and divide them by the number of elements.
<p>4. How have the COM, FH and the BTI evaluated Turkey's RoL from 2005 to 2012?</p> <p>How have they evaluated Turkey's individual RoL elements for each year?</p>	executed based on coding manual	<ol style="list-style-type: none"> 1. Calculate each index's RoL element's score for each year. 2. Calculate each index's aggregated RoL score for every year by adding all element's scores and divide them by the number of elements.
5. How similar/different are the three indices' results with view	SSQCA results compared on one	<ol style="list-style-type: none"> 1. Compare calculated scores in index pairs as well as across all

to the overall RoL evaluation for each case study?	scale, depicted in table and diagram	indices. 2. Analyze based on number of identical/very similar, mixed and very different results.
6. How similar/different are the three indices' results with view to each RoL element's evaluation?	SSQCA results compared on one scale, depicted in table and diagram	1. Compare calculated scores in index pairs as well as across all indices. 2. Analyze based on number of identical/very similar, mixed and very different results.

Table 17 - Overview of SSQCA Approach

Execution

First, each indices' RoL conceptualization will be tested using the inductive qualitative content analysis approach. The written Reports of each index will be scrutinized regarding their specific criteria, thus providing a clearer picture of each index' RoL understanding. This can then be used to compare all three RoL concepts to one another. The inductive approach will show which RoL elements the three indices have in common and what those elements consist of in each index. Their shared RoL elements will then be used to carry out the actual comparison using the scaled, structuring qualitative content analysis.

The inductive approach combs through the entire material and uses an open coding process to highlight anything relevant to the research question (e.g. "What constitutes the RoL concept for the COM/FH/BTI?"). The highlighted passages are then being grouped into main and sub-categories which are collected on a coding sheet. A repeated working through the material and re-organization of the categories helps to create a final set of categories. Each main and sub-category will then be defined and complemented with anchoring examples. Using Mayring's analysis software (QCAmap) it is calculated which categories are represented how often in each index. Only those categories found in each of the indices (the designated threshold of representation is at least 75%) will then be used as the underlying RoL concept for the subsequent scaled, structuring qualitative content analysis which in turn is aimed at comparing the indices' specific evaluations.

The procedure behind the inductive approach consists of eight steps. First, the material, research question, the goal of the analysis and its underlying theory need to be defined. Secondly, category definitions (selection criteria) and abstraction levels for the inductive category formation must be determined. Third, an incremental inductive category formation within the material in relation to the definition and abstraction level has to be carried out. Subsequently, old categories have to subsumed into new ones, thus reducing their number and consolidating their meanings. After 10-15% of the material, the categories have to be revised and a formative reliability test (evaluation and improvement of process) has to be employed. Afterwards, the entire material will be reviewed based on the revised categories. Then, a summative reliability test will be carried out. Lastly, the entire material will be assessed and the main categories will be formed by conducting a quantitative analysis. The process is illustrated in the graphics below:

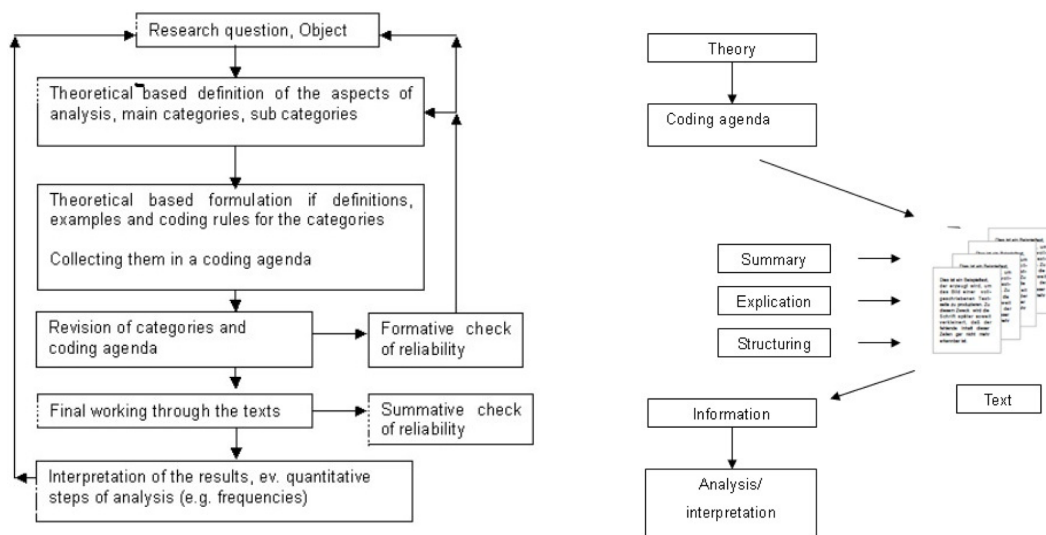


Illustration 9 - Mayring's Qualitative Content Analysis (2000)

Material, research question, goal of analysis and underlying theory

The material used are all text passages of the three indices' Reports that evaluate or assess the case study's RoL status quo or development for each year between 2005 and 2012. The Reports comprise of the European Commission's Progress Reports, Freedom House's Freedom in the World analyses and Bertelsmann's Bertelsmann Transformation Index. Specifically, this includes Chapter 23 of the European Commission's Progress Reports on the "Judiciary and Fundamental Rights", Freedom House's "Political Rights and Civil Liberties"

section as well as Bertelsmann's "Rule of Law" section (which is the third sub-component of its "Democracy" chapter). All Reports cover the development of both case studies, Croatia and Turkey, from 2005 (the beginning of their accession negotiations) until 2012 (the acceptance of Croatia as a new EU member state). Because the Bertelsmann Reports are only issued every two years, said Reports only exist for the even years (2006-2012). Thus, a total of 40 Reports is being analyzed using the QCA method.

The research question reads: What are the individual RoL elements each index uses to evaluate the case study's RoL status quo/development for the respective year? Consequently, the goal of the analysis is to identify the RoL elements of each index. Given the backgrounds of the indices at hand, those elements should be consistent with a Western liberal RoL concept as outlined in II.1.3, and thus, similar to one another.

Category definitions, selection criteria and abstraction levels

The main category is the RoL concept and its sub-components – its individual constitutive elements which are used to evaluate the case studies' RoL status quo and/or development. The abstraction level is:

- the concrete mentioning and/or description of the RoL-relevant categories or their specific aspects with view to the case study's status quo and/or development
- the citing and/or evaluation of RoL-elements related policies

Regarding the analytical units, the minimal coding units comprise of meaningful phrases and the maximum context unit of the entire text on the respective country's RoL status quo/development within the given year. The assessment unit consists of the complete text on the RoL situation (with multiple codings for different aspects, but no multiple mentionings).

Subsuming and revisions of categories and main category formation

In the first category formation phase, all possible categories will be identified. The second category formation phase then summarizes the categories from the first phase into topics on a higher abstraction level, thus leading to the final categories. Those categories will be defined by their constitute elements and respective anchoring examples from the texts at hand.

Scaled, Structuring Qualitative Content Analysis (SSQCA)

The scaled, structuring qualitative content analysis approach aims to answer a specific research question (e.g. "Do COM, FH and BTI evaluate Croatia consistently between 2005 and 2012") by comparing only the RoL elements which are shared by all three indices. These

turned out to be *Judicial Independence*⁷⁷ (JI), *Efficiency*⁷⁸ (E), *Crime* (C), and *Rights & Liberties* (R&L). These elements constitute the RoL concept used for the comparison. The concept has been assembled similarly to the inductive approach: first, the text passages within all the Reports are highlighted that include statements (meaningful phrases, semantic judgments) on the four RoL categories. Then, each category has been sub-categorized and defined using examples from the individual Reports. The concept will be introduced in the following section.

RoL Concept used for comparison

Judicial Independence on the institutional level requires a self-governing body for the selection, appointment, evaluation, promotion and disciplinary proceedings. On the individual level, tenure as well as an adequate remuneration is required. With view to the budget, sufficient funds need to be granted and the self-governing body must be able to prepare its own budget as well as being consulted before its finalization. *Efficiency* can be dissected into fighting the backlog of cases through optimized delegating models as well as modern technological solutions. The *Crime* category consist of anti-corruption measures and organized crime, police and prison as well as war-crime aspects. Anti-corruption and organized crime provisions include measures taken to prevent the abuse of professional for personal gains including respective laws and institutions. Police and prison focus on the humane treatment of prisoners and war crime provisions are targeted at cooperation with war tribunals. The *Rights and Liberties* section addresses Fundamental and Human Rights as outlined by international conventions, civil liberties as outlined by the European conventions as well as Minority protections as outlined by both international and European legislation. Table 18 below summarizes the detailed aspects of each sub-category:

RoL Cat.	Sub-Categories	Specific requirements (equals full compliance)
JI	Institutional	Self-governing body for selection, appointment, evaluation, promotion, disciplinary proceedings consisting of: <ul style="list-style-type: none"> • Majority of judges selected by peers • Minority of legal scholars appointed by Parliament • Independence of Ministry of Justice

⁷⁷ Judicial Independence is defined as the freedom from undue influences from anyone, decisions made solely based on the law and professional interpretation thereof.

⁷⁸ Efficiency is defined as the optimal usage of infrastructure and resources.

		<ul style="list-style-type: none"> • In charge of admin. tasks • Objective selection criteria • Clear rules against nepotism/patronage • Non-discriminatory
	Individual	<ul style="list-style-type: none"> • Adequate remuneration • Tenure
	Budgetary	<ul style="list-style-type: none"> • Right to prepare budget • Rights to be consulted prior to finalization • Sufficient funds to carry out functions (2-4% of GDP)
E	Fighting Backlog	<ul style="list-style-type: none"> • Speedy proceedings to clear old/ avoid new backlog • Delegation of non-judicial work to other institutions
	Computerization	<ul style="list-style-type: none"> • Integrated Court Management System
C	Anti-Corruption & Organized Crime	<p>Measures taken that prevent the abuse of professional power for personal gain including:</p> <ul style="list-style-type: none"> • Ratification and implementation of relevant Conventions/Protocols/domestic legal framework • Code of conduct/ethics • Reporting • Avoidance/disclosure of a conflict of interest • Refusal of gifts/favors • Effective, efficient and economical use of public and official resources
	Police & Prison	<ul style="list-style-type: none"> • Respect for human rights (no violence, no overcrowding) • Adequate staff and capacities/resources • Adequate health conditions
	War Crimes	<ul style="list-style-type: none"> • Cooperation with War Tribunals
R & L	Fundamental/ Human Rights	<ul style="list-style-type: none"> • Prohibition of torture/inhuman/degrading treatment • Freedom of thought/conscious/religion • Freedom of expression
	Civil Liberties	<ul style="list-style-type: none"> • Freedom of assembly and association • Access to justice/legal aid • EU citizen's rights • Right to education
	Minorities	<ul style="list-style-type: none"> • Protection of minorities (religious, ethnic, socially vulnerable, disabled, gender, age) • Anti-discrimination law enforcement • Protection of cultural rights

Table 18 - Specific elements of RoL sub-components, gathered through initial qualitative content analysis of COM, FH and BTI Reports (2005-2012)

Coding Manual

In order to generate scores based on the RoL concept outlined above, a coding manual has

been created. This manual consists of categories, definitions, anchoring examples, coding rules and a numerical score that corresponds to each level. The categories represent the level of compliance with the specific requirements outlined in the table above, e.g. “fulfillment”, “general functionality” or “regress”. The definitions outline when a certain level of compliance is met, e.g. “fulfillment” applies if there is a full compliance with the requirements which means that the necessary institutions are in place and rules are being fully implemented. Anchoring examples provide textual examples from each of the indices’ Reports which correspond with the the definition provided. Coding rules provide the reader/other coder with the generalized rules as to what linguistic markers indicate the attribution of a certain level of compliance, e.g. “fulfillment” can be attributed if the language of the Reports is characterized by invariably positive assessment or the usage of superlative, positive grade particles such as “substantial” or “fully”. The numerical score awarded for a certain level of compliance represents on ordinal scale from 1-5, with 1 point indicating “regress” and 5 points demonstrating “fulfillment” of the RoL requirements. Within each category, the respective sub-categories (e.g. Crime: anti-corruption, war crimes, police, prison, organized crime) will be scored according to the coding manual. The total sum of the sub-categories’ scores will be divided by the number of sub-categories to create the category’s final score. For each report, there will be four scores corresponding to each of the RoL elements tested as well as one final score which is the sum of categorical scores divided by the number of categories addressed within the report. This calculation method was chosen based on the fact that most indices (including FH and Bertelsmann) weigh each RoL element/question equally, although in reality one may argue that specific elements are more influential in a country’s RoL functionality than others.

Table 19 below summarizes the coding manual used for the deductive qualitative content analysis:

Category	Definition	Anchoring examples	Coding rule	#
Fulfillment	Institutions in place and being implemented Full compliance with requirements/goals	<u>EC</u> : “substantial progress” “no problems have been reported in the implementation of these provisions”	Invariably positive assessment Superlative,	5

		<p><u>FH:</u> “death penalty fully abolished”</p>	<p>positive grade particles (<i>substantial, fully</i>).</p>	
<p>General Functionality</p>	<p>Institutions in place and being implemented</p> <p>Working well but no complete compliance</p>	<p><u>EC:</u> “good progress”</p> <p>“generally satisfactory, although further improvements can be made”</p> <p>“no indications of systemic problems”</p> <p><u>FH:</u> “generally meet acceptable international standards”</p> <p>“implementation has improved”</p> <p>“recent reforms had...positive effect”</p> <p><u>BTI:</u> “clearly indicates a move toward increased credibility and signals an obvious improvement”</p> <p>“generally respected”</p> <p>“notable progress”</p> <p>“have largely been achieved”</p>	<p>Positive trend (<i>increased, improved</i>) assessment with semantic restrictions (<i>largely, generally</i>).</p>	4
<p>Limited Functionality</p>	<p>Institutions in place</p> <p>No or insufficient implementation</p>	<p><u>EC:</u> “limited progress/some developments”</p> <p>“Considerable progress on the legislative front, but this must now be matched with pro-active enforcement”</p> <p>“several successful measures have been introduced...still a significant problem”</p> <p><u>FH:</u> “do not fully meet international standards”</p> <p>“inadequate implementation”</p> <p>“established a variety of bodies...., but so far, impact has been limited”</p> <p><u>BTI:</u> “relatively independently”</p>	<p>Functional. assessment always qualified through restriction.</p> <p>Restrictive adverbs (<i>limited, relatively</i>)</p> <p>Modal particles (<i>but, still, however, although</i>)</p>	3

		<p>“still face major difficulties”</p> <p>“limited reaction”</p> <p>“impact not yet evident”</p>		
Non-Functionality	<p>Required institutions non-existent</p> <p>Required means not provided</p>	<p><u>EC</u>:</p> <p>“no progress”</p> <p>“needs to be urgently reformed”</p> <p>“need to be reviewed...lacking the appropriate competence and experience”</p> <p><u>FH</u>:</p> <p>“neglect”</p> <p>“legal impediments ...remain”</p> <p>“other groups lack legal status and...are subject to legal challenges”</p> <p><u>BTI</u>:</p> <p>“continues to constrain their effective operation”</p> <p>“still far from EU standards”</p> <p>“insufficiently competent”</p>	<p>Invariably negative assessment.</p> <p>Superlative, negative grade particles (<i>no, lacking</i>)</p> <p>Usage of conditional language (<i>needs to, should, must</i>).</p>	2
Regrets	<p>Status quo regarding requirements/goals has deteriorated since the last evaluation</p>	<p><u>EC</u>:</p> <p>“average criminal trial period increased”</p> <p>“The growth of the prison population is leading to serious overcrowding, which is hampering attempts to improve detention conditions. A complete overhaul of the...system is needed.”</p> <p><u>FH</u>:</p> <p>“Transparency has improved...government seemed to be reversing that trend”</p> <p>“controversial bill adopted...introduces interview...which...could make space for further executive interference”</p> <p>“scandal revealed official wiretapping of judges...accusations of political interference”</p>	<p>Negative trend signaled through comparative adjectives (<i>deteriorated, increased, decreased</i>).</p>	1

		<u>BTI:</u> “deteriorated” “violations...continued to increase ” “complaints.... doubled ”		
--	--	---------------------------------------------------------------------------------------------------------------------------	--	--

Table 19 - Coding Manual

*Institutions refers to policies (laws, regulations, guidelines) as well as to physical institutions (committees, bodies in charge of implementation) and resources (financial and materialistic granted for implementation)

Display & Analysis of Results

The results for questions one through four (RoL elements and shared RoL elements, RoL evaluation for Croatia and Turkey) will be provided in respective tables in the **VII. Results** Chapter. Their consistency (questions five and six) will be tested first in pairs (COM + FH, COM + BTI, FH + BTI) and then across all three indices. Their comparison is based on the following cut-off points on the scale of 1 through 5 points:

- ‘Consistent results’: 0-0,3 points deviation
- ‘Mixed results’: 0,4-0,9 points deviation
- ‘Inconsistent results’: deviation exceeds one point

This differentiation and attribution is based on the fact that each point represents a completely different level (quality) of compliance with the respective RoL requirements. Consequently, a 1-point deviation automatically places an evaluation result on another score level and thus, produces “inconsistent results”. Given the subjective element introduced by human authors (coders) for each index, a minimal deviation must be allowed for the slightly different interpretation of otherwise ‘consistent results’ which is why the 0,3 deviation was introduced. Given that a 0,4-0,9 deviation may still keep an individual evaluation within the same level of compliance while demonstrating a potentially very different quality, the intermediate deviation range was termed “mixed results”.

After counting the pairs and overlapping evaluations of all three indices, conclusions can be drawn as to the actual level of consistency between the COM’s Progress Reports and those of the other two indices⁷⁹.

The findings of said comparison will then be interpreted and discussed based on the conceptual framework as well as the semi-structured, in-depth expert interviews with representatives of each index. The methodological approach of the latter will subsequently be explained.

VI.II Expert Interviews

In order to explain the outcome of the analysis, the second tool used are explorative, in-depth, semi-structured expert interviews with representatives of all three indices who took part in creating the analyzed Reports. EU officials have been asked to explain the Progress Report creation process as this has not become public knowledge yet. During the interviews, a number of internal documents were referred to which were later accessed and included into the analysis and explanation of the comparison's results (documents included into the Annex). Furthermore, all interviewees – including FH and Bertelsmann officials – have been asked to give their opinions on three aspects. First, they were asked to present their ideas on why scores across indices could either be very similar or very different. Then, they shared their opinions on the relevance of the independent variables from the theoretical framework in order to examine their agreement with theory-based explanation approaches. Last, they were asked to reflect on the political bias claim vis-à-vis the COM's Reports. Each of the analytical tools will be described in detail in the following sections.

Aim

The aim of the expert interviews is to provide the reader with contextual information and explanations concerning the results of the qualitative content analysis and the comparison of index scores. First, specific procedural knowledge on the creation and formation of the COM's Progress Reports is generated which has not become public knowledge as of now. Understanding the COM's methodological approach is important in order to compare it to those of the other indices tested. Secondly, the experts offer explanations for (dis-)similarities between their institution's results. Third, their reflection on the theory-based variables explaining possible (dis-)similarities is disclosed. Finally, the experts' opinions on the political bias accusation vis-à-vis the COM – which ultimately is the origin of this thesis – is gathered

Interview Research Design

Several representatives of each institution issuing the Reports were chosen who have been personally involved either in the authoring or the revision of said Reports for the respective

case studies. Interviewing an author as well as a ‘handler’ of the report ensures that both a reflection on the original evaluations as well as their processed results could be obtained. Since this handling is carried out by single individuals (regional coordinators) at FH and Bertelsmann and since the authorship is conducted by very few individuals (one at FH, two people at BTI, 10 at the COM) who usually remain anonymous, the small number of representatives per institution is sufficient to answer the questions. The topic guide for the interview as well as its specific questions are attached to the thesis in the Appendix. The interviews address the following general categories:

- I. General methodological approach of report creation process
- II. Interviewee’s personal explanation of consistent evaluation results among different indices
- III. Interviewee’s personal explanation of inconsistent evaluation results among different indices
- IV. Interviewee’s reaction to theory-based variables
- V. Interviewee’s conceptual understanding of and opinion on political bias accusation vis-à-vis the European Commission and its own index (if not COM representative)

The explorative, semi-structured, in-depth interviews lasted between one and two hours and were carried out in person or via video-conference. All interviews have been audio-taped and partially transcribed (see Appendix).

Execution

The expert interviews were aimed at providing substantive results, i.e. shedding light on the broader (social) reality that exists outside of the data gathered through the analysis in order to provide contextual explanations. With view to the COM’s progress report creation process, the interviews have been inductive and non cross-sectional, although the underlying information have later been used to compare the COM’s methodology to that of the other indices. Regarding the political bias aspect as well as explanations for potential (in)consistencies across the Reports, the interviews have been subject to an inductive thematic cross-sectional analysis which compares and contrasts the different accounts of the interviewees. In terms of their reflections on the theory-led explanations, a deductive thematic cross-sectional analytical approach has been used.

Display & Analysis of Results

The findings of the interviews will be presented along the lines of the five aims outlined above. Explicit and implicit explanations will be offered as to why all the Reports tested demonstrate such a high level of consistency as well as how outliers can be explained.

VI.III Review of Internal COM Documents

During the Interviews with the COM's DG NEAR Enlargement officials – which asked to remain anonymous – a number of internal documents pertinent to the Progress Report creation were mentioned. These comprise of the 'Guide to the main Administrative Structures required for implementing the Acquis', which forms the basis for the subsequent 'Guidance Notes' packages the Reports' authors receive which in turn are complemented with specific guiding questions for different RoL elements (including the four elements tested in the analysis). The salience of these documents lies in the specific (linguistic and qualitative) instructions to the authors which are reflected in their evaluations. These can help explain the origin and persistence of critique issued towards the neutrality of the Progress Reports, as will be detailed in the Results and Discussion Chapter. Access to said internal documents from 2005 to 2016 has been requested and granted by the NEAR ACCDOC. The documents are included into the Appendix.

VII. Results

The thesis' methodology employed three tools: the qualitative content analysis (VII.1), the expert interviews with index authors and coordinators and the revision of internal COM documents (subsumed under VII.2) whose results will be presented in the following sections.

VII.1 Qualitative Content Analysis Results

The structuring qualitative content analysis consisted of an inductive content analysis approach to generate the indices' shared RoL elements (VII.1.1) and a subsequent scaled, structuring qualitative content analysis approach (SSQCA) that enabled the indices' scoring on one, common scale (VII.1.2). The results of both analytical steps will be presented subsequently.

VII.1.1 Inductive Qualitative Content Analysis Results

The inductive qualitative content analysis aimed to answer two questions:

- 1. Which RoL elements are represented and tested in the indices?**
- 2. Which RoL elements do all three indices have in common?**

In the first category formation phase, the following 16 categories and sub-components could be identified:

- C1: Judicial Independence (inst: selection, appointment, evaluation, promotion, disciplinary; ind: pay, tenure; budgetary: budget)
- C2: Judicial Impartiality (actual or perceived bias in judgment)
- C3: Judicial Accountability (judicial review, peer review, civil society review)
- C4: Professionalism & Competence (training standards & facilities)
- C5: Efficiency (infrastructure & resources: computerization, backlog)
- C6: Anti-Corruption
- C7: Fundamental Rights (Human Rights)
- C8: Civil Liberties (Freedom of speech, association & assembly, privacy, fair court trial etc.)
- C9: Anti-Discrimination (Minorities/gender/ethnic groups/LGBT)
- C10: Organized Crime
- C11: Checks and Balances (Separation of Powers, Judicial Review)

- C12: Equal Access to Justice (e.g. legal aid, interpreter)
- C13: War Crimes
- C14: Police & Prison Conditions
- C15: Data Protection
- C16: Alternative Dispute Resolution (ADR)

The second category formation phase summarized the categories outlined above into topics on a higher abstraction level and led to the nine final categories:

- C1: Judicial Independence (institutional: selection, appointment, evaluation, promotion, disciplinary; individual: pay, tenure; budgetary: budget)
- C2: Judicial Impartiality (actual or perceived bias in judgment)
- C3: Judicial Accountability (judicial review, peer review, civil society review)
- C4: Professionalism & Competence (training standards & facilities)
- C5: Efficiency (infrastructure & resources: computerization, backlog, ADR)
- C6: Rights & Liberties (Human Rights, Fundamental Rights, Anti-Discrimination, Protection of Minorities, Equal access to Justice)
- C7: Crime (Police and Prison Conditions, Organized crime, war crimes)
- C8: Checks and Balances (Separation of Powers, Judicial Review)
- C9: Data Protection

The inductive QCA showed that a total of nine RoL elements (categories) can be found in all of the indices' Reports⁸⁰. The frequency of each category within the total sum of Reports is depicted in Table 20 below:

Category	Category occurs in % of all Reports
Judicial Independence	90
Judicial Impartiality	57,5
Judicial Accountability	32,5
Professionalism & Competence	42,5
Efficiency	95
Rights & Liberties	100
Crime	100
Checks & Balances	17,5
Data Protection	37,5

Table 20 - appearance of RoL categories in all three indices

For the next step - the deductive qualitative content analysis (dQCA) - only RoL categories that lay in the fourth quartile (occurring in at least 75% of the Reports or more, colored above) have been used to ensure that only shared criteria will be compared. Those shared RoL elements comprise of the four categories *Judicial Independence*, *Efficiency*, *Crime*, and *Rights*

⁸⁰ This includes all three indices' Reports on both Croatia and Turkey between 2005-2012, limited to textual evidence that is related to the RoL.

& Liberties.

VII.1.2 Scaled, Structuring Qualitative Content Analysis Results

The results of the scaled, structuring qualitative content analysis based on the inductively generated RoL concept (Judicial Independence, Efficiency, Crime, Rights & Liberties), were first qualitatively interpreted and then quantitatively assessed and calculated for each of the four RoL categories. Each category of every index has been scored on a scale of 1-5 points ('regress' – 'full compliance'). The overall score for each case study's annual RoL evaluation per index was then calculated by dividing the sum of the four categories' scores by the number of categories addressed. While most Reports address all four categories, some year's Reports only refer to two or three categories which is reflected and outbalanced in the new scores.

Thus, the remaining four questions can be addressed:

- 3. How have the COM, FH and the BTI evaluated Croatia's RoL from 2005 to 2012?**
 1. How have they evaluated Croatia's individual RoL elements for each year?
- 4. How have the COM, FH and the BTI evaluated Turkey's RoL from 2005 to 2012?**
 1. How have they evaluated Turkey's individual RoL elements for each year?
- 5. How similar/different are the three indices' results with view to the overall RoL evaluation for each case study?**
- 6. How similar/different are the three indices' results with view to each RoL element's evaluation?**

The following two sections address question three and four and present the analysis' results on Croatia and Turkey.

Results for Croatia

The indices' evaluation of Croatia's RoL 2005-2012

The comparison of the overall RoL scores awarded by all three indices for Croatia's shows some instances of 'mixed results' (0,4-0,9 points deviation, colored yellow) and fewer instances of 'consistent' evaluations (0-0,3 points deviation, coloured in green). No substantially 'inconsistent results' (more than one point) could be found. The specific scores are assembled in Table 21 below as well as illustrated in Diagram 1:

Croatia	COM	FH	BTI
2005	3,2	2,7	-
2006	2,9	2,3	3,5
2007	2,9	3	-
2008	2,9	3	2,3
2009	2,8	3	-
2010	3,1	2,6	2,7
2011	3,5	3	-
2012	3,3	3,1	2,9

Table 21 - Table of Croatia's aggregated, overall RoL scores

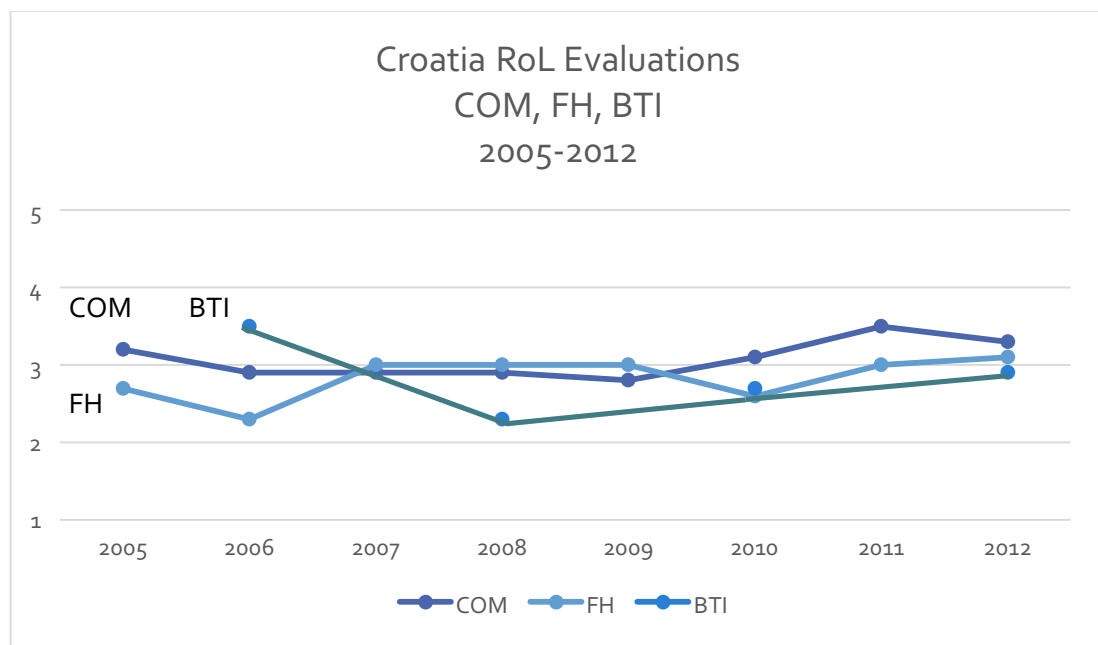


Diagram 1 - Croatia's RoL Evaluations (COM, FH, BTI)

The indices' evaluation of Croatia's Judicial Independence (RoL) 2005-2012

With view to *Judicial Independence*, the three indices cover a spectrum of evaluations ranging from 'non-' to 'limited functionality'. Particularly similar observations have been made by the COM and BTI Reports between 2007 and 2009. Significant deviations arise and persist between 2009 and 2012. Due to a lack of data for BTI in the uneven years and FH between 2006 and 2009, no authoritative conclusion can be drawn for this RoL category.

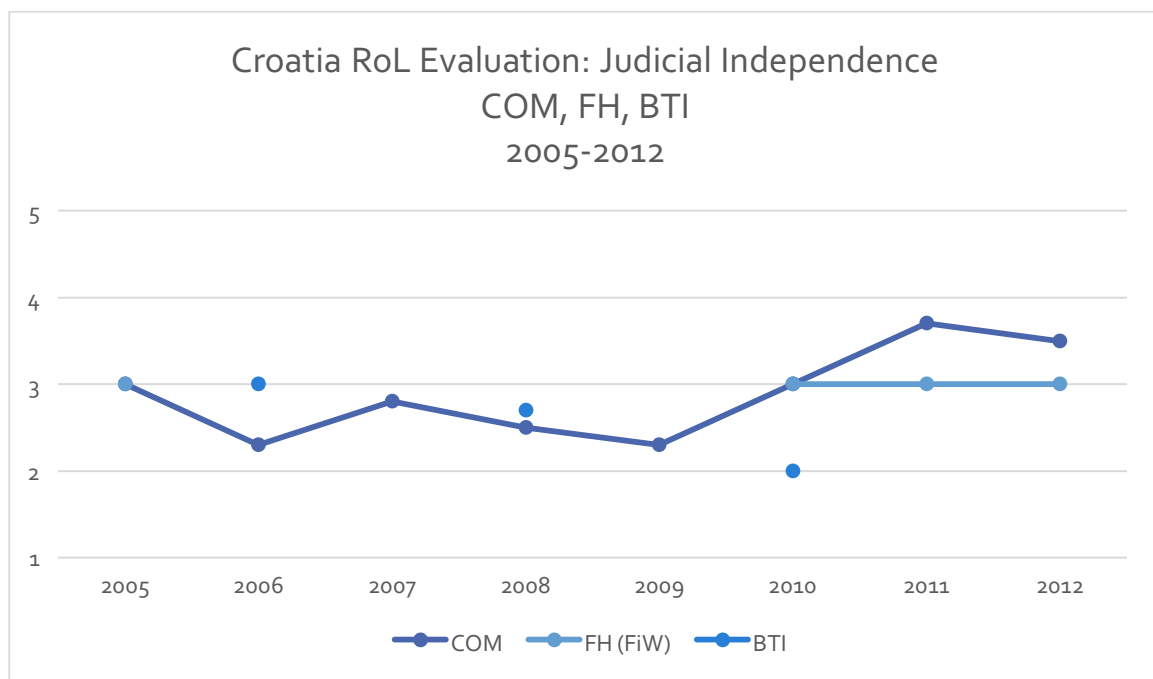


Diagram 2 - Croatia's Judicial Independence Evaluations (COM, FH, BTI)

The indices' evaluation of Croatia's Efficiency (RoL) 2005-2012

As opposed to the *Judicial Independence* category, the indices' results for *Efficiency* are more consistent, moving along the line of 'limited functionality' (with one FH-outlier showing 'regress' in 2006). Especially between 2007 and 2010, all three indices provide similar results.

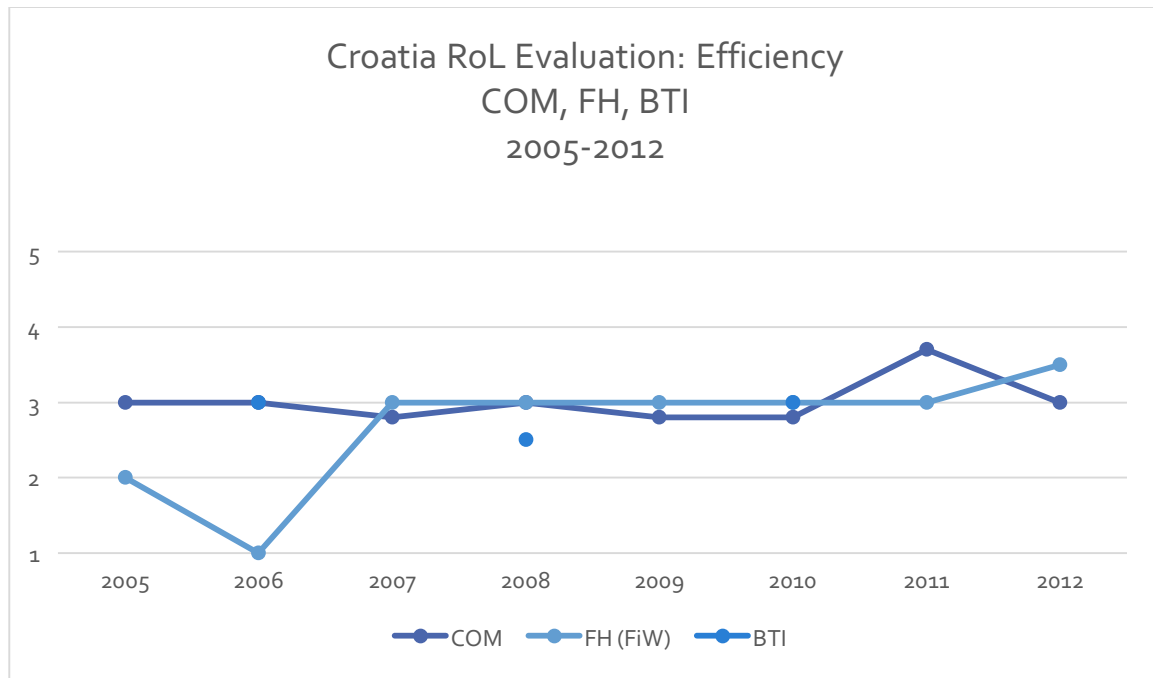


Diagram 3 - Croatia's Efficiency Evaluations (COM, FH, BTI)

The indices' evaluation of Croatia's Crime (RoL) 2005-2012

Within the *Crime* category, all three indices move around the line of 'limited functionality' with two outliers signaling either 'regress' (BTI, 2008) or 'non-functionality' (FH, 2010). Very similar evaluations can be found between 2005 and 2009 within the COM's and FH Reports, whereas the results of the BTI Reports vary significantly in 2006 and 2008 and converge towards the other indices since 2010.

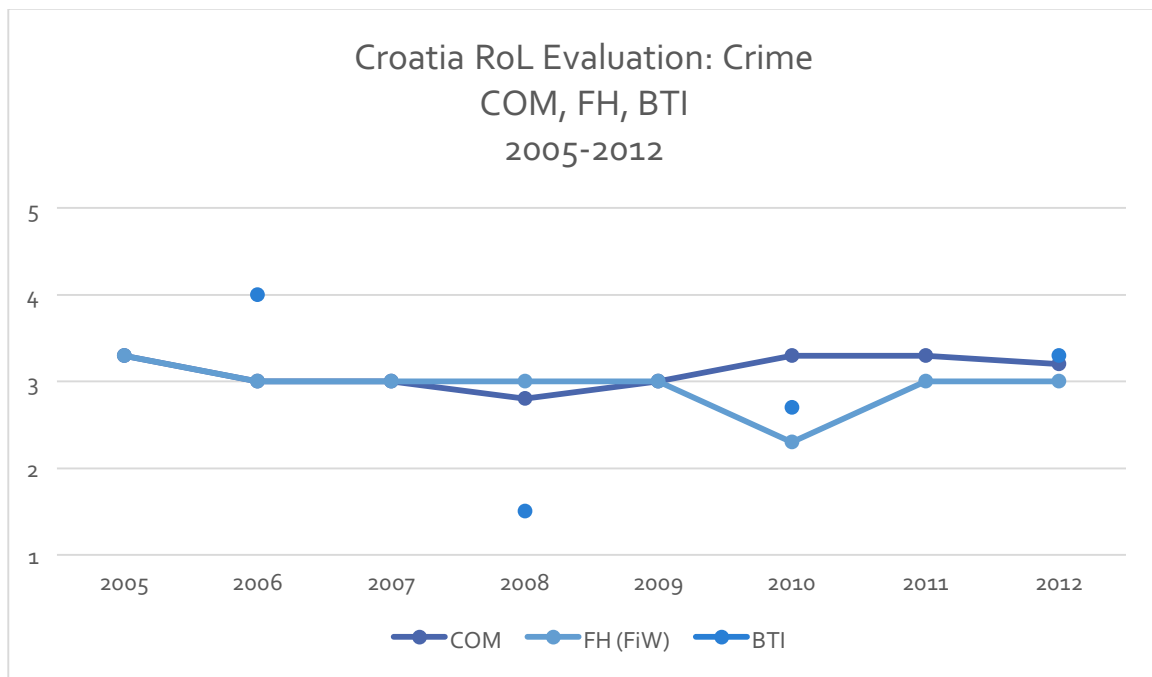


Diagram 4 - Croatia's Crime Evaluations (COM, FH, BTI)

The indices' evaluation of Croatia's Rights & Liberties (RoL) 2005-2012

The *Rights & Liberties* category provides consistent results across all three indices along the line of limited functionality. Very similar results can be found between 2006 and 2009 for COM and FH Reports, whereas BTI results show outliers in 2006, 2008 and 2012.

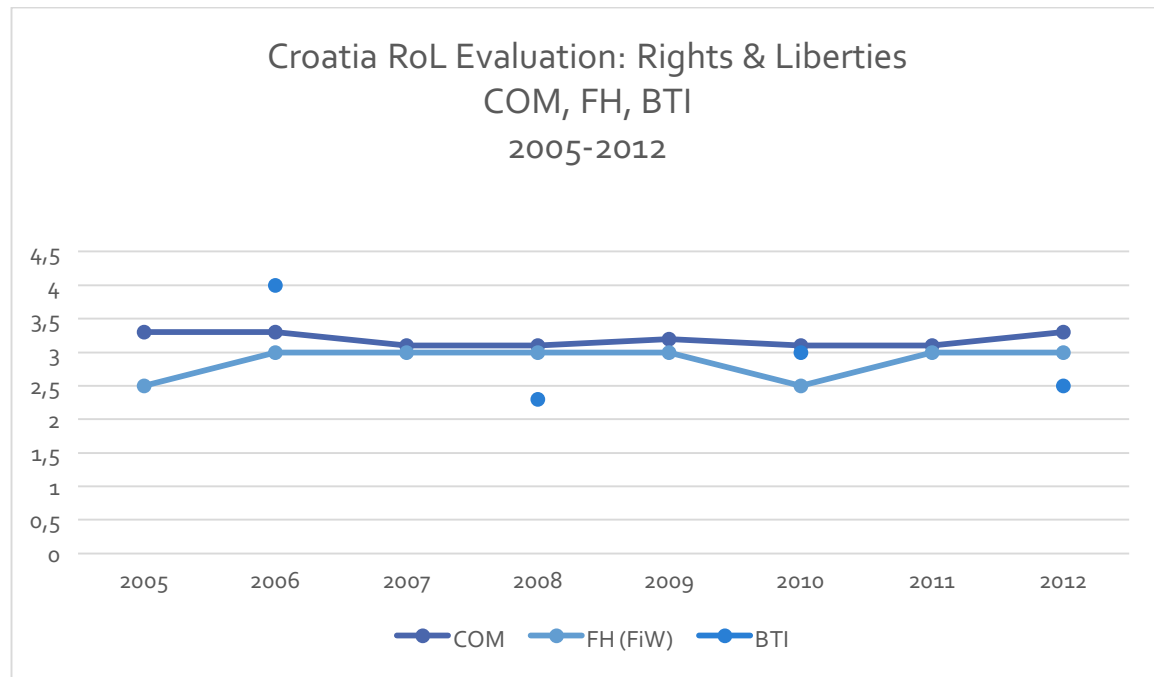


Diagram 5 - Croatia's Rights and Liberties Evaluations (COM, FH, BTI)

Index Consistence for Croatia

Index Consistence: COM

The entire spectrum of the COM's RoL evaluations for Croatia ranges between 'non-' and 'limited functionality' with peak performances in all categories in 2011, which is the year of Croatia signing the EU Accession Treaty. Until 2012, a stagnation regarding *Crime* and *Rights & Liberties* as well as a deterioration of performance with view to *Judicial Independency* and particularly *Efficiency* can be observed. A particular ascent in performance can be traced for *Judicial Independence* between 2009 and 2011.

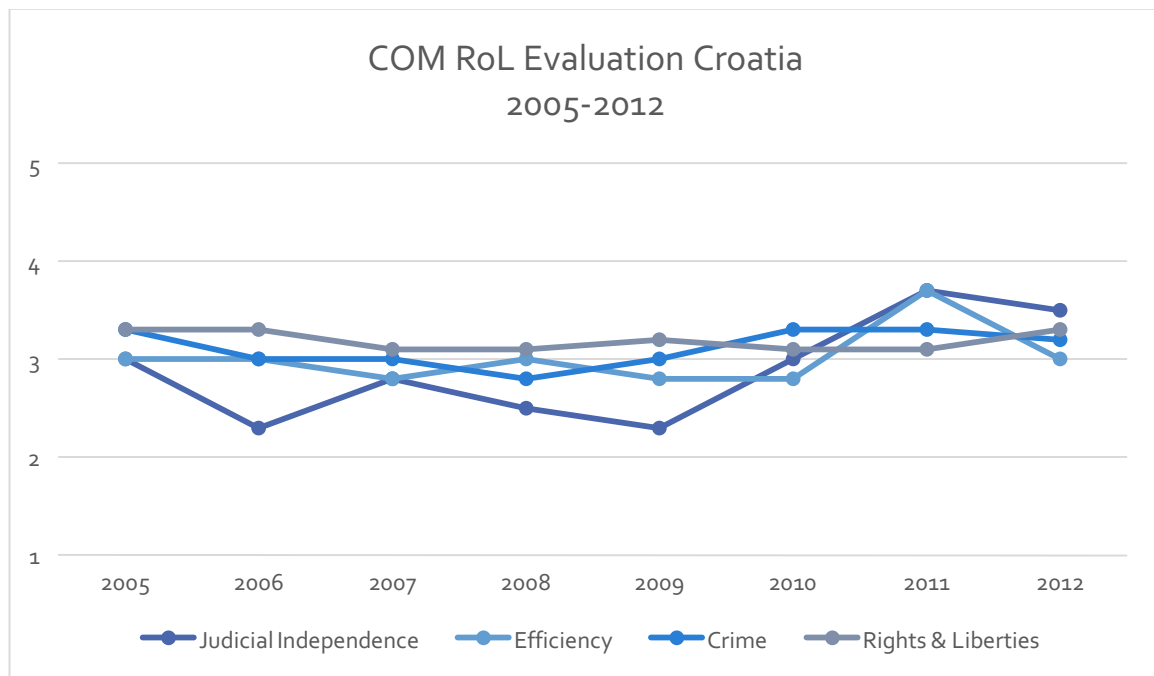


Diagram 6- Croatia's RoL Evaluation (COM, all four sectors)

Index Consistence: BTI

Compared to the COM's evaluations, the BTI scores for Croatia's RoL development cover a broader range ('non-' to 'general functionality'). Whereas the development of the *Judicial Independence* category is very similar, different assessments can be found with view to the other three categories, mostly *Crime*, which received a much lower score in 2008. Overall, the BTI evaluations are less consistent during the entire time period than those of the COM's Reports.

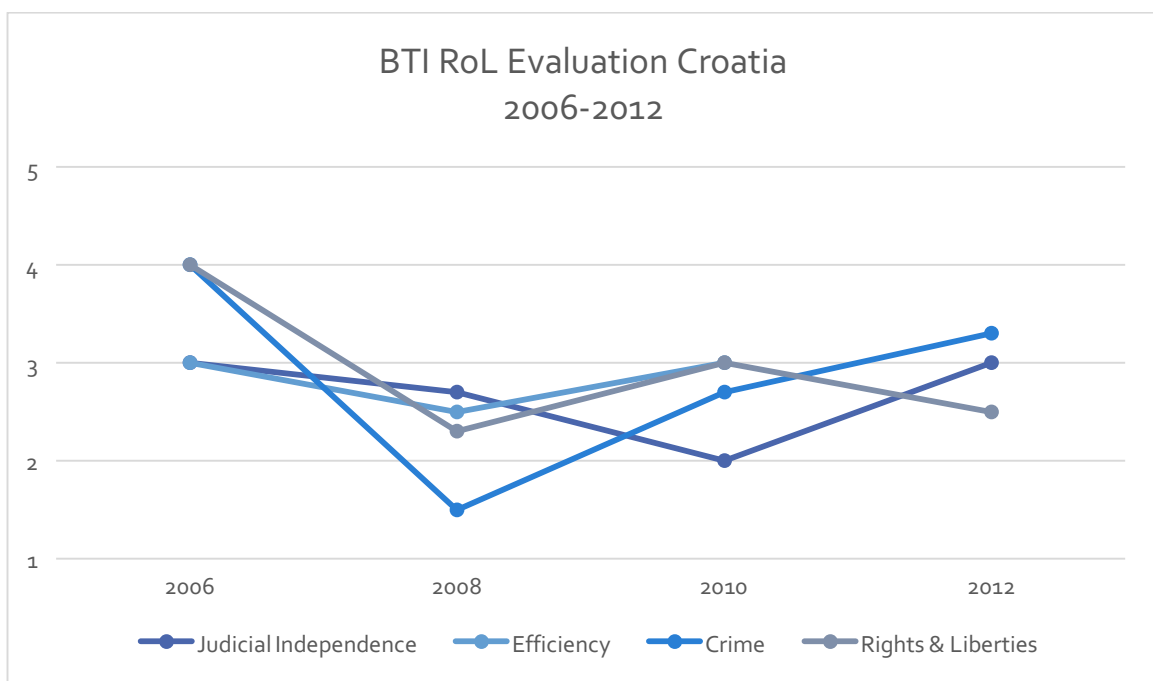


Diagram 7 - Croatia's RoL Evaluation (BTI, all four sectors)

Index Consistence: FH

Similar to the BTI findings, the FH evaluations for Croatia vary significantly during the time period analyzed, particularly between 2005 and 2007 as well as 2009 until 2012. From 2006 until 2009, stagnating scores for three of the four categories can be found (*Efficiency, Rights & Liberties, Crime*) whereas scores for *Judicial Independence* are missing for that time period.

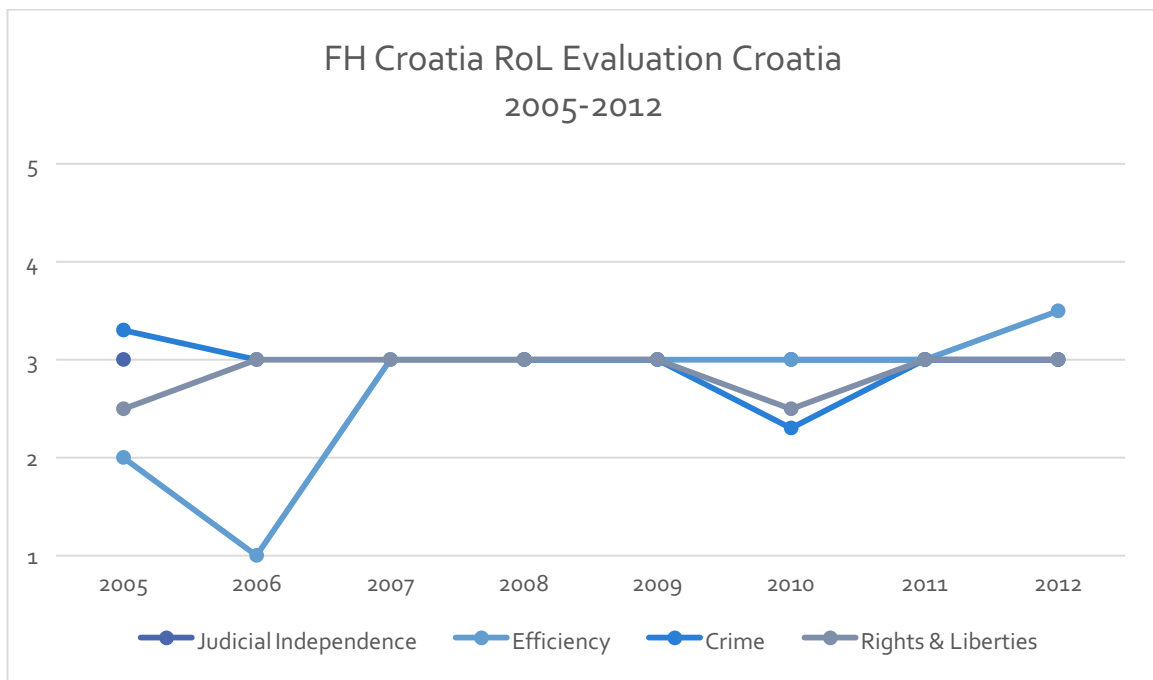


Diagram 8 - Croatia's RoL Evaluation (FH, all four sectors)

Results for Turkey

The indices' evaluation of Turkey's RoL 2005-2012

The comparison of the overall RoL scores awarded by all three indices for Turkey shows mostly 'consistent' evaluations (0-0,3 points deviation, colored green) and fewer instances of 'mixed results' (0,4-0,9 points deviation, colored yellow). No substantially 'inconsistent results' (more than one point) could be found. In comparison to Croatia's overall RoL evaluation, there is a higher level of consistency in the scores provided by all three indices. The specific scores are assembled in Table 22 below as well as illustrated in the following Diagram 9:

Turkey	EC	FH	BTI
2005	3	2,8	-
2006	3	2,8	3,5
2007	2,8	2,7	-
2008	3	2,3	2,5
2009	2,7	2,8	-
2010	3,2	2,5	2,8
2011	2,9	2,5	-
2012	2,8	2,5	2,6

Table 22 - Table of Turkey's aggregated, overall RoL scores

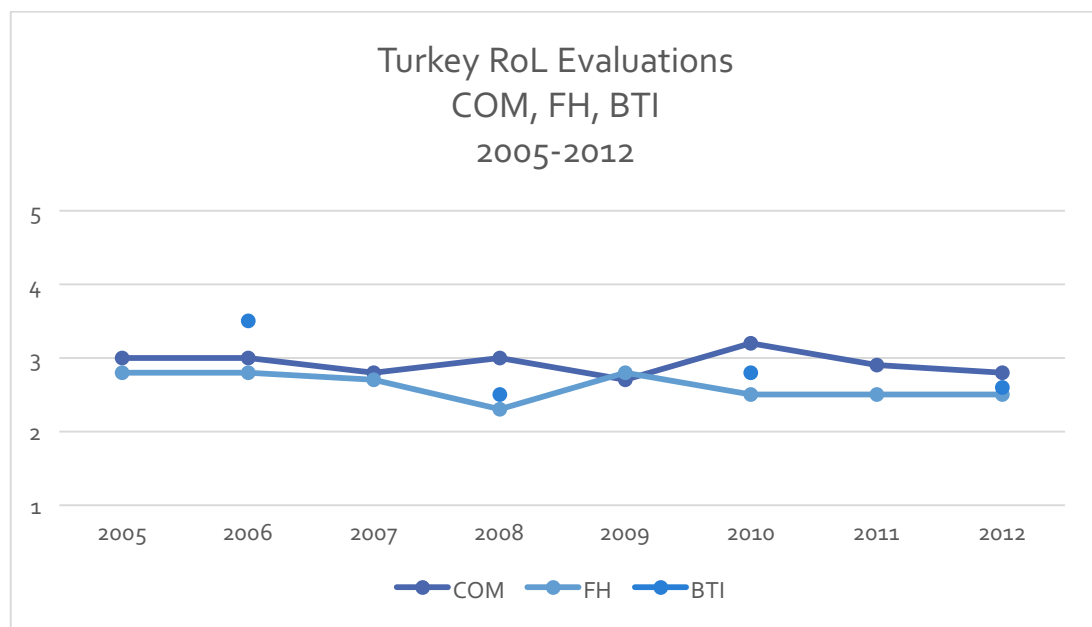


Diagram 9 - Turkey's RoL Evaluations (COM, FH, BTI)

The indices' evaluation of Turkey's Judicial Independence (RoL) 2005-2012

Similar to Croatia's evaluation of the *Judicial Independence* category, Turkey's results cover a large range ('regress' – 'general functionality') and show significant deviations across all three indices between 2008 and 2012. Some parallels in evaluations can be found between 2005 and 2007 for COM and FH Reports, although on significantly different levels (FH: 'limited functionality', EC: low 'non-functionality' scores). Similar parallels can be found for BTI and COM Reports between 2010 and 2012 (COM: 'limited functionality', BTI: 'non-functionality' – 'regress scores').

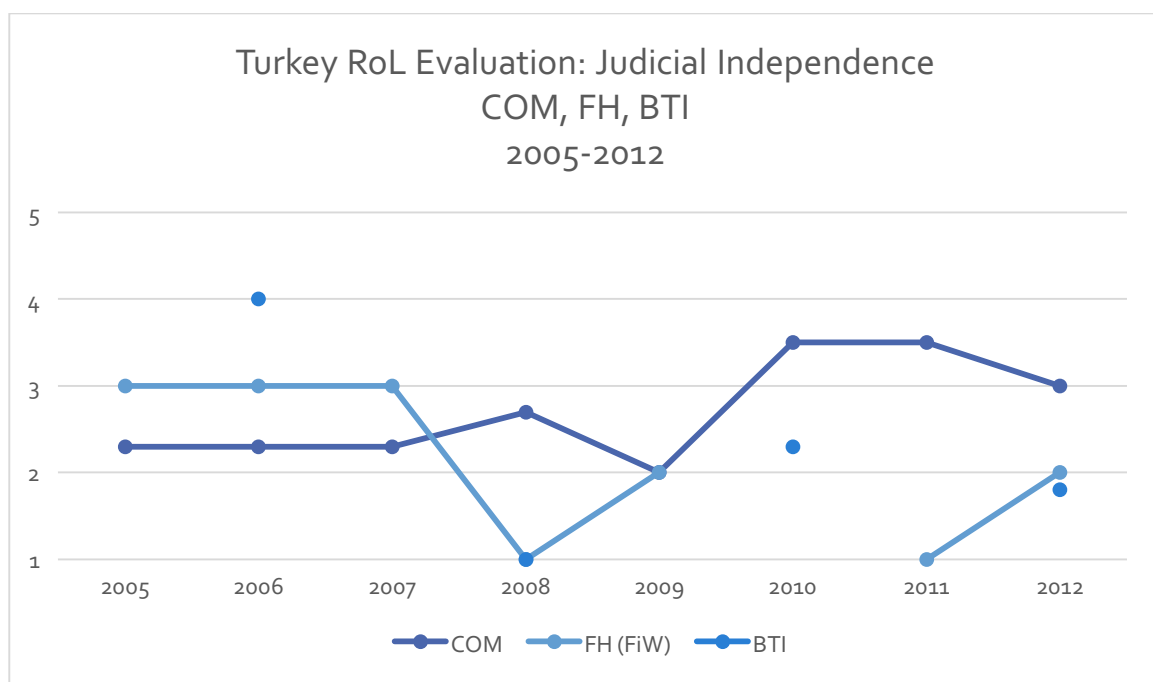


Diagram 10 - Turkey's Judicial Independence Evaluations (COM, FH, BTI)

The indices' evaluation of Turkey's Efficiency (RoL) 2005-2012

More consistency across the three indices exists for the *Efficiency* category with evaluations ranging from 'limited' to 'general functionality'. Particularly similar results can be found for FH and BTI Reports between 2008-2012. During the same period, the COM's Reports consistently award higher scores to Turkey. From 2011 to 2012, all three indices award the same scores along the line of 'limited functionality'.

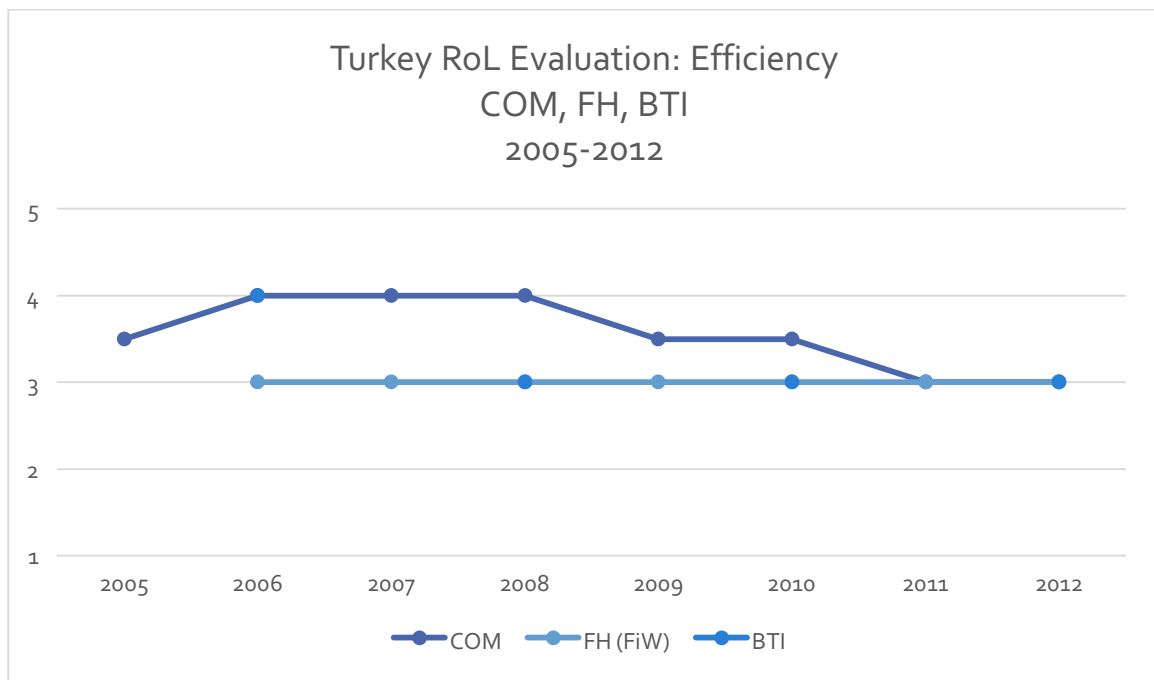


Diagram 11 - Turkey's Efficiency Evaluations (COM, FH, BTI)

The indices' evaluation of Turkey's Crime (RoL) 2005-2012

Within the *Crime* category, COM and FH evaluations run parallel. Between 2005 and 2008, COM scores range higher whereas FH scores range higher between 2008 and 2011. For the second time period the scores are more consistent than for the first. Between 2011 and 2012, the trend is being reversed again. During the entire period, BTI scores remain consistently along the 'limited-functionality' line, thus corresponding to COM evaluations in 2006, 2008 and 2010 and 2012. Since BTI scores are only available for the even years, this suggests consistency for the entire time period.

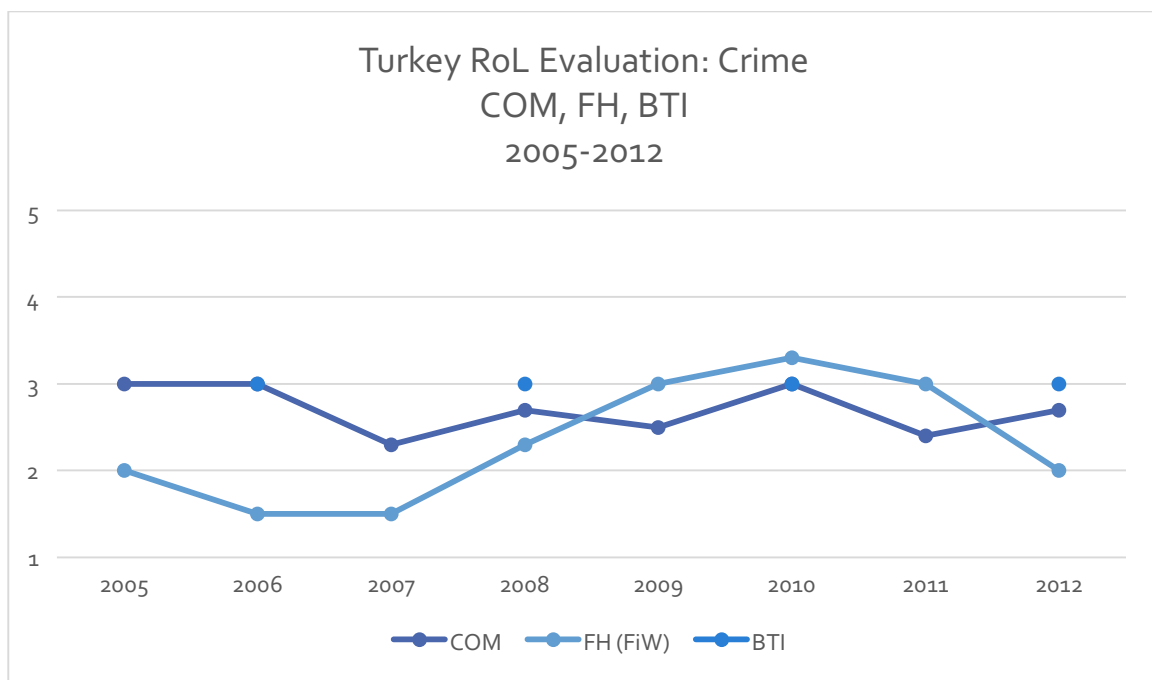


Diagram 12 - Turkey's Crime Evaluations (COM, FH, BTI)

The indices' evaluation of Turkey's Rights & Liberties (RoL) 2005-2012

The *Rights & Liberties* category also delivers 'consistent results' across all three indices along the 'limited functionality' line with FH scores being slightly above COM and BTI evaluations. Particularly similar results can be found across all indices between 2008 and 2011.

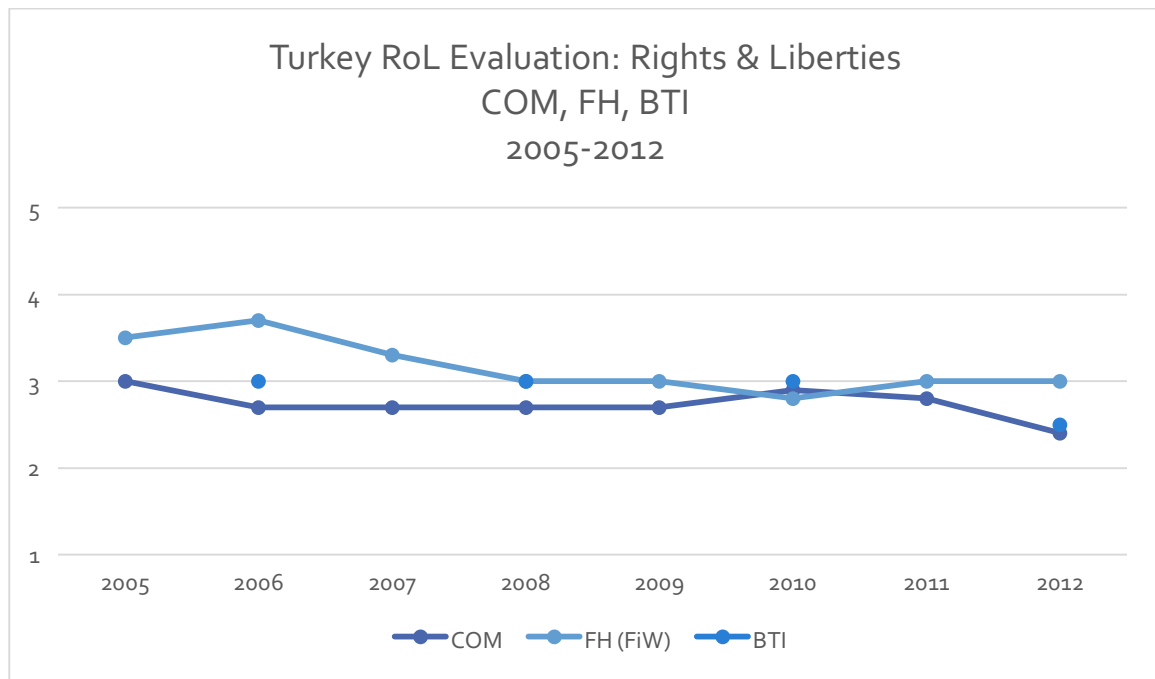


Diagram 13 - Turkey's Rights and Liberties Evaluations (COM, FH, BTI)

Index Consistence: COM

Turkey's scores are partially consistent between 2005 and 2012. Strong consistency exists particularly regarding the *Efficiency* and *Rights & Liberties* categories from 2005 to 2009. Significant deviations can be found in the results for *Crime* and *Judicial Independence*, especially between 2009 and 2012. While *Rights & Liberties*' scores are stagnating, *Judicial Independence* scores rose significantly between 2009 and 2010. A constant deterioration can be found with view to *Efficiency* from 2009 until 2012. Alternating scores exist for the *Crime* category.

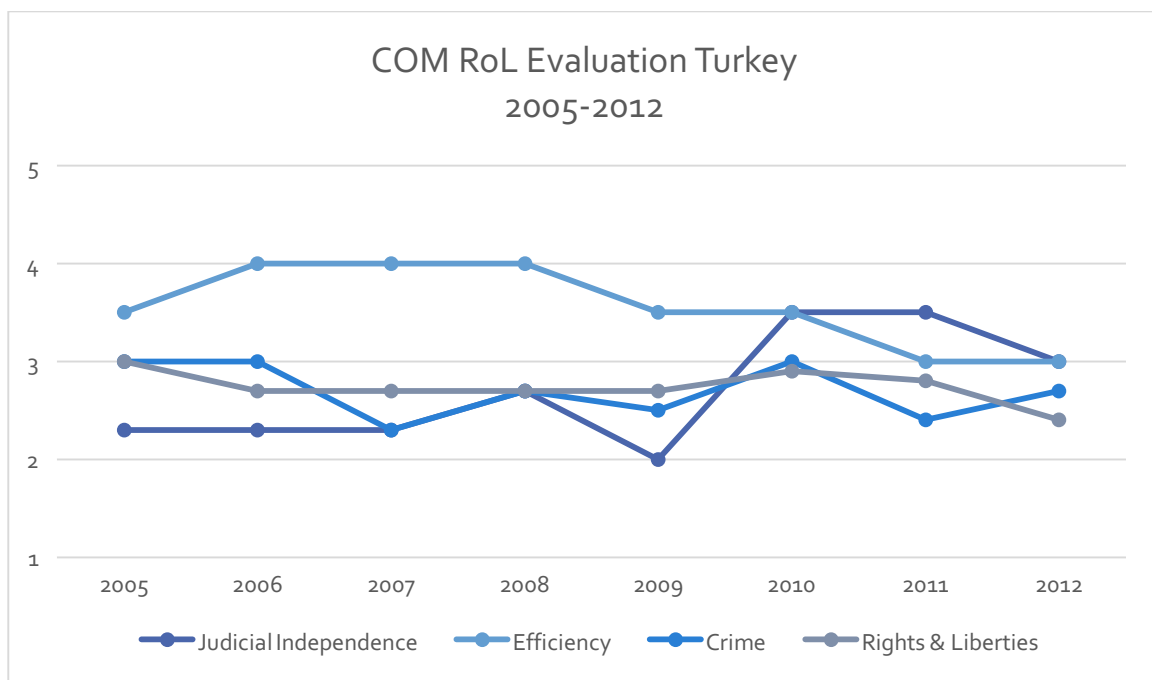


Diagram 14 - Turkey's RoL Evaluation (COM, all four sectors)

Index Consistence: BTI

BTI evaluations for Turkey are very consistent for the categories *Efficiency*, *Crime*, and *Rights & Liberties* which all center along the line of ‘limited functionality’. However, they cover a broader range than the COM’s Reports (‘regress’ – ‘general functionality’). Stagnating scores can be found regarding *Efficiency* and *Crime*, whereas a deterioration is shown for *Rights & Liberties*. *Judicial Independence* scores alternate from year to year.

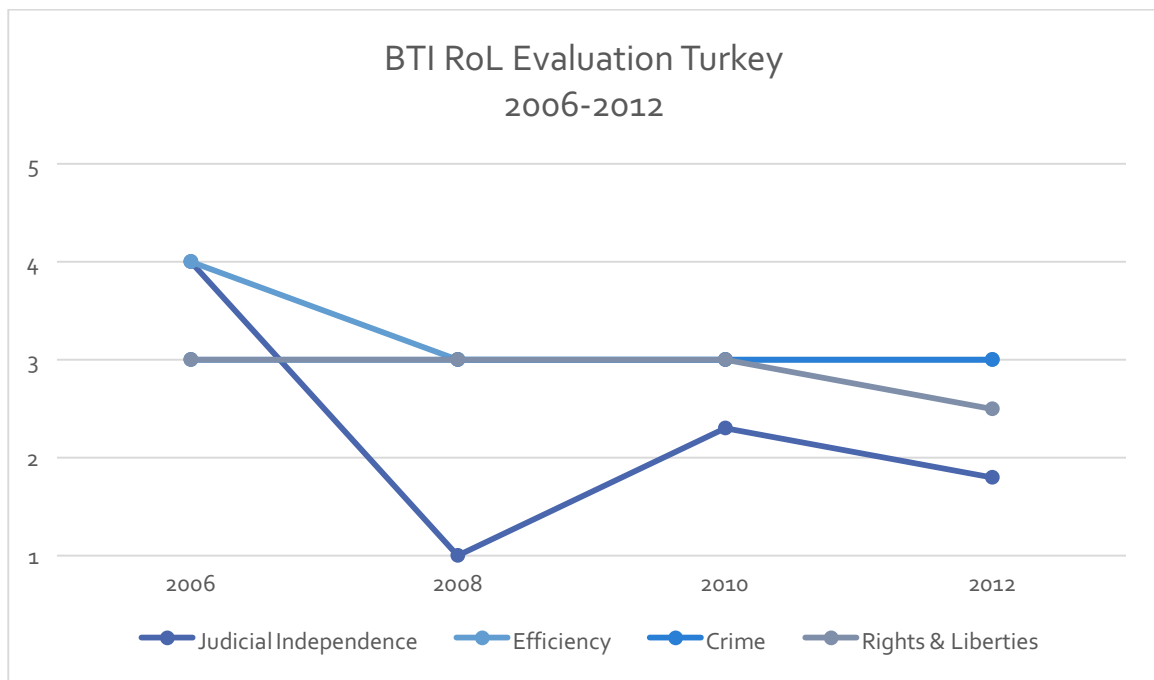


Diagram 15 - RoL Evaluation (BTI, all four sectors)

Index Consistence: FH

FH's evaluation for Croatia show the biggest range of scores of all three indices (regress – limited functionality) as well the strongest deviations of scores within one category. While Efficiency and Rights & Liberties scores are rather consistent between 2006 and 2012, Judicial Independence and Crime scores vary significantly, particularly between 2007 and 2012. The 2007 to 2008 drop and 2008 to 2009 rise of the Judicial Independence score correlates with BTI's findings, whereas the Crime curve is inconsistent with the findings of COM and BTI Reports.

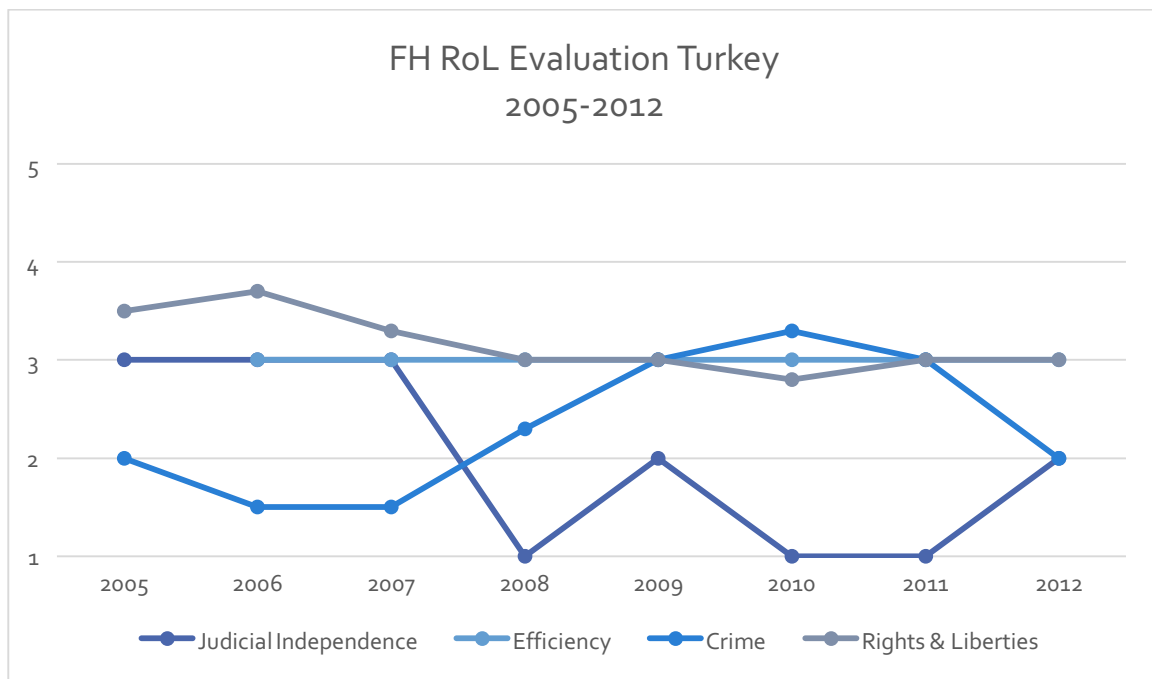


Diagram 16 - RoL Evaluation (FH, all four sectors)

After presenting the findings on questions three and four, the last two questions will be addressed subsequently:

5. How consistent are the three indices' results with view to the overall RoL evaluation for each case study?

6. How consistent are the three indices' results with view to each RoL element's evaluation?

Similarities and Differences in overall RoL assessment for both case studies

Comparing the case studies' scores shows that the indices' findings mostly correlate with one another. The majority of scores are either 'consistent' (0-0,3 points deviation) or have less than a 1-point deviation among them which places them either in the same or the adjoining evaluation level ('non-functionality' or 'limited functionality'). The evaluation of Turkey is more consistent among all three Reports during the entire time of analysis (2005-2012), whereas stronger deviations can be found for Croatia between 2005 and 2007. Each case study's overall RoL score is assembled in Table 23 below:

Croatia	COM	FH	BTI	Turkey	COM	FH	BTI
2005	3,2	2,7	-	2005	3	2,8	-
2006	2,9	2,3	3,5	2006	3	2,8	3,5
2007	2,9	3	-	2007	2,8	2,7	-
2008	2,9	3	2,3	2008	3	2,3	2,5
2009	2,8	3	-	2009	2,7	2,8	-
2010	3,1	2,6	2,7	2010	3,2	2,5	2,8
2011	3,5	3	-	2011	2,9	2,5	-
2012	3,3	3,1	2,9	2012	2,8	2,5	2,6

Table 23 – Croatia and Turkey's assembled RoL scores (2005-2012)

Similarities and Differences in individual RoL elements' assessments for both case studies

The following scores have been generated for each of the four shared RoL elements for the respective case studies:

Country 1/Year	COM	FH	BTI	Country 2	COM	FH	BTI
Croatia 2005	C: 3,3 R&L: 3,3 Jl: 3 E: 3	C: 3,3 R&L: 2,5 Jl: 3 E: 2	No Report issued	Turkey 2005	C: 3 R&L: 3 Jl: 2,3 E: 3,5	C: 2 R&L: 3,5 Jl: 3 E: -	No Report issued
Croatia 2006	C: 3 R&L: 3,3 Jl: 2,3 E: 3	C: 3 R&L: 3 Jl: - E: 1	C: 4 R&L: 4 Jl: 3 E: 3	Turkey 2006	C: 3 R&L: 2,7 Jl: 2,3 E: 4	C: 1,5 R&L: 3,7 Jl: 3 E: 3	C: 3 R&L: 3 Jl: 4 E: 4
Croatia 2007	C: 3 R&L: 3,1 Jl: 2,8 E: 2,8	C: 3 R&L: 3 Jl: - E: 3+	No Report issued	Turkey 2007	C: 2,3 R&L: 2,7 Jl: 2,3 E: 4	C: 1,5 R&L: 3,3 Jl: 3 E: 3	No Report issued
Croatia 2008	C: 2,8 R&L: 3,1 Jl: 2,5 E: 3	C: 3 R&L: 3 Jl: - E: 3	C: 1,5 R&L: 2,3 Jl: 2,7 E: 2,5	Turkey 2008	C: 2,7 R&L: 2,7 Jl: 2,7 E: 4	C: 2,3 R&L: 3 Jl: 1 E: 3	C: 3 R&L: 3 Jl: 1 E: 3
Croatia 2009	C: 3 R&L: 3,2 Jl: 2,3 E: 2,8	C: 3 R&L: 3 Jl: - E: 3	No Report issued	Turkey 2009	C: 2,5 R&L: 2,7 Jl: 2 E: 3,5	C: 3 R&L: 3 Jl: 2 E: 3	No Report issued
Croatia 2010	C: 3,3 R&L: 3,1 Jl: 3 E: 2,8	C: 2,3 R&L: 2,5 Jl: 3 E: 3+	C: 2,7 R&L: 3 Jl: 2 E: 3	Turkey 2010	C: 3 R&L: 2,9 Jl: 3,5 E: 3,5	C: 3,3 R&L: 2,8 Jl: 1 E: 3	C: 3 R&L: 3 Jl: 2,3 E: 3
Croatia 2011	C: 3,3 R&L: 3,1 Jl: 3,7 E: 3,7	C: 3 R&L: 3 Jl: 3 E: 3	No Report issued	Turkey 2011	C: 2,4 R&L: 2,8 Jl: 3,5 E: 3	C: 3 R&L: 3 Jl: 1 E: 3	No Report issued
Croatia 2012	C: 3,2 R&L: 3,3 Jl: 3,5 E: 3	C: 3 R&L: 3+ Jl: 3 E: 3,5	C: 3,3 R&L: 2,5 Jl: 3 E: -	Turkey 2012	C: 2,7 R&L: 2,4 Jl: 3 E: 3	C: 2 R&L: 3 Jl: 2 E: 3	C: 3 R&L: 2,5 Jl: 1,8 E: 3

Table 24 - Croatia and Turkey's RoL scores by sector (2005-2012)

The indices' results for each RoL element/category were compared, first in pairs, then across all three indices. This comparison shows, that the majority of evaluations across the three indices are either 'consistent' (Croatia: three, Turkey: four instances) or 'mixed' (Croatia 10, Turkey: nine instances), whereas only a minority of evaluations are 'inconsistent' (Croatia: one, Turkey two instances). This refutes the hypothesis that the COM's Reports's evaluations

differ significantly from the others. This was one of the two premises for assuming political bias. In contrast, across all three indices, there are significantly high levels of agreement between the COM's and FH's evaluations for Croatia (20 same/similar results as opposed to eight for Turkey). At the same time, particularly high levels of disagreement can be found also between the COM's and FH's evaluations, this time for Turkey (10 as opposed to three cases of 'inconsistent results' for Croatia). The nature of both of these findings will be examined and interpreted in Chapter VIII. **Discussion.** Table 25 below summarizes the comparison's findings both in pairs and across all three indices for each of the four RoL sub-components:

Case study	COM + FH pairs	COM + BTI pairs	FH + BTI pairs	COM + FH + BTI
Instances of 'consistent results' (0-0,3 points deviation)				
Croatia	20	5	4	3
Turkey	8	8	7	4
Instances of 'mixed results' (0,4-0,9 points deviation)				
Croatia	6	5	5	10
Turkey	3	3	4	9
Instances of 'inconsistent results' (>1 point deviation)				
Croatia	3	3	5	1
Turkey	10	5	5	2

Table 25 - Consistency of results for Croatia and Turkey

Limitations of the analysis

Apart from the missing data of the BTI for the uneven years (2005- 2011), there are two other limitations to the analyzed data: the difference in length and detail of the Reports as well as circle-referencing among them. First, the length of the Reports varies significantly from a median one page by Freedom House over a median two pages by Bertelsmann to a median five pages by the European Commission (raw text). Naturally, the longer the analysis, the more comprehensive it becomes regarding the number of RoL elements covered as well as their depth. For example, Bertelsmann and the European Commission include more examples supporting their evaluations than Freedom House. This issue is partially rectified by only

comparing the RoL elements all indices have in common and understanding examples as part of the coding unit for a specific RoL sub-component.⁸¹

BTI and FH were chosen as test-indices for the European Commission because of their similar backgrounds and evaluation focus as well their independence from the accession process, thus making the comparison more reliable. However, circle-referencing may affect said neutrality because it is difficult to distinguish whether an index based its findings merely on someone else's factual information or their interpretation thereof. Those cannot easily be separated. In fact, factual information may be selected (unconsciously) based on what message or evaluation the author wants to communicate.

Both BTI and FH have included the COM's Reports as their sources. All BTI Reports on Turkey between 2006 and 2012 have relied on the respective COM's Progress Reports two years prior to theirs, e.g. the COM's 2004 Progress Report on Turkey was used for Bertelsmann's 2006 Report. The same is true for three BTI Reports on Croatia (2008-2012 quote COM Reports 2006-2010). Freedom House specifically refers to the COM's Progress Reports for Croatia one year prior to theirs for the years 2006-2010 and 2012, e.g. the COM's 2006 Progress Report on Croatia was used for FH's 2007 Report. If the COM's Reports were treated as decisive for the Bertelsmann and Freedom House evaluation instead of being one among many sources, this could compromise the comparison and possibly lead to similar, but delayed results. For example, if the COM's Progress Report on Turkey for 2008 was crucial for Bertelsmann's 2010 evaluation, similar results should be found in both Reports. Such strong correlation (same evaluation or 0,3 deviation) can be found in half of the respective BTI Reports and most of the Freedom House Reports as highlighted in the Table 26 below.

Both FH and BTI (although with fewer instances) Reports appear to draw significantly from the COM's findings. The instances of time-shifted consistent evaluations between the indices is summarized in the tables below:

⁸¹ As outlined in Chapter VI.1.2, the comparison is conducted for the elements Judicial Independence, Efficiency, Crime, and Rights & Liberties. There are only five instances where a report does not cover all of them (Judicial Independence in the FiW report for Croatia, 2006-2009) and Efficiency in the FiW report for Turkey, 2005).

Turkey	QCA Score	BTI	COM
2008/2006	Crime	3	3
	Rights & Liberties	3	2,7
	Judicial Independence	1	2,3
	Efficiency	3	4
2010/2008	Crime	3	2,7
	Rights & Liberties	3	2,7
	Judicial Independence	2,3	2,7
	Efficiency	3	4
2012/2010	Crime	3	3
	Rights & Liberties	2,5	2,9
	Judicial Independence	1,8	3,5
	Efficiency	3	3,5

Table 26 - Turkey's score consistence between the COM and BTI

Croatia	QCA Score	BTI	COM
2008/2006	Crime	1,5	3
	Rights & Liberties	2,3	3,3
	Judicial Independence	2,7	2,3
	Efficiency	2,5	3
2010/2008	Crime	2,7	2,8
	Rights & Liberties	3	3,1
	Judicial Independence	2	2,5
	Efficiency	3	3
2012/2010	Crime	3,3	3,3
	Rights & Liberties	2,5	3,1
	Judicial Independence	3	3
	Efficiency	-	2,8

Table 27 - Croatia's score consistence between the COM and BTI

Croatia	QCA Score	FiW	COM
2007/2006	Crime	3	3
	Rights & Liberties	3	3,2
	Judicial Independence	-	2,3
	Efficiency	3	3
2008/2007	Crime	3	3
	Rights & Liberties	3	3,1
	Judicial Independence	-	2,8
	Efficiency	3	2,8
2009/2008	Crime	3	2,8
	Rights & Liberties	3	3,1
	Judicial Independence	-	2,5
	Efficiency	3	3
2010/2009	Crime	2,3	3
	Rights & Liberties	2,5	3,2
	Judicial Independence	3	2,3
	Efficiency	3	2,8
2012/2011	Crime	3	3,3
	Rights & Liberties	3	3,1
	Judicial Independence	3	3,7
	Efficiency	3,5	3,7

Table 28 - Croatia's score consistence between the COM and FH (FiW)

VII.2 Interview and Internal COM Document Revision Results

The interview results will be presented in the following sub-sections: the progress report creation process (VII.2.1), the experts' opinions on the political bias accusation vis-à-vis the COM (VII.2.1), the experts' explanations for (dis-)similarities between their institution's results (VII.2.3), and their reflection on the independent variables explaining the (in-)consistency among results (VII.2.4). The results will be portrayed in order of relevance for the thesis. Consequently, the respondents' opinions on the political bias claim will be presented first before continuing with the other interview topics.

VII.2.1 Progress Report Creation Process

The interviewees' accounts of the Progress Reports' production have been summarized and included in the III.3.1.

VII.2.2 Experts' Opinions on the Political Bias Accusation vis-à-vis the COM

The interviewed experts were asked about their understanding of the concept of political bias, their awareness of the allegations that the COM's Progress Reports were politically biased and if/why they personally believed those allegations to (not) be true.

The shared understanding of the concept of political bias with view to development Reports was the lack of objectivity either on the personal level of the report's author(s) or on the institutional level through in- or external influences (such as "instructions"). Its consequence was considered to be the more or less favorable evaluation of a country compared to its de facto developments.

"Decision are made that are not entirely based on facts that have been presented...once facts have been presented there is an extra calculation of is this what we wanna report or say for a particular political end." (FH official)

"[It] is always there...it is hard to remain completely objective. This is why we employ an elaborate calibration process". (DG NEAR official)

Apart from one DG NEAR official, all interviewees confirmed their awareness of the political bias accusations vis-à-vis the European Commission, offered similar explanations for the perception but demonstrated differing levels of belief in the claim. A repeated supposition was that possible alterations do not take place at the level where the Reports are originally drafted but during a later stage when the Commissioners (political appointees and their staff) are included in harmonizing the Reports.

"My guess is that the people who are doing the research are doing their job just as much as the people doing the research for Freedom House...I've met people who work in these country offices in the EU...I think they are absolutely doing it in good faith. I think it comes in

the report writing phase...I am not saying that they are misrepresenting the facts in their Reports but that they are more careful about criticism in certain areas in certain countries... so maybe they found something damning and maybe they decided that they were not going to highlight that while Freedom House would choose to highlight that. Maybe they would choose to write something vague...” (FH official)

Particularly the language used in the COM’s Progress Reports to interpret certain events or developments was pointed out to be disputable.

“People always say: this is what the EU Progress Reports said but we need to be careful because there is a certain amount of politics in the Reports...My personal impression from sitting in our meetings and what the country experts and regional experts feels is that ... We were looking at...Croatia...we were looking at minority groups and how they are treated. And somebody said ‘Oh well, the EU report it says no problem’ and somebody else said ‘Well, but the EU progress report would never criticize’. So I think there is the impression that there are certain places where language is very judicious where it would never actually say that there is a problem...people might learn to read between the lines if they read a lot of those Reports.” (FH official)

“If it is politically opportunistic that [the Reports] are critical, this will be incorporated...if the mood is that it is more sensible to be commendatory, then that...” (BTI official)

This view has been supported and amended by the notion that the Member States’ reaction to the Commissions draft Reports has in the past aligned with the general attitude in the Council towards accession. During the accession round of 2004, where EU expansion was welcomed, the European Commission was not exposed to allegations by the Member States of being “*too positive*” whereas today expansion is generally viewed more critically wherefore it is “*in their political interest depicting positive Reports as too positive*” (DG NEAR official). As has been repeatedly mentioned, no formal influence can be exerted but there are some informal options. Since the Council has unofficial meetings with officials from the COM in preparation of the Reports, their comments and concerns may or may not be included in the drafts. Also, the candidate country’s Ambassador can try to convince the General Directorate to “*depict this or that a bit more positively*”. An alternative view denied direct influences but acknowledged that “*mistakes are possible*” and given the fact that the COM is a “*living institution with*

political parties...there might be people who try to indirectly influence something” (DG NEAR official).

The dissenting opinion offered by a DG NEAR official stressed that he was not aware of any political bias accusations yet critique was voiced with regards to the Reports being “*too bureaucratic, not easy to read, [and] retrospective*”. However, he did also acknowledge that accusation have been made by civil society groups of drawing too positivistic pictures of the candidate countries. He also suggested that perception of political bias may be based on personal feelings:

“For the political criteria you cannot always back up everything with facts. If the judiciary has been laboriously restructured this is a big success for the candidate. Yet the country may receive the evaluation “little progress” because corruption remains high. Then, countries feel mistreated. They also feel personally affected if it concerns their units, e.g. administrative capacity, they feel undervalued.”

Apart from potential influences within the Union, external influences named included international Actors who assert “*extreme pressure*” (USA) as well as politicians and media from the candidate countries themselves who feel like they have been “*evaluated too negatively*” or that others have been “*supported more strongly than them*”. One interviewee argued that this line of criticism is particularly easy because it does not need to be substantiated but is yet believed by the public.

With view to the credibility of the political bias claim, a range of different views was expressed. Those who agreed with the notion felt that the COM’s officials are “*unconsciously subjectively biased*”. Moreover, they persisted that the European Council of Ministers can assert its institutional and Member States their individual interests vis-à-vis the Commission which decreases the level of objectivity in the interpretation of the Reports’ results. However, even those respondents who opposed the claim acknowledged that based on the different interests of candidate countries, governments, civil society and international organization involved in the report production there persist at least a high level of potential influence.

VII.2.3 Experts' Explanations for (Dis-)Similarities between their Institution's Results

Similarities

Reasons given for similarities between indices' results included the accurate ascertainability of developments in the respective country, similar underlying values and concepts as well as shared sources. All of these aspects correspond with the theory-led independent variables introduced before and will now be examined in more detail.

Adequate Assessment

All interviewees agreed that the reality of a country's status quo and development process can be accurately assessed through indices, because "*the country is no secret*" in terms of the apparentness of its problems as well as the transparency of political developments which are decisive for the scores awarded.

"Since we evaluate the same countries, their problems are relatively obvious. For example, the Western Balkans and Turkey still have such extensive issues that those can be hardly overlooked or concealed".

Similar underlying concepts

Even if the factual basis is congruent, each index interprets the findings which influences its results. Multiple interviewees suggested that the indices' shared underlying concept of a democracy ruled by law leads to "*similar conceptions of the situation*".

Shared Sources

Similar evaluations could also be based in shared sources which on the one hand refers to the same informational base, e.g. the usage of Reports of international Organizations such as Amnesty International, Transparency International or the Worldbank/IMF. On the other hand, a mutual utilization of the indices' Reports will eventually cause evaluations to converge.

"[We are a] secondary user of EU analyses particularly with view to Turkey versus the other Middle Eastern countries", BTI regional coordinator Dr. Völkel explained, though emphasizing that a deliberate denunciation of this procedure has been chosen for the latest BTI Report of 2016.

Dissimilarities

Aim of the Report

The most strongly emphasized reason for dissimilar report findings was the report's aim. *“If the aim is different, the same subjects are looked at and assessed in different ways.”* (BTI official). All three indices pursue very different goals. Until it changed methodology in 2015, the COM never stressed the comparability of the individual candidate countries through the Reports but treated them as individual cases that have to be measured against the Acquis. In contrast, the BTI and FH specifically aim at making countries all across the world comparable through their evaluations which is the reason for employing numerical scales. Furthermore, the COM's Reports are checking the adoption and implementation of Acquis rules, thus providing a basis for decision-making for the Member States and the Council. In contrast, BTI and FH are not catering to any specific other/higher institution (although their Reports are being used by government officials for policy decisions). While the BTI only examines transitioning countries in order to derive best practices for democratic development, FH's approach is to cover a maximum of countries worldwide and simply portray their level of freedom. Although one interviewee stated that FH has geopolitical interests that could cause deviations, this view cannot be substantiated. Additionally, the interpretation of the factual findings is crucial. As has been mentioned, both FH and BTI try to assess the ‘overall picture’ of the respective country whereas the COM focuses particularly on its progress from the last year. The latter can lead to a more positive or negative evaluations. While the overall picture may remain very similar - although specific measures have been taken - those measures could be viewed by the COM as significant progress (or regress) and thus affect the assessment. Since the COM's Reports are also the basis for financial support decisions (IPA), their wording and assessments are guided to incentivize the candidate countries to continue their progress in fulfilling the Acquis goals.

In contrast, BTI's regional coordinator Dr. Völkel asserts:

“The BTI has no autonomous opinion regarding Turkey's EU accession und generally no overriding goals. There is no political exercise of influence, no expectations and the BTI board of Professors decides based on academic standards [...] if need be, it makes politically unpopular decisions, such as declaring Russia as either a dictatorship or democracy.”

Different Concepts, Methods and Guidelines

Based on the different aims, different concepts (of the RoL) as well as different evaluation methods and guidelines to interpret the findings have been named as a cause for deviating results. While FH and BTI issue specific questionnaires that their authors respond to both quantitatively and qualitatively, there is no such practice for the COM. Naturally, the more or less specific the addressed questions/topics are, the more or less the findings will vary. Apart from the different thematic focuses of the questions, the “*conceptual status of a question within the entire theoretical framework*” (BTI official) can heavily influence its salience for the final score awarded. Also, specific guidelines for the interpretation of findings can be influential. For example, assessments in the COM’s Reports should be in line with the carrot and stick approach while FH and BTI have no similar requirements for their authors.

“Positive steps are mentioned first to encourage and congratulate candidate states on their progress...[we can later] say: well, but things remain to be done...”. (DG NEAR official)

Given the nature of BTI and FH, they employ an approach that is closer to the requirements of academic research. In contrast, the COM’s Reports’ methodology is tailored to its aim which in the eye’s of the interviewed DG NEAR official does not affect its quality:

“In general, our work is very transparent. We do not do any academic work here but we use our freedom to collect and use all information but digressing from academic work we do not cite every source. This is a methodological approach one can criticize...and in this regard it is a political report that follows another methodology. All I can say is that there has been no critique regarding the quality of the methodology from any of the Member States.” (DG NEAR official)

Range of Resources used in Production

The range and depth of resources used in the production of the Reports has been emphasized as very decisive for the quality of the Reports by all interviewees. This is where both FH and BTI representatives saw the COM’s advantage because it disposes of broad sources, Country Delegations and expert trips to the country in question as well as close contacts to the country’s authorities which makes it easier to get and verify information which is “*not the same thing as a mission who goes for one week to a country*” (BTI official).

“[We have] direct access and colleagues working all year on the subjects with continuous contact to the Turkish authorities [as well as the] Turkish Mission in Brussels. Our contact and cooperation with Turkey is very close”, a DG NEAR official explained.

In contrast, the BTI and FH have to make *“trade-offs between costs and returns”* both with view to the number of people involved in producing a report as well as with view to the time spent on it. The more countries an index evaluates, the less resources can be spent on every individual one which is reflected in the length and depth of the final Reports. Another aspect mentioned was the usage of older sources for new Reports which is specifically pertinent to the BTI. Since it is published every two years as opposed to the COM’s and FH’s annual Reports, it may include information that could be outdated by the time of publication.

Subjectivity and Fluctuation of Authors and Coordinators

A view repeatedly presented was that the subjectivity of the Reports’ authors is natural, unavoidable but by and large controllable causing only *“slight differences”*. Even though authors are presented with (more or less) clear guidelines on how to assess and evaluate the countries, different people have their *“own style”* and *“see things in a different light”*.

“If I get a new analyst this year and the new analyst sees the country very differently than the analyst from last year they are going to change a lot of scores and this happens all the time. And they’ll just change scores because they don’t think it’s a 2 for independent judiciary but a 3. And there is no way to know whether this analyst has a more accurate view or whether they just have a different understanding of the process or a less good understanding of the process because they haven’t done it before and they haven’t sat in the meeting...so there is in a way a bias...but I don’t think that you can really get around that and still have a process that compares a country over time. You have to make some decision.” (FH official)

While this might only have a limited impact on individual scores or evaluations, once these scores are added (evaluations are summarized) they can add up and thus have a *“seemingly bigger impact”* resulting in a *“more improved or deteriorated evaluation than is adequate”*. All interviewees emphasized that such instances are balanced either through the regional coordinators or the Chapter Desk (COM) whose role it is to *“ensure a maximum homogeneity”*.

“Biases come out in these meetings and that’s why we have so many people involved....There is bias in every process” (FH official)

“Everybody has their own biases that they bring to the table so I’m sure that’s part of everybody’s evaluation and that’s just part of the process and then having a methodology that is standard for all countries and that has quite a lot of detail and then having a lot of people in the room who have evaluated a lot of countries on the same indicators that’s just how you even it out.” (FH official)

VII.2.4 Reflections on the Independent Variables explaining the (In-)Consistency of Results

The independent variable that potentially explains differing results between the indices is different (re-)sources. The independent variables that potentially explain similar results between the indices include similar (re-)sources and path dependence. Apart from path dependence, the interviewees’ accounts corresponding with these variables have been summarized above, too.

All interviewees confirmed that path dependence could in fact occur in their Reports given that their previous Reports (usually the one from the last year, in case of BTI the one from two years ago) will be consulted for their current evaluations. With view to the COM’s Reports, this procedure is self-explanatory since progress is defined by the positive development made from a certain point onwards. AS a DG NEAR official explains, theoretically, the Reports should take a fresh look at the candidate country each year, yet *“practically, one looks at what did I write last time? So one cannot say there is one-to-one comparability from year to year and it is hard to say what the differences are in absolute terms”*.

Also, the foregoing report is used to evaluate if initially criticized problems have been solved. With regards to FH and BTI, the numerical scores awarded in the previous year are considered a *“good basis for orientation”*, especially given the fact that *“a range of questions are temporally invariant (such as social capital)”* wherefore it is sensible to revert to the previous texts. Previous events or processes are then being included in the new report if they are relevant to understanding the current situation, such as constitutional reforms. If textual elements are being recycled, they are amended by new examples.

“Well it is an anchoring effect but I think this is intentional. The only way to make the scores comparative is to...if there is a decline to decline it from what happened from last year.” (FH official)

With view to their shared resources or using each other as references, all indices’ interviewees stressed that the other indices’ evaluations or scores did not influence their own interpretation of events. Merely factual information may be borrowed which in turn were checked before including them into one’s own report.

“We don’t take anything from others that wasn’t corroborated by another source...[We use] conversation with people, news coverage, [we check] whether the source is a government source...It’s a process of triangulation. We take a lot of subjective material and thus make it as objective as possible.” (FH official)

“We all look at each others things but if we use each others indices we are caught in a loop. We try to be aware of what other sources are saying but not determine our scores on what they are saying. When they come up, they tend to be raised by new analysts [which is] discouraged by our staff [and thus] outbalanced.” (FH official)

“We don’t include everything one-to-one. We have the big ones [indices], we consult with them but generally we are more broadly positioned than the others. We cover a bigger overall picture and are not as topic-specific as the others. We might interpret nuances differently because of the bigger overall picture.” (DG NEAR official)

VIII. Discussion

The Discussion Chapter consists of four parts. First, the research question is recapitulated and the major findings are outlined (VIII.1). Secondly, the salience of the previously introduced independent variables in explaining the analysis' results is analyzed (VIII.2). Third, the findings will be related back to academia's critique of the Progress Reports, thus supporting or rebutting specific arguments made against the quality of the documents (VIII.3). Last, the thesis will be embedded in the academic debate and reflected critically as to its relation to previous, similar studies, potential alternative explanations for findings and its limitations (VIII.4). A conclusive answer to the research question will be provided and the implications of the aforementioned findings will be outlined including recommendations for future research.

VIII.1 Research Question, Hypotheses and Key Findings

The following section will reiterate the research question and its salience both within academia and the *Realpolitik* of the European Union's accession process. It will then outline the specific findings related to the initial hypotheses.

VIII.1.1 Research Question and Relevance

The European Commission has been repeatedly and continuously accused of producing politically biased Progress Reports on its candidate countries from the candidates themselves, academia and even Member States (cf. Chapter IV.). Candidates suspected that they received evaluations worse than their actual progress, Member States repeatedly claimed the Commission's Reports to be too positive fearing premature accessions and scholars criticized the lack of conceptual clarity, transparency and uniformity of the evaluation process. The thesis therefore stipulated that if such political bias was embedded in the COM's Reports, its results should significantly and systematically differ from those of other, independent indices that test the same criteria for the same countries during the same time period. Hence, the underlying research question addresses the consistency of the COM's Progress Reports with the results of two other, independent indices (Freedom House and Bertelsmann).

Being able to make academically founded statements on the Reports' creation and consistency (and eventually, the political bias claim) is important in three ways. First, the more transparent and substantiated the report creation process, the more likely it is to provide accurate results

based on which the Member States can make informed accession decisions. The Member States must be able to trust the accurate assessment of the candidate's readiness in order to not endanger the Union's political or economic stability. Secondly, trustworthiness of the evaluation process and outcome strengthens the EU's credibility vis-à-vis the candidates. If the same standards are applied equally to all candidates, they are more likely to comply with the criteria. In contrast, candidates who feel as if they have been treated unfairly reduce their compliance efforts or resort to window-dressing or rent-seeking behavior, as previous accession rounds have shown.⁸² Third, transparency and reliability can re-establish the trust of the European citizens in European institutions thus addressing the perception of arbitrary and inscrutable decision-making processes.

VIII.1.2 Contribution to Research

The thesis closed research gaps in the following ways. First, it uncovered the COM's Progress Report creation process which has not been published in detail before. Outlining the process makes it possible to compare its approach and methodology to the other indices for the first time. This information has been gathered through semi-structured in-depth interviews with COM DG NEAR Neighbourhood and Enlargement Negotiations officials as well through revising internal documents (cf. Appendix). Secondly, it provided a methodological approach which addresses all the critical aspects found in previous, similar comparative attempts (Veebel, 2011) to make the COM's qualitative findings comparable to the quantitative and qualitative assessments of other indices (cf. Introduction). The scaled, structuring, qualitative content analysis first uncovered all indices shared RoL sub-components and then transformed their findings onto one, common scale to make them quantitatively comparable. Third, in-depth expert interviews with representatives of all three indices generated further explanations for (in-)consistent results.

VIII.1.3 Key Findings

The thesis generated four key findings. First, in contrast to Thesis 1, the three indices' Reports' results are not significantly inconsistent wherefore no political bias can be substantiated on the part of either the European Commission. Secondly, while no significant inconsistencies could be substantiated, some systematic differences were found. These can be

⁸² Window dressing refers to fiscal-political measures that aim to manipulate an evaluator into getting a favorable opinion of the actor to be assessed. Those measures do not sustainably or systematically improve the balance sheet but only alter it optically or short-term based. Rent-seeking behavior describes actions that are targeted at obtaining benefits for oneself through political interactions, such as subsidies or special classifications, that is, gathering wealth without creating new wealth (Henderson, 2016).

explained by a time-shifted circle-referencing of the indices among each other. Third, the similarity of results can be traced back to three specific factors: overlapping sources, circular citations among the indices, and anchoring effects within each index that eventually lead to path dependence. Fourth, the overlapping use of sources as well as the circular citations cause a high level of interdependence of the respective indices and disproportionately empower those who dispose of the most capacious resources.

Political Bias Claim

The political bias claim vis-à-vis the European Union could not be verified through this analysis. In contrast, both at the drafting level (where expert interviews were conducted) as well as on the outcome level (actual Reports) no evidence for intentionally politically altered evaluations could be found. While perceptions of potential unofficial involvement of political actors at later stages of the reporting phase have been voiced by representatives of FH and BTI, those could not be factually verified.

All three indices' representatives explained that their results were mainly consistent because of the observability of their evaluation targets. They argued that the chosen RoL sub-components were objectively visible and measurable leading to similar interpretations of the country's level of RoL compliance. Each interviewee supported the notion of neutrality on the part of the European Commission, at least at the drafting level, thus referring to the factual information presented in the Reports. FH and BTI representatives even stressed the high quality of the Commission's work in terms of collecting an unprecedented amount and depth of information wherefore they use it as a source for their own evaluations.

However, this eventually leads to circular citations when the COM cross-checks its results with FH and BTI. A German MEP even claimed that the COM was in fact "too unpolitical" and "extremely hesitant to call a spade a spade" (MEP wishes to remain anonymous, all quotes stem from a conference in Brussels in July 2016). Specifically, he criticizes that "the COM always uses a language that is extremely diplomatic so as not to ruffle any feathers" wherefore it was up to the Parliament to call out the candidate countries for their lack of compliance. The critique of the COM being too soft on candidate countries has previously been voiced by representatives of the FH and BTI as well but can be explained through the COM's focus on progress (an immanently positivistic approach) and the related guidelines on

their linguistic expression of progress, i.e. an encouraging tone (cf. Chapter III.3.1).⁸³ The MEP explained the COM's perceived diplomacy with its institutional role: "The people who write the Reports want to keep their jobs. In order to negotiate one has to be nice to his clients or customers". If the COM does evaluate all candidate countries better than their actual performance, this rebuts the political bias claim stating that some countries receive preferential treatment while others are treated prejudicially. In fact, the analysis substantiates the claim that the COM issued slightly better scores than FH and BTI for both Croatia and Turkey from 2005 to 2012.

Systematic Differences

The analysis showed several instances where *systematic* (not individually based) deviations between the indices' evaluations could be found ('mixed results') (cf. VII.1.2). These can be attributed to the fact that FH and BTI use the COM's Reports permanently as one of their most important sources for their factual information. Because FH and BTI use the COM's Reports one and respectively two years prior to theirs, a traceable time shift in overlapping 'consistent results' accrues from this procedure (cf. Chapter VII.1).

Similarity of Results

The similar results of all indices can be explained best by two factors that are mutually reinforcing. First, they broadly use the same and one another as sources. All three indices retract their information from major, often identical institutions such as the Worldbank/IMF, Transparency International, Amnesty International, and other NGOs and civil society activist groups. Because of its superior availability of resources (direct access to country authorities, own Country Delegation), both FH and BTI continuously draw their factual information from the COM's Reports and the COM in turn consults with FH during its drafting phase. This ensures a similar base of factual data to be interpreted. In that respect, FH and BTI's results can be understood as a quality control for the COM's findings and interpretation thereof. The similarity of their evaluations suggests that the COM's Reports are widely reliable, even in the eyes of the academically-based indices. While the interviewees of FH and BTI stated that they believed the COM's Reports to be political tools they still praised the methodological approach and density of information gathered by the drafting team.

⁸³ An internal guideline on the creation of the Progress Reports includes instructions regarding the language regimen to be used, which is a positivistic one, especially in the introductory sentences of each section. This can be attributed to the aim of the Reports. Lauding their progress is supposed to motivate the candidates to increase their investments for further progress.

Secondly, all three indices base their current evaluations on their previous Reports, thus falling prey to the anchoring effect (cf. Chapter V.2.6). The anchoring effect refers to a psychological phenomenon where initial information (the last report's evaluation) is highly influential on the evaluation of new information (the current Reports). This effect can over time lead to path dependence, that is stable scores. The combination of using each other as reference points (or even circular citations) while relying on one's own previous findings leads to converging evaluation results among the Reports that will eventually reach an equilibrium. An indication for this assumption is found in the data. An equilibrium has set in for the overall RoL score for both case studies during the time of analysis around the higher 2 or lower 3 mark (ranging from a 2,3 to 3,5 for both Croatia and Turkey).

Indices' Interdependence

As has been described above, the usage of the same sources and the respective other indices gives rise to an entirely different question than political bias, namely how independent the individual indices really are – which in turn was the premise for comparing them to one another. The initial comparison of the Reports set-up and backgrounds conducted in Chapter III suggested that FH and BTI are no stakeholders in the accession process and generally methodologically independent enough to use them as test indices. The in-depth qualitative content analysis discovered that both indices refer to the COM's Reports repeatedly for both case studies during the entire time period. The expert interviews then revealed that the COM, too, consults with FH during its drafting phase to gather information and benchmark its findings against those of the other institution. As has been argued before, this approach can be viewed both as quality control as well as an infringement on the indices' independence. With view to the political bias issue, if the COM Reports were politically biased and the other indices relied too heavily on their content, they would be likely to unconsciously import that bias into their own evaluations. Consequently, the more resources are available to an institution creating evaluations, the more influence it has on other Reports using it as a source. Thus, a capacious institution can become highly influential in shaping the general tone of evaluation. This enables a general, system-immanent critique of democracy measuring indices rather than singling out the European Commission as a potentially biased analyst.

Hypotheses

In the following section, the thesis' hypotheses will be addressed and discussed based on the tests introduced in the Conceptual Framework.

Thesis 1: The COM's Progress Reports are politically biased towards the candidate countries as indicated by significantly inconsistent scores when compared to those of Bertelsmann and Freedom House.

The thesis' initial premise was that if the COM's Reports were politically biased, they should be significantly and/or systematically different from the results of independent indices. Surprisingly, significant inconsistencies could not be substantiated (thus contradicting the results of Veebel's analysis). In fact, a high level of consistency could be found for both case studies between FH and the COM's Reports and the majority of results between BTI and the COM are either consistent or mixed, too. Given that the COM in some aspect produces better evaluation results for both case studies than do FH and BTI can be attributed to the COM's methodology in creating the Reports, particularly the positivistic language regimen requested in the Guidance Note given to the authors (cf. VII.2, Appendix). This finding supports the notion of a German MEP that the Reports are in general too complaisant, regardless of the candidate country in question. This in turn contradicts the political bias claim which implies that there is a preferential or discriminate treatment of individual candidates.

According to the interviewed experts, the high level of consistency among the Reports can be attributed to a shared RoL understanding combined with the notion that the case studies' RoL development can and has been adequately assessed by all three indices. However, they emphasize that while relying on the same factual information for the respective RoL sub-components, the indices do differ slightly in their interpretation thereof. Other factors that explain the consistent results include the usage of overlapping resources and path dependence as will be discussed in detail below.

Hypothesis 1: The European Commission issues politically biased Reports in order to uphold a specific reputation.

This Hypothesis could not be verified. As has been expressed during the expert interviews, the COM is just as concerned with a high level of neutrality as are FH and BTI because otherwise it would not be accepted as an adequate measuring organ by either the Member States or the candidate countries. This emphasized level of professionalism and neutrality on the part of the COM – at the report drafting level - has been supported by representatives of FH and BTI as well.

Hypothesis 2: The European Commission issues politically biased Reports to incentivize its candidates.

This Hypothesis could not be verified. The linguistic regime required in the pre-2015 Guidance Notes to the COM Reports' authors is –according to the expert interviews – based on motivating the candidates to increase their reform efforts. It requires positivistic formulations and the focus on progress rather than stagnancy or even regress which may portray and candidate in a more favorable light. Since FH and BTI have no motivational intentions towards their case studies, this could lead to the differences in the interpretation of factually identical information among the indices. However, the political bias definition rests on the notion that evaluations are intentionally altered depending on how much a candidate is (un-)wanted to become a EU member. Since the linguistic regime applies equally to all candidates and both case studies, Croatia and Turkey, receive slightly better evaluations from the COM than from the other two indices, the hypothesis cannot be substantiated.

Hypothesis 3: The level of report consistency depends on the usage of (re-)sources by the indices. The more overlapping resources are used, the more consistent the scores should be and vice versa.

This Hypothesis could be verified and is highly applicable in this case. As is outlined below in detail, the vast range of overlapping sources among the indices accounts strongly for the high level of consistency, particularly with view to referencing one another.

Hypothesis 4: The level of report consistency depends on the level of inclusion of their respective earlier findings. The more the indices include their old evaluations, the more prone they are to the anchoring effect resulting in path dependence which causes a convergence of their results within their own Reports over an extended period of time. The less they do so, the more independent they remain in their yearly evaluations.

This Hypothesis could be verified and is a strong explanation for the high consistency among the results as outlined below in detail. All three indices are geared to their previous Reports which in turn are based on the other indices' factual information. Consequently, their results become stable over time within their own Reports, and eventually among them, too. This in turn endangers the independence of the Reports.

VIII.2 Impact of Independent Variables

The variables examined in the following sections include the Independent Variables introduced for Dependent Variable I, Political Bias (VIII.2.1), and Dependent Variable II, the Consistency of the Indices' Results (VIII.2.2). Additionally, explanations offered by the Interviewees will be discussed (VIII.2.3).

VIII.2.1 Independent Variables for Political Bias

As has been explained in Chapter V. **Conceptual Framework**, Reputation and Incentivization are independent Variables that have to be tested if the COM's reporting results differ significantly from those of FH and BTI which is not the case. Consequently, only a brief explanation will be provided at this point.

Incentivization

Incentivization can generally be substantiated for the COM. The interviewed DG NEAR officials affirmed that the positivistic language regime was developed to motivate the candidate countries in their reform efforts. Hence, the COM's evaluative results are slightly better than those of FH and BTI. However, their scores are still mostly 'consistent' or 'mixed'. Significant deviations - as could have been expected if the COM was trying hard to motivate its candidates - could not be substantiated. Given the high similarity of evaluation results, it remains unclear if interpretative deviations can be attributed to linguistic Incentivization efforts or merely to the inevitable level of subjectivity that is pertinent to every indices' evaluation.

Reputation

All Indices profit from being perceived as fair and reliable evaluators. FH and Bertelsmann are information providers to policy makers and aid donors who rely on the credibility of their information. Whereas the same is true for the European Union, it also has to uphold the reputation as a strict enforcer of the rules enshrined in the Acquis. Additionally, the COM should abide by the Council's general position on accession. During the time of analysis (subsequent to the last two accession rounds in 2004 and 2007), this general position was rather dismissive. Consequently, one could expect the COM to produce more negatively connoted Reports than FH and BTI. This could not be substantiated. In fact, their reports produced slightly more favorable evaluations of the case studies which can be attributed to the

positivistic language regime imposed on the authors. Consequently, Reputation has been designated a low-impact variable on the comparison's outcome.

VIII.2.2 Independent Variables for Consistency of the Indices' Results

(Re-)Sources

One of the key factors explaining the similarity among the indices tested are their sources. Although the range of sources used differs greatly among the Reports based on the range of resources available to them, the COM, FH and BTI draw their key information from the same institutions and even one another.

The European Commission has by far the greatest amount of resources available to carry out its evaluations. It employs both a Country Delegation and Country Unit responsible for each candidate country which is in direct contact with the country's authorities as well as NGOs and civil society representatives during the entire year. The broadness and depth of information gathered by the COM is valued by BTI and FH which repeatedly reference their Reports on both Turkey and Croatia. Regarding the RoL, BTI's report on Turkey (2006-2012) each time references the COM's respective report from two years before, e.g. BTI 2006 includes information from the COM's report of 2004. This may also be the best explanation for the slight evaluative differences between the COM's and BTI Reports. Since the BTI is including – by then – 'outdated' COM's results, its evaluations 'lag behind' regarding its scores, as can be seen in Chapter VII.1.2. The same applies to the BTI's Reports on Croatia between 2008 and 2012 which respectively reference the COM's Reports from 2006 until 2010, sometimes even twice within the RoL relevant sections. Freedom House also draws from the COM's Reports, although from the ones one year prior to theirs. As BTI regional coordinator Dr. Brusis explains, the COM's Reports are usually more detailed than other sources because the COM employs more people within the country to gather information. Consequently, "factual information is being used but copying is avoided". In FH's Reports on Croatia from 2006-2010 as well as 2012, references are made to the COM's Reports from 2005-2009 and respectively 2011. FH's Reports on Turkey, however, do not specifically mention the COM's Reports. As is the case with the BTI evaluations, slight differences in evaluations can be traced back to the time shift between FH's evaluations and the year of the COM's Reports.

There is no specific mentioning of the BTI as a source for either the COM's Reports nor those of FH.

Although the COM does not specifically reference FH in its Reports, the Guidance Notes instruct the report's authors to conduct consultations with other think tanks regarding specific questions though not citing them explicitly. A DG NEAR ENLARGEMENT official confirmed that FH Reports are being used "but not for the first Draft and more in order to classify and review if the results generally concur or deviate". This statement clashes with another DG NEAR official's description of the process claiming that such consultations are already taking place during the drafting process. In case their findings deviate from one another, another factual check ensues. As BTI regional coordinator Dr. Brusis verifies, both FH and the COM's Reports are being used for the BTI.

Apart from using one another as sources or 'reference value', all three indices borrow from the same international organizations such as Transparency International, Amnesty International, the Worldbank, IMF, and Human Rights Watch (COM, BTI, FH) as well as the respective country's Ombudsman Reports (COM, FH). The fact that all three indices reference one another (although time shifts do cause some systematic differences) partially explains the similarity between their results. Another aspect that underpins and amplifies both the similarity and stability of the scores is path dependence.

Path Dependence

Path dependence exists if certain decisions – such as the evaluations within the Reports – remain constant and if future decisions are significantly dependent on those from the past. Pierson has introduced a path dependence test which checks for four criteria: multiple equilibria, contingency, a critical role of timing and sequencing, and inertia which will be applied henceforth.

The multiple equilibria criterion applies to all three indices. Initially, numerous different options of evaluating a country both qualitatively and quantitatively (BTI, FH) are possible. The contingency criterion also applies to all three indices because even small events within the country, such as the adoption of a new law or the termination of a judge, can cause enduring consequences like the improvement or deterioration of assessment scores. These in turn may influence the country's accession process (EU/COM) or the amount of development aid given by International Actors (which is often based on developing indices). The timing

and sequencing criterion is also fulfilled with view to the COM's report. As all interviewed DG NEAR ENLARGEMENT officials have explained, much more significant and visible progress with view to law creation and institution building can be made in the beginning of the accession process as opposed to later stages. Consequently, once an evaluation is rather (un)favorable, it is more likely to remain so than to change significantly. With view to FH and BTI, the fulfillment of timing and sequencing is not as clear-cut. Their aim in assessing the countries is not to come to a final decision regarding its 'overall maturity' for EU accession. Hence, they can open-endedly go back and forth between scores which is why scores of the past should not matter as much for future Reports. However, the fact that all three indices base their future evaluations on their previous Reports does support the notion that timing and sequencing matters for FH and BTI Reports as well. In fact, all three indices use text modules from their earlier versions which sometimes only change slightly in terms of linguistics and examples and sometimes even remain the same entirely. The inertia criterion suggests that once a stable equilibrium has been reached it does not change any more. This cannot be fully substantiated for the indices' Reports. On the one hand, one could argue that an equilibrium has set in around the higher 2 or lower 3 mark given that the overall RoL scores for both countries are relatively stable between 2005 and 2012 (ranging from a 2,3 to 3,5 for Croatia and Turkey as well). On the other hand, scores do change up to 1,2 points which on a 1-5 scale can be considered significant enough to discard the equilibrium argument. Given the fulfillment of the first three criteria in combination with the fact that all three indices do not fully create new Reports each time but advance their previous ones does substantiate the path dependence hypothesis which in connection with the circle-referencing of the Reports is the strongest explanation for their similar results.

VIII.2.3 Explanations introduced by Interviewees

The interviewed experts reflected both on their opinions of the political bias claim as well as the high level of consistency among the reports. Their summarized ideas will be presented in the following section in order to complement the Independent Variables introduced in the Conceptual Framework.

Experts' Explanations for Political Bias Perception

The assumed motives behind politically biased reports, reputation and incentivization, could not or only partially be substantiated. Therefore, the question arises where the political bias

perception originates from. During the expert interviews, the interviewees suggested a differentiation between political bias on the individual and institutional level and delivered explanations which are summarized below.

Political bias on the institutional level

The institutional set-up of the the Progress Report creation process can partially explain the political bias perception. While the experts dismissed any form of direct influence targeted at changing their assessments, they agreed that the COM is a “living institution” while consisting of political parties “that might indirectly try to influence something” Also, an Ambassador may try and talk to a General Directorate representative to try and “make him present this or that a bit more favorable”. However, these attempts at influencing the evaluations are informal and neither their occurrence nor success rate can be evaluated at this point. Apart from the candidate countries, the Member States may have vested interest in a country’s (non-)accession. This is where the Council comes into play. As has been confirmed in the expert interviews, while different Member States may have their specific reasons for (not) wanting another country to become a EU member, more important is the “general attitude concerning accession” (DG NEAR ENLARGEMENT official). In 2004, this general attitude was more positive than today. Consequently, the COM was not “blamed to be too positive [in the Progress Reports] because accession was desired” and it “was in their political interest to highlight positive Reports” (Ibid.). However, the Council’s influence on the COM is limited. During the summer, informal meetings are held with Council’s working group representatives without any written input or exchange of report drafts. Only after the COM published the Reports, the Council can submit its Conclusions where the Member State “can depict their viewpoints” (Ibid.). Since these Conclusions “are political statements, they may vary in their tonality from the Progress Reports” and are “in tendency more negative” at least within the “years after Bulgaria’s accession up until 2013/14” (Ibid.). During the last two years, the Council’s Conclusions have become “more relaxed” and “less critical” which the DG NEAR official attributed to Juncker’s message that there won’t be any more accessions within the next five years.

More importantly, the DG NEAR official raised another interesting point on the institutional level which may add to (the perception of) politically biased evaluations: the double role of the COM. While the COM is responsible for objectively assessing the candidate countries it is also in charge of distributing IPA money (financial development assistance aid) among them.

“This double role is not easy” DG NEAR official admits “some countries concern oneself more concretely” and sometimes it needs a “targeted push for countries to do something” which “sometimes comes across as if there was an excessively positive evaluation”. However, the decision over IPA-funds is not made by the Political Desks who carry out the analysis but by other segments of the Commission, thus reducing the bias potential. While they consult with one another, financial support is generally not allocated to countries as a reward for positive developments but rather as a support mechanism for those who lack in progress. This in itself is problematic because the carrot-and-stick approach suggests that compliance is rewarded while non-compliance is punished. Moreover, this may send the wrong signal to other candidates who feel treated unfairly when candidates receive ‘incentives’ for non-compliance which may stimulate them to reduce their reform efforts.

Political Bias on the individual level

As opposed to BTI or FH, there are no individual authors for the COM’s report. A Country Unit consists of approximately 10 people where individual officials are responsible for single Chapters (out of 33-35) which are combined into one draft. Once the initial draft is finished it goes through the chain of command where it may undergo further linguistic changes. Consequently, even if the individual analysts of the COM were to (un)consciously include personal interests or beliefs into their assessments, the set-up of the evaluation system would outbalance an individual’s opinion with those of his multiple colleagues’. Since the COM’s DG NEAR ENLARGEMENT officials all come from different countries and have been socialized differently, they are likely to have differing personal beliefs as well. Naturally, working for the same institution may in turn create similar interests/beliefs. As has been confirmed in the expert interviews, there may be “slight differences” because after all, there are “people at work” instead of machines which have their own “style” and may “see things in a different light...despite the guidelines”. However, the DG NEAR ENLARGEMENT official has emphasized that “all colleagues make an honest effort to be constructive, impartial, and objective...but mistakes are possible”. Consequently, at the Country Unit level, politically biased assessments understood as intentional tempering with evaluation results can be largely dismissed.

Another potential source for individual political bias is the risk of “going native” as a DG NEAR official describes the process of an analyst identifying too much with the country he is assessing. One of his employees agrees: “Unconsciously one has intensively engaged with a

country and invested so much of one’s own energy that one wants to see progress” which may affect one’s personal objectivity. “While we try to do this as much as possible, absolute objectivity cannot be guaranteed but it can be increased through involving many people who are less engaged with specific countries”. His supervisor emphasizes that it is important to keep the balance between having highly motivated staff that are interested in accompanying a country during the accession negotiation process and keeping a “healthy distance”. Institutionally, this distance is enforced through a rotation mechanism which ensures that (politically) sensitive posts are changed every four to five years.

Experts’ Explanation for High Level of Results Consistency

Adequate Assessment

All interviewed experts suggested and supported the notion that adequate assessment of the case studies’ developments was both feasible and decisive for converging index results. This perception was based on the conviction that the Reports consisted mostly of factual information that cannot be contested, e.g. the adoption of laws or the creation of institutions. Given their focus on the same RoL-elements in combination with objectively determinable developments is a strong explanation for the similar results. FH and BTI live off their reputation as neutral and credible observers of democratization processes and consequently make use of their discretion to interpret said information according to their standards. Since they still end up producing very similar results to the COM, this corroborates the hypothesis.

The table below summarizes the findings detailed above.

(Independent) Variable	Findings
Political Bias	
Reputation	<ul style="list-style-type: none"> • Accession-fatigue of Council should result in more negative evaluations by the COM to keep up ‘strict EU’ reputation • In contrast, evaluations slightly more positive than FH and BTI
Incentivization	<ul style="list-style-type: none"> • COM employs incentivizing linguistic regime (pre-2015) • Potential personal biases of COM staff outbalanced by large number of evaluation team members and rotation principle
High Level of Result Consistency	
Sources	<ul style="list-style-type: none"> • All three indices share several same sources • Use each other as sources causing circular citation: <ul style="list-style-type: none"> → COM as source for FH and BTI → very similar results BUT with time shift (1-2 years) → FH as source for COM and BTI → BTI not specifically mentioned as source for either

Path Dependence	<ul style="list-style-type: none"> All indices use previous report as a departure point for current evaluation → likely to lead to stable evaluations b/c of anchoring effect
Interviewees' Explanation	
Adequate assessment	<ul style="list-style-type: none"> Notion supported by all interviewees True for level of fact-checking but interpretation may still differ because indices' criteria for interpretation are either not clear or not the same
Institutional Political Bias	<ul style="list-style-type: none"> Opaque involvement of political actors and double role of COM (assessment and distribution of financial means, IPA) Perception of bias prevalent in other indices, yet they regard the COM's information as credible and use it
Individual Political Bias	<ul style="list-style-type: none"> Unlikely on the individual level because personal assessments outbalanced by team Team size bigger than at FH and BTI which reduces risk of individual bias inclusion

Table 29 - Impact of Independent Variables

VIII.3 Validity of Academia's Critique of Progress Reports

Academia's critique of the Progress Reports has been described in detail in Chapter IV.4.1 and categorized into five lines of argumentation. One group of academics criticized that accession is predominantly a political decision wherefore the Progress Reports are not as relevant for accession decisions and discriminates candidate countries based on the political will to (not) include them into the Union. The second group stated that the criteria used for evaluations were insufficiently clear to be used as benchmarks (either over-inclusive or lacking depth and consistency) which they believed to be intentional on the part of the Commission to grant political flexibility. The third group denounced the inconsistent application of the criteria. The fourth group criticized that the Member States required the candidate countries to fulfill criteria they partially do not meet themselves. The fifth group admonishes the application of double standards in the evaluation of different candidate countries. Each of these allegations will be tested in the following paragraphs.

VIII.3.1 Political Motives of COM's Reports

Formally, the European Council makes its accession-related decisions based on the recommendations provided by the Commission which in turn are based on the Progress Reports. Within the academic debate of the Progress Reports, it has been claimed repeatedly that the Reports have de facto either not much impact on the accession decision or are being used as a tool to express political motives (Ludwig, Maresceau, Zalewski, Saatcioglu). These notions can neither be substantiated through the qualitative analysis nor the expert interviews conducted with the COM's DG NEAR officials. The lack of impact can be repudiated based on the fact that so far the Council has always followed the COM's recommendation with view to accession. However, a DG NEAR official expresses, the further along a candidate country is within its accession process, the less important the Progress Reports become in favor of the opening and closing benchmarks. These benchmarks are detailed roadmaps of the type of changes and reforms a candidate has to address in order to fulfill the Copenhagen criteria. Its performance in meeting those goals is then summarized in the Progress Reports. According to him, the benchmarks as well as the Monitoring Reports are taken very seriously by the candidate countries and are being distributed within the Ministries of the respective candidate country, not least as a means for gathering more financial and/or political support for internal reforms. He further emphasized that the Member States generally commend the Commission for the high quality of its Reports but regularly express concerns that the Commission may be 'too soft' on the candidates. Since the 2007 enlargement round there has been a general consensus on non-accession. Hence, if the Reports were merely a political tool to fulfill the Council's wishes, this should be reflected in more negative evaluations. This cannot be substantiated for the case studies analyzed. However, evaluations starting from 2015 may become more critical given that the 2015 Guidance Note introduces the Progress Reports' authors' to the option of negative evaluative terms such as "no progress" or even "backslide" (cf. VII.2).

VIII.3.2 Aptitude and Usability of Copenhagen Criteria

The aptitude of the Copenhagen Criteria for making adequate evaluations was challenged based on their over-inclusiveness as well as their lacking depth and consistency. With view to the RoL for the case studies examined, this notion cannot be substantiated. The qualitative content analysis clearly showed that the same nine RoL sub-components were tested in the COM's Progress Reports for both case studies, where applicable. These components have been made clear across a number of Commission-issued documents, although the critique is

valid that no single document codifying the actual criteria exists. During the expert interviews, this was justified with the explanation that once specific criteria are named, the flexibility to address other, additional criteria would be harmed (cf. VI.2 Interview Results). Additionally, the interviewed DG NEAR officials emphasized that the notion of inconsistency can be traced back to the fact that while the general performance of a country is evaluated, the strong focus is on its actual progress. This means that some RoL areas are scrutinized and highlighted stronger than others just because more progress has been achieved by the respective country. Additionally, the Reports – as opposed to those of the other democracy indices – are not meant to compare the countries to one another but to decide whether these individually fulfill the Acquis. Therefore, their consistency is more important within one candidate countries than across all of them. The Progress Reports themselves do not lay claim to make cross-country comparisons, at least not until their methodological revision in 2015 which addresses their comparability.

VIII.3.3 Application of Copenhagen Criteria

The same logic applies to the application of the Criteria. The Reports monitor the progress within one country from year to year highlighting the fields in which progress was traceable. Consequently, a consistent application of the Criteria across case studies under said premises is not the primary goal of the COM's Reports and therefore, unlikely. Hence, the related academic critique may be factually right, but given the goal of the Reports – unjustified.

VIII.3.4 Double Standards between Member States and Candidate Countries

Based on the empirical data gathered in the thesis, no statement can be made as to the potential double standards between Member States and candidate countries.

VIII.3.5 Double Standards between Candidate Countries

The academic critique suggests that countries are being evaluated more or less favorable than others based on how badly the Member States want to include them into the Union. It has been argued that while Croatia had significant support of a large number of Member States, Turkey faced severe opposition among most members, apart from the UK (cf. II.4 Suspicion of Political Bias). If said double standard existed, the Reports should reflect them in terms of differing evaluations from those provided by FH and BTI. Specifically, the score awarded to Croatia by the Commission should be better than those of the other indices and respectively lower for Turkey. This cannot be substantiated through the analysis. In fact, the scores awarded by the Commission to Croatia were slightly worse than those of BTI and FH which

supports the antithesis, namely that the Commission does carry out its evaluations neutrally, at least with view to the case studies analyzed.

VIII.4 Embedding and Critical Reflection of Thesis within Academic Debate

VIII.4.1 Relation to Similar Studies

While there are a number of studies that compare and scrutinize democracy indices', there is only one previously published paper whose conceptual approach resembles the one of this study.⁸⁴ Viljar Veebel (2011) compared the COM's Progress Reports' findings for several Acquis elements to the calculated median scores of six other indices, including FH and BTI for Bulgaria, Croatia, FYROM and Romania between 2004 and 2006. Veebel concluded that he could both substantiate significant and systematic differences between the Reports which he explained with politically biased evaluations by the COM. In comparison to the study at hand, the paper reviews older data for only one of the case studies used for this comparison while spreading its attention to multiple Copenhagen Criteria rather than focusing in-depth on one aspect. This thesis employed a superior analytical and methodological approach than Veebel resulting in deviating results. Methodologically, the thesis took on aspects that limited Veebel's study's reproducibility aiming to not only make this study's results easier to substantiate by other researchers but also provide them with a comprehensive analytical tool to conduct future comparisons. The aspects addressed include the assumed independence of the test indices, the calculation of qualitative into quantitative data as well as causal explanation provided for findings offered by Veebel which were neither embedded in theoretical nor empirical evidence.

This thesis includes an analysis of the indices' independence suggesting their comparability.⁸⁵In order to circumvent the problem of not being able to compare the COM's qualitative findings with the quantitative data of the other indices, Veebel created an ordinal ranking between the case studies based on the language that is used to describe their progress in the Reports. Since the COM itself does use a scaled language depending on the level of progress, this approach at first glance seems appropriate. However, as has been confirmed through the expert interviews, until 2015, the COM neither publicly nor internally provided clear guidelines regarding when to attribute what language. In fact, the comparability of two

⁸⁴ i.e. Bollen (1980,1990), Inkeles (1991), Beetham (1994), Mayer (2008), Coppedge et.al. (2011), Alexander et.al. (2012), Lindberg et.al. (2014), Schmidt (2016)

⁸⁵ Later, through expert interviews, the fulfillment of this very important premise was called into question because a significant amount of sources is shared by and all three indices circle-reference each other.

different countries' assessments prior to the methodological change of 2015 is called into question by a DG NEAR senior official. He explained that if country A progresses from 90 to 95% of fulfilling the Acquis it will only receive the notion "some progress" whereas if country B progressed from 20 to 40% it may receive "substantial progress" (cf. VI.2 Interview Results). Therefore, different linguistic evaluations for two countries can in fact describe the same level of progress while similar linguistic evaluations may in turn not describe the same absolute level of progress. Given the specificities of the Reports, it becomes even more important to provide a detailed coding manual with anchoring examples for inter-coder reliability as well as clear explanations on how scores are being calculated. Such detailed accounts have therefore been provided in the respective Chapters. In order to scrutinize the underlying political bias claim, this study offers and tests a range of theory-led variables and provides additional explanatory aspects derived from the expert interviews. While Veebel provides a mono-causal explanation for his findings (significantly different findings can be explained through political bias on the COM's side). This thesis uncovered that the indices' results are mostly similar and that the slight but systemic differences cannot be traced back to the COM's political bias but are in fact rooted in a number of other (interacting and mutually reinforcing) factors as outlined in the previous sections.

VIII.4.2 Alternative Explanations of the Findings

The indices' results are mostly consistent but show some systematic differences. Explanations other than those offered in the key findings will be introduced at this point. First, the systematic differences found may not be systematic after all but coincidental and based on the fact that the Reports' aims, methods and authors vary sufficiently enough to produce partially differing results. However, there is proof of a clear pattern of FH and BTI showing strongly consistent results with the COM one or two years after her Reports were issues, thus discrediting this alternative explanation.

With view to the COM's Reports, their consistence with other Reports may not be any indicator for their quality at all. In contrast, one could argue that although the facts aggregated may be superior, the entire process of highlighting and interpreting specific events is a reflection of politically biased decisions. This could already play a role at the level of drafting and may be supported by internal instructions or implicit pressures that none of the interviewees of DG NEAR would admit to - given their political sensitivity. However, the general resource-based superiority in range and depth vis-à-vis the other indices is

emphasized by FH and BTI staff and although they could use their discretion to highlight and interpret the same factual findings differently, the analysis shows that this is hardly the case, thus speaking to the high quality of the COM's evaluations.

Another explanation for the findings is mere coincidence: what may be true for Croatia and Turkey may be completely different for other case studies, other test periods, other Acquis elements or other test indices. However, the analysis suggests that a high level of source-based interdependence can be expected of other indices, too, that rely on the COM's findings. Consequently, similar result patterns are to be expected of other indices, too.

Last, the origin of the bias claim may have nothing to do with the COM's double role or its language regime given that at least the general public is not familiar with these aspects. It could simply be rooted in the general distrust of political institutions, the Member States' skepticism of expansion, the candidates' fear of being misjudged or academia's critical approach to anyone carrying out evaluations. Therefore, it may be impossible to be overcome even through a range of detailed policy recommendations. However, the COM personnel itself reflects critically upon its evaluation practices and eventually amended them to become more similar to those of academically accredited indices. This is proof for acknowledging both the issues behind the political bias claim and the belief that it can be addressed which – with the help of the introduced policy recommendations – could be done and communicated even more effectively.

VIII.4.3 The Study's Limitations

There have been some conceptual and methodological limitations to the study. Conceptually, the relevance of the Reports' reliability has been argued to be grounded in their salience for accession decisions. However, EU-accession is a politico-economic decision that is unlikely to be purely merit-based, as the Progress Reports suggest. It remains unclear to what extent political motives and the Progress Reports' findings play into the final accession decision. Given that the Council has always agreed with the COM's accession recommendations – that in turn were based on the Reports which the study proved to be unbiased – this objection may be disregarded at this point.

Also, it became clear during the expert interviews with COM staff that politically appointed officials are openly and unofficially involved in the Report creation process after the initial

level of drafting. However, the extent of their influence in terms of changing the draft remained unclear.

Methodologically, testing the RoL criterion is a good start to test the political bias claim because after ‘Democracy’ it is the most salient accession criteria of the entire Acquis. Nevertheless, there remain numerous other Acquis criteria which can be tested. The same argument is true for the case studies. Both case studies chosen support the notion that the COM’s Reports are not politically biased. Based on the EU’s critical stance regarding accession – Turkey should have received more negative evaluations from the COM than FH and BTI. This was not the case, in fact, the COM’s scores were slightly better. Based on the EU’s public support of Croatia’s accession it should have received better evaluations from the COM than from FH and BTI. This could be substantiated but can be traced back to the positivistic language regime required of the COM’s Progress Reports authors which also led to slightly more positive assessments for Turkey. However, there may exist other candidate countries whose Reports include stronger evidence for or against political bias.

VIII.4.4 Finding’s Implications

The main question of the thesis was how consistent the COM’s Progress Reports are with the findings of other, independent sources. The analysis unfolded that the COM, BTI and FH produced mostly consistent results for Croatia and Turkey within the time period examined. However, their de facto level of independence has to be questioned based on their shared sources, their mutual use of each other for (at least) factual information and the anchoring effect of previous evaluations which causes path dependence. The fact that the degree of independence among the indices is smaller than expected mostly affects the institutions who base their policy and aid decisions on said Reports, at least for the countries who have been evaluated by the COM prior to BTI and FH. For example, FH Reports are used by the Millennium Challenge Corporation (a foreign aid fund separate from USAID), the US State Department, US AID and the US National Security Council who – if politically biased evaluations do exist in the Reports - may unwittingly support European interests by giving or not giving financial or policy support to specific countries. To that extent, it would be highly interesting to conduct further research on the level of interdependence among the different democracy indices given their broadly shared sources.

As has been argued by BTI and FH interviewees, if politically biased evaluations are included into the COM's Reports, this must take place not at the level of drafting but writing and finalizing them. This perception is based on a double standard to the disadvantage of the COM. Both FH and BTI - as well as most other indices - distinguish between the stage of initial drafts by the authors (country experts) and the writing/interpretation/finalization stage conducted by a committee whose goal it is to harmonize the Reports (narratives and scores) across countries, regions or even worldwide. This process is deemed legitimate and necessary for the other indices while considered arbitrary for the COM. However, it can be explained through the nature of the evaluating institutions: while FH and BTI are initially given credit for being independent and trustworthy think tanks, the COM is prejudged as partisan. In fact, it may be within the interest of other indices to convey the picture of politically altered COM Reports in order to justify their own existence and importance. Given that both FH and BTI use the Progress Reports and admit to the COM's superiority in terms of access to resources, why should an institution rely on their Reports if the depth and range of the COM's findings are so much broader? Consequently, academia being critical of the Reports may be both necessary and self-serving.

IX. Summary

Aim and research question

The European Commission annually produces Progress Reports on the EU candidate countries to track and evaluate their progress in meeting the Acquis requirements which are the legal prerequisites for joining the Union. A candidate country's accession decision is supposed to be based on the Reports. Given the clout of the accession decisions, the Reports have been contested on many fronts. The Commission has repeatedly and persistently been accused of producing politically biased Progress Reports on its candidate countries based on certain preferences regarding which country (not) to include. Some propagators of this claim are parties to the accession process. Some candidate countries feel that they have been evaluated unfairly or received less support than others. Some Member States believe that certain candidates were assessed too complaisant compared to their actual developments. Others, including a number of academics, have engaged in offering a wide range of critique as well.

The thesis aimed to address and answer the question whether the political bias claim can be substantiated. It developed a line of argumentation based on which a Conceptual and Methodological Framework was created. The main idea was that politically biased evaluations by the COM should automatically lead to different assessment results than those of independent institutions who evaluate the same countries using the same parameters. Consequently, two independent democracy measuring indices were chosen as test indices (Freedom House and Bertelsmann Transformation Index). The countries used as case studies were Croatia and Turkey because they both started their accession negotiations with the EU in the same year (2005) and were attributed similar shortcomings, particularly with view to one salient Acquis-criterion: the RoL. Given that their accession negotiations took part during the same time period, aspects that could otherwise explain different evaluations (e.g. the Union's general stance towards accession) could be ruled out. While Croatia was finally admitted into the EU in 2013, Turkey's accession has become even more unlikely given Erdogan's recent restrictions of the RoL due to the alleged military coup of July 2016 (Rankin, 2016). Consequently, the COM's Progress Reports on both countries from 2005-2012 (year of the last report produced on Croatia) were chosen and their RoL assessments were compared to those of Freedom House's Freedom in the World and Bertelsmann's Bertelsmann Transformation Index.

The structure of the thesis

The thesis first set out to provide the reader with a comprehensive overview of external democracy promotion as is carried out by the European Commission. Understanding the motives, underlying concepts and methods of external democratization actors (specifically the EU's enlargement) helped to contextualize the meaning and creation of democracy measuring indices. All indices used throughout the analysis were examined in detail and tested for their academic quality using Munck/Verkuilen's analytical framework. Additionally, the reader was presented with a summary of the academic debate surrounding the RoL concept, thus demonstrating both the magnitude of said criteria for a country's democratic development as well as the risks of measuring two completely different things while using the same name for it. After the elaborate theoretical foundation, the thesis focused on the actual analysis and comparison of the indices' results for Croatia and Turkey. The comparison aimed to detect significant and systematic inconsistencies among the Reports which would build the premise for the political bias hypotheses. After discovering that the results are mainly consistent, it continued to examine the reasons for similar evaluations, both conceptually developed and with the help of explorative in-depth expert interviews and internal COM documents. The interviews unearthed a number of institutional factors that explain the origin of the political bias perception. Taking into account the insights of the analysis, the thesis concludes with a set of policy recommendation addressing measures fighting the political bias perception.

Design and Methods

The thesis is based on a range of qualitative methods (inductive and deductive qualitative content analysis, scaled structuring qualitative content analysis, semi-structured, in-depth expert interviews). This choice lies in the nature of the research question and the material used to answer it. Political bias in itself is a phenomenon that cannot be quantified but rests on perceptions regarding the Progress Reports' production process, language and imagery. It also depends on both its' authors' and consumers' potential underlying values or goals. The complexity of the phenomenon can only be examined qualitatively. Since the material used consists of written Reports, a qualitative approach had to be chosen before quantifying the data on one, common scale.

Procedurally, first the indices were examined for their shared RoL criteria to build a foundation for the comparison using the inductive qualitative content analysis. Based on these four criteria (Judicial Independence, Efficiency, Crime, Rights and Liberties) the indices'

results were compared using the structuring, scaled qualitative content analysis. This analysis discovered that all indices have – surprisingly - mainly consistent findings for the respective case studies, thus destabilizing the political bias claim. The indices’ results’ consistency was then elevated to become the second Dependent Variable. Subsequently, conceptually developed Independent Variables as well as interview-gathered opinions and explanations from the index experts were examined. These were contextualized with the help of internal Com documents that that were mentioned during the expert interviews and later accessed by the author. Based on these findings, policy recommendations were constructed which are presented at the end of this chapter.

The two test indices used to conduct the comparison are FH’s Freedom in the World and Bertelsmann’s Transformation Index. Their selection was based on their shared underlying value system (Western liberal democracy) and similar methodology while remaining institutionally independent from the EU and the accession process. While a preliminary test of the indices supported the notion of independence, the content analysis in combination with the expert interviews unearthed a surprisingly high level of interdependence among them. Croatia and Turkey have been chosen as case studies because they both started their accession negotiations in the same year and both suffered from similar RoL deficiencies in the beginning. Moreover, they are the candidates most likely to be affected by politically biased evaluations given the Member States’ strong support for Croatia’s accession as opposed to the Member States’ strong aversion against Turkey’s inclusion. Had such bias existed in the past, this should have been reflected in the COM’s evaluations. For the in-depth expert interviews, representatives of each institution have been chosen who have either authored the respective Reports themselves or coordinated them as part of the harmonizing procedure before its finalization⁸⁶.

Key themes identified & categories of response

The key finding of the comparative analysis was that politically biased evaluations by the COM could not be substantiated. In fact, the Progress Reports’ results are mainly consistent with those of FH and Bertelsmann for both countries (although there is a slightly higher

⁸⁶Two members of the political desks of the COM DG NEAR as well as their coordinator and negotiation-leader were interviewed on the Commission’s part. Two regional coordinators from BTI as well as one from FH were interviewed.

consistency for Turkey than for Croatia). Their consistency can be attributed to a range of factors.

First, all three indices share the majority of their key sources. Secondly, the indices consistently circle-reference each other. While FH and BTI strongly relying on the COM's Reports for their factual elements, the COM cross-checks its results with those of the others. Third, all three indices base their evaluations on their previous year's scores and assessments. This is done to ensure coherence but eventually leads to a high degree of path dependence in their ratings. Consequently, the indices' evaluations remain very constant individually and through circle-referencing, converge over time.

Other explanative aspects mentioned by the experts include adequate assessment, similar methods and overlapping RoL conceptualizations. It was argued that the evaluated elements are mostly objectively ascertainable (e.g. adoption of a law, creation of an institution). Additionally, it was emphasized that all three indices shared the same understanding of democracy and the RoL. Furthermore, it was noted that using similar methods to carry out the evaluation eventually leads to a similar interpretation of factual information. While these aspects were anticipated by the author, they could not be tested as Independent Variables. The 'adequate assessment' of a case study cannot be evaluated unless a new, independent analysis was conducted which is impossible resource-wise.⁸⁷ The thesis also proved that both the selection of these sources as well as their actual level of independence must be called into question. If all indices compared used very different sources and still produced consistent results, this explanation could withstand. However, the tested indices' most important resources overlap wherefore the adequate assessment claim cannot be tested. As for the similar RoL conceptualization and methodology, these variables are not independent. In fact, the indices were chosen because their similarities make them comparable. Consequently, claiming that their results are consistent because of their similarity would constitute circular reasoning.

Dissimilarities between the results can be attributed to different ranges of resources available to the institutions, diverging interpretations of factual elements and varying aims of the indices that lead to their deviating interpretation of the same findings.

⁸⁷ Even if it were possible, it would be prone to the same potentially bias-causing theoretical and methodological decisions as have been criticized in the tested indices. One could argue that by including a range of independent sources, the indices get as close to 'the scientific truth' as possible.

Systematic inconsistencies can be attested to FH's and the BTI's use of earlier COM Reports which lead to time-shifts in their evaluations. Evaluations by the COM repeatedly resemble scores given by FH one year later, respectively two years later in the case of BTI (which is published only every two years).

Furthermore, although the experts from the other indices were aware of the political bias claim and did believe that politically motivated interpretations of the findings were possible, they trusted the general reliability of the COM's Reports' factual elements based on the professionalism of the Country Units and the vast range of resources available to the Commission in gathering this information. Although the DG NEAR officials conceded that the process is prone to mistakes just like any other evaluation process carried out by humans, its procedural net is so tight-knit that subjective elements are reduced and outbalanced in the harmonization process. Given the structured and procedural style of the officials interviewed equaling that of the political scientist at the other indices, the impression was gained that - at the level of report drafting -biased evaluations are rather unlikely.

However, a comprehensible explanation for the persisting political bias perception was provided by the interviewees pertaining to the COM's double role as both assessor and distributor of financial development support. The longer a staffer oversees a candidate, the higher the risk of personal subjectivity grows in wanting a country to succeed. Additionally, the unpublished guidelines on how DG NEAR ENLARGEMENT officials should linguistically write their Reports stipulates to name positive developments first to congratulate and incentivize the candidate countries for further progress. This is based on the general progress-focus (as opposed to overall-picture) which can lead to a contorted perception of how good the overall performance of the country really is. The same problem arises from the fact that it is much easier for a country to show significant progress in the beginning of the process than at a later stage. As a DG NEAR ENLARGEMENT official explained, if a country progresses from 90 to 95% of Acquis fulfillment, it linguistically is evaluated with less progress than a country moving from 20 to 40% of Acquis adoption. Since what is being assessed is both the adoption/creation of institutions and their implementation/functionality, countries can initially easily adopt laws and create institutions, thus receiving positive evaluations. Making those institutions sustainably functional is much harder and usually leads to less enthusiastic Reports the longer the process continues. Consequently, the structural and

situational set-up of the Reports is crucial in explaining why its results are perceived as arbitrary.

Research contribution

As has been explained in the **Literature Review**, the RoL is a widely debated and contested topic with neither a single definition to be found in academia nor used by the European Commission. Although there is no single codified RoL concept, the European Commission de facto used certain benchmarks during the time of analysis for its accession evaluations. These have been outlined in internal documents (Guidance Notes) and “fleshed out” by reviewing Directives, Opinions and other formal writings issued by EU institutions on the matter. A synopsis of said documents was created to present the reader with a comprehensive EU accession RoL concept. Thus, this thesis contributed to a better understanding of *what* exactly is being tested by the Commission. Moreover, the Commission has been criticized repeatedly for the intransparency of its evaluation process. In order to shed light on the previously unpublicized *how* of Progress Report creation process, in-depth expert interviews with EU DG NEAR (Enlargement) officials were conducted. Furthermore, the Commission’s underlying RoL concept as well as its method were compared to those of independent, academically-based democracy indices. All three of them were analyzed based on their academic quality so as to uncover each index’ strengths and weaknesses. By using the scaled, structuring qualitative content analysis, one common scale was created on which the results for each case study’s RoL elements for each respective year could be measured and compared. This comparison showed that – unexpectedly – the scores were mostly consistent, thus disproving the political bias accusation vis-à-vis the Commission. The few (systematic) inconsistencies were analyzed as to their origin. With the help of internal documents and in-depth expert interviews, aspects that enable and strengthen the political bias perception could be identified.

Future Research

The thesis uncovered some aspects that comprehensibly explain why there is such a strong perception of politically biased assessments. These aspects include the intransparency of the COM evaluation procedures, the informal consultations with Council staff and the involvement of politically appointed officials in the finalization of the Reports. More academic scrutiny could be applied to these aspects and their actual influence in altering the original drafts. For example, one could detect and interview politically appointed officials

finalizing the Reports as well as participants of unofficial meetings. Comparing the COM's draft Reports to the final versions could also be highly interesting as well as comparing reports created under the old guidelines to those drafted after 2015 based on the new, more-structured regulations. Also, one could compare the COM's pre-2015 and post 2015-scores to those of truly independent indices to see whether they converge even stronger after introducing the option of negative assessment language.

Another focus for future studies should be the independence of democracy measuring indices. Previous research has focused either on the quality of their methodology or their ideological independence from their supporters. This thesis proved the high level of interdependence between the COM, FH and BTI. Consequently, it should be scrutinized to what extent supposedly independent indices may be influenced by other, more resourceful ones and to what degree circle-referencing causes streamlined evaluations among them.

Policy recommendations

The analysis discovered a range of problems with the progress report creation process which should (and since 2015 have started to be) addressed by the COM in order to fight the political bias perception. These problems are the lack of a codified RoL concept used as a benchmark for the evaluation, the intransparency of the methodological procedure carried out by the DG NEAR officials, the informal meetings held with the Member States' representatives prior to the report finalization, the involvement of politically appointed officials in finalizing the Reports, and the double role of the European Commission in both evaluation and distributing financial development support to the candidate countries. The following four policy recommendations address these issues.

Policy Recommendation 1: Create a codified RoL benchmark

As outlined in the beginning, the Member States have valid reasons for not choosing to codify what specifically they understand to be the RoL. This is due to their own national versatility in shaping their RoL, the fear to be held accountable to a certain standard they themselves may not fulfill as well as the trepidation of creating binding legislation that may not include everything necessary in adequately evaluating a country's RoL situation. The latter can be attributed to changing realities which may require attention in areas that have previously been of less importance (e.g., corruption, human trafficking, refugees). However, the clearer and more detailed the concept becomes, the more likely an institution can produce reliable and comparable evaluations. This is why all other indices have specific sets of questions their

authors must adhere to. As has been voiced by a DG NEAR official, there is some fear that once clear-cut RoL criteria are introduced, the candidate countries will only focus on those while disregarding other important aspects. However, it can be argued that it is the job of the EU institutions to create an adequate RoL checklist that is sufficiently inclusive to address that issue.

This aspect has to some extent been acknowledged and rectified by the COM. In 2015 a new codification mechanism was introduced by the COM which is not available to the general public. It consists of specific guidelines, a certain structure and more detailed checklists for sectors of particular importance (the RoL, anti-corruption, organized crime, public administration) with which the level of a candidate's engagement/alignment is being assessed. For example, the functioning of the judiciary is measured based on specific statistics. In 2016, additional sectors were included into the new mechanism which focus on the economic fundamentals of accession. The mechanism is not supposed to be broadened to all Chapters in order to keep candidates from focusing on "side issues" while neglecting the "fundamentals". The introduction of the new mechanism is multi-causal. The new Commissioner took into account both external critique for lack of transparency as well as the "changing realities of accession" (all quotes DG NEAR official) in comparison to the previous enlargement rounds. While the old system was suitable for quick accessions the new one is targeted at the Western Balkans and Turkey, where both developmental pro- and regress have to be taken into account and where less "quantum leaps" can be expected than with the 2004 candidates, said a DG NEAR official. Consequently, negative language options ("no progress", "backslide") have been introduced to do their developments justice both in the case of pro- and regress.

Policy Recommendation 2: Publish evaluation process and criteria

People distrust things they do not understand. This truism applies to politics and academia as well. Decisions that are made behind closed doors are expected to be untrustworthy. Methods that are insufficiently disclosed are expected to hide faults. Consequently, if a researcher does not or insufficiently describes his methodological process, his results will be questioned. Even indices like Freedom House and Bertelsmann - which release very detailed accounts of their analyses - are being criticized either for the omission of disclosing specific steps or the lack of explanation for why aspects are interpreted or calculated in a certain way. The European Commission has published very little on its progress report creation process, although – as became evident in the expert interviews – these things are "not really secret" (DG NEAR

official) and can be accessed through the European Commission's Secretariat-General Transparency upon request (SG.B.4). If there is no need to protect the information because they adhere to academic standards – as the author's impression was after interviewing the officials – then they should be released to the public in order to enable professionals of the respective fields to methodologically review them. Moreover, this would send a clear signal to the candidate countries and the general public that no secrecy is desired.

Policy Recommendation 3: Limit the influence of politically appointed officials or their staff

Based on the same logic as the section above, informal meetings, especially with people who are known to have vested interests, are not helping the Reports' credibility. The voluntary and informal offer of additional information on candidate countries by Member States exemplifies such objectionable procedures. The consultation of Member States' representatives or candidate countries' Ambassadors in informal meetings may be purely informational but it provides a futile breeding ground for bias claims. Ideally, these meetings should be made formal or attended by neutral third parties confirming that their content has not altered the Reports findings inadequately. The adequacy-restriction is based on the fact that sometimes the COM's staff is presented with new factual information (such as the adoption of a new law) which then could legitimately alter the report.

More importantly though, politically appointed officials are part of the formal report harmonizing process, such as the Directorate General, and the Commissioners forming the College of Commissioners. Since the accession is a political decision based on political interests and motives, the inclusion of politically appointed officials in finalizing the Reports is comprehensible and necessary, maybe even eligible in terms of democratic legitimacy. However, two preconditions must be met: First, it needs to be secured that these Commissioners are as neutral as possible vis-à-vis the candidate country - which is not self evident given conflicts between Member States and candidate countries, such as Greece and Turkey over Cypriot territory. Secondly, it must be ensured that certain Commissioners – maybe due to the power of their country of origin – do not assert a much greater level of influence than do others. Given political realities, it is acknowledged by the author that both of these recommendations are difficult to enforce.

Policy Recommendation 4: Dissolve the COM's double role

A parent should not be its child's teacher because it is likely to not treat and assess it equally to its class mates because of his personal involvement. The COM is somewhat facing the

same dilemma when it distributes IPA funding to the very countries it evaluates. As has been recognized by a DG NEAR official, if a Country Unit is working closely with the candidate country, it is more likely to want to see progress even if a neutral observer may not agree with the assessment. In that case, he may want to support and incentivize the candidate ('carrot') where punishment ('stick') would have been more in line with the official conditionality principle. Although the Political Desks do not make the funding decisions, they consult with one another and award cash injections to countries lagging behind their reform schedules. Thus, the degree of subjective elements included in the decision-making process remains unclear. In order to return to the actual 'carrot and stick' approach - rewarding successful reformers and punishing reluctant ones rather than handing out 'pity payments' - it is important to ensure that the people in charge of the financial support are widely independent from the Country Units.

Conclusion

When assessing corruption, there is a major difference between the level of actual, provable corruption and perceived corruption (as measured through Transparency International). In many countries, the level of perceived corruption is much higher than what can be proved. On the one hand, there is an unspecific dark figure due to the nature of corruption. On the other hand, people gravely overestimate the degree of criminal energy of the authorities once they decided to distrust them. The same is true for political bias, particularly in the European Union. Due to its nature of complexity and opacity, it is a phenomenon that cannot be easily examined or conclusively attested to the COM's Progress Reports. However, the basic understanding of all actors who criticize the Reports (candidate countries, Member States, academia, representatives of other indices) is the intuitively sound logic: because the European Union is a political institution all its decisions must be politically motivated at the expense of neutrality, meritocracy and fairness. This notion is supported by a number of institutional decisions regarding the creation of the COM's Progress Reports. The Reports have been and - to some extent even after the methodological changes of 2015 - still are lacking clear criteria or rules on how to evaluate the RoL. Additionally, the positivistic language regulations used until 2015 enabled misleading interpretations of events which other indices do not share. Moreover, the COM holds a double role as both assessor and sponsor of development to the candidates, thus enabling the perception of individual and institutional biased evaluations. Most importantly though, the unspecified and unregulated inclusion of

politically appointed figures before finalizing the Reports leaves plenty of room for speculation regarding their influence on the Reports' outcomes.

However, no proof of such political bias could be found in the analysis at hand. Consequently, the main implication of the findings is that it is within the power and discretion of the COM to both include politically biased assessments in the Progress Reports as well as to change the perception of political bias by making the evaluation process more consistent, transparent and reliable.

In order to fully understand the political bias dilemma, the old serenity prayer is helpful. It seeks the serenity of mind to accept things which cannot be changed; the courage to change things which can be changed, and the wisdom to know the one from the other. As this thesis laid out, the political bias perception stems from a combination of psychological factors and practical decisions surrounding the Reports' creation. What cannot be changed and must be accepted is that there will always be actors or institutions (e.g. candidates) who feel skeptical about decisions made by political institutions they do not fully trust (e.g. the COM). Thus, the psychology of the political bias perception is unlikely to disappear. What can be changed though, is the report creation process, particularly after its weaknesses resulting in political bias claims have been identified. Those include more methodological transparency, less institutional linkages to political actors as well as less structurally overlapping activities within the Commission. This may not erase all doubts but should soften the degree of critique the Commission is facing regarding the reliability of its Reports.

Luckily, the COM acknowledged the political bias perception dilemma and reacted to it by introducing serious methodological changes in 2015. However, while the perception of political bias can be softened through these changes it will never be fully alleviated by improving the evaluation method because "there is no such thing as a perfect evaluation tool" (FH official). As the same FH representative put it, all indices "take a lot of subjective material and thus make it as objective as possible" which always leaves room for discretion and explains why all other democracy measuring indices have also been criticized for their methodological choices. The goal of the European Commission must therefore be to better

communicate its own process to the public, particularly the parties involved in the immediate accession process, in order to fight the political bias perception.

Bibliography

- Abreu, J M (2001): „Theory and Research on Stereotypes and Perceptual Bias: A Didactic Resource for Multicultural Counseling Trainers.“ *The Counseling Psychologist*, 29(4), pp.487-512.
- Alexander, A C, et. al. (2012). “Measuring effective democracy: A defense“. *International Political Science Review* 33 (1), pp. 41-62.
- BBC (2010): „BBC News - Eurozone approves massive Greece bail-out“. [News.bbc.co.uk](http://news.bbc.co.uk/2/hi/business/8656649.stm). Accessed 22. 07. 2015, Available: <http://news.bbc.co.uk/2/hi/business/8656649.stm>
- Beetham, D, ed. (1994): „Defining and measuring democracy“. Vol. 36. Sage
- Benvenuti, S (2011): “Judicial Training in Turkey in Light of Constitutional Traditions and Europeanization.” *Journal of Civil Law Studies Volume* 4(2): 309–30.
- Bertelsmann Stiftung (2014): „BTI Methodology“. Available: <http://www.bti-project.org/en/index/methodology/>
- Bertelsmann Stiftung (2012): „BTI 2012 | Turkey Country Report“. Available: http://www.bti-project.org/fileadmin/files/BTI/Downloads/Reports/2012/pdf/BTI_2012_Turkey.pdf.
- Bertelsmann Stiftung (2010): „BTI 2010 | Turkey Country Report“. Available: http://www.bti-project.org/fileadmin/files/BTI/Downloads/Reports/2010/pdf/BTI_2010_Turkey.pdf.
- Bertelsmann Stiftung (2008): „BTI 2008 | Turkey Country Report“. Available: http://www.bti-project.org/fileadmin/files/BTI/Downloads/Reports/2008/pdf/BTI_2008_Turkey.pdf.
- Bertelsmann Stiftung (2006): „BTI 2006 | Turkey Country Report“. Available: http://www.bti-project.org/fileadmin/files/BTI/Downloads/Reports/2006/pdf/BTI_2006_Turkey.pdf.
- Bertelsmann Stiftung (2010): „BTI 2010 | Croatia Country Report“. Available: http://www.bti-project.org/fileadmin/files/BTI/Downloads/Reports/2010/pdf/BTI_2010_Croatia.pdf.
- Bertelsmann Stiftung (2008): „BTI 2008 | Croatia Country Report“. Available: http://www.bti-project.org/fileadmin/files/BTI/Downloads/Reports/2008/pdf/BTI_2008_Croatia.pdf.
- Bertelsmann Stiftung (2003): „BTI 2006 | Croatia Country Report“. Available: https://www.bti-project.org/fileadmin/files/BTI/Downloads/Reports/2006/pdf/BTI_2006_Croatia.pdf.

- Bertelsmann Stiftung (2012): „BTI 2012 | Croatia Country Report“. Gütersloh. Available: http://www.bti-project.org/fileadmin/files/BTI/Downloads/Reports/2012/pdf/BTI_2012_Croatia.pdf
- Blitz, BK (2008): “Democratic Development, Judicial Reform and the Serbian Question in Croatia.” *Human Rights Review* 9(1): 123–35.
- Bollen, K A (1993): “Liberal Democracy: Validity and Method Factors in Cross-National Measures.” *American Journal of Political Science* 37(4): 1207–30. Available: http://www.irss.unc.edu/content/pdf/Bollen_1993_AJPS.pdf
- Bollen, K A (1980): „Issues in the comparative measurement of political democracy“. *American Sociological Review* (): pp. 370-390.
- Bollen, K A (1990): „Political democracy: Conceptual and measurement traps“. *Studies in Comparative International Development*, 25(1), pp.7-24.
- Boudling, K (1958): “National Images and International Systems.” *Conflict Resolution* 3(2): 120–31.
- Burnell, P (2008): “From Evaluating Democracy Assistance to Appraising Democracy Promotion.” *Political Studies* 56(2): 414–34. Available: <http://doi.wiley.com/10.1111/j.1467-9248.2007.00653.x> (December 30, 2014).
- Butler, I (2014): „Policy Brief“. Available: <https://www.opensocietyfoundations.org/sites/default/files/how-monitor-rule-law-democracy-and-fundamental-rights-eu.pdf>
- Carothers, T (1998): “The Rule of Law Revival.” *Foreign Affairs* (March/April): 95–106. Available: <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=935b72e4-8c86-4e7e-9def-19f7118a839a%40sessionmgr198&hid=124>.
- Cengiz, F,; Hoffmann, L (2013): “Turkey and the European Union.” *Turkey and the European Union: Facing New Challenges and Opportunities* 23(1): 1–232. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84917450682&partnerID=tZOtx3y1>.
- Comer, R; Gould, E (2010): „Psychology around us“. John Wiley & Sons.
- Committee of Ministers (2000): „Recommendation No. R (2000) 10 of the Committee of Ministers to Member States on Codes of Conduct for Public Officials“. Available: <https://wcd.coe.int/ViewDoc.jsp?id=353945&Site=CM&BackColorInternet=9999CC&BackColorIntranet=FFBB55&BackColorLogged=FFAC75>.
- Coppedge, M, et al. (2011): „Conceptualizing and measuring democracy: A new approach“. *Perspectives on Politics* 9 (2), pp. 247-267.

- Coppedge, M, et al. (2011): “Conceptualizing and Measuring Democracy: A New Approach.” *Perspectives on Politics* 9(02): 247–67. Available: http://www.journals.cambridge.org/abstract_S1537592711000880 (July 11, 2014).
- Council of Europe: Committee of Ministers (1994): „RECOMMENDATION No. R (94) 12 OF THE COMMITTEE OF MINISTERS TO MEMBER STATES ON THE INDEPENDENCE, EFFICIENCY AND ROLE OF JUDGES“. Available: http://www.judicialcouncil.gov.az/Law/eAS_NK_12sayli.pdf.
- Crescenzi, M, et al. (2015): “Reliability, Reputation, and Alliance Formation1.” *International Studies Quarterly* 56(2): 259–74. Available: <http://doi.wiley.com/10.1111/j.1468-2478.2011.00711.x> (July 20, 2015).
- Daugirdas, K (2015): “Reputation and the Responsibility of International Organizations.” *European Journal of International Law* 25(4): 991–1018. Available: <http://ejil.oxfordjournals.org/cgi/doi/10.1093/ejil/chu087> (May 11, 2015).
- Donnell, G (2004): “WHY THE RULE OF LAW MATTERS.” *Project Muse* 15(4): 32–46. Available: http://home.comcast.net/~lionelingram/odonnell_WhyRuleOfLawMatters.pdf.
- Doucouliagos, C (1995): „Institutional bias, risk, and workers’ risk aversion“. *Journal of Economic Issues* 29(4), pp. 1097-1118.
- Downs, G; Jones, M (2003): “The University of Chicago The University of Chicago Law School.” *Political psychology* 24(3): 449–68.
- Eldem, T (2005): “The End of Turkey’s Europeanization?” *Turkish Policy Quarterly* 12(1): 125–36.
- Escritt, T; Deutsch, A (2014): „Dutch rightist Wilders blames EU for Ukraine crisis; hints at UKIP alliance“. *Reuters UK*. Accessed 06. 02. 2015. Available: <http://uk.reuters.com/article/2014/04/17/uk-eu-election-dutch-wilders-idUKBREA3G1CT20140417>
- EurActiv (2013): „EU to reopen accession talks with Turkey on November 5“. *EurActiv.com*. Accessed 25. 11. 2014. Available: <http://www.euractiv.com/video/eu-reopen-accession-talks-turkey-november-5-9232>
- EUR LEX, (o. J.): „Suspension Clause“. *Eur-lex.europa.eu*. Accessed 20. 11. 2014. Available: http://eur-lex.europa.eu/summary/glossary/suspension_clause.html?locale=en
- European Commission (2016): „Agenda 2000 - For a stronger and wider Union. Document drawn up on the basis of COM (97) 2000 final, 13 July 1997. Bulletin of the European Union, Supplement 5/97 - Archive of European Integration“. *Aei.pitt.edu*. Accessed 13. 05. 2015. Available: <http://aei.pitt.edu/3137/>
- European Commission (2015a): „Draft Reporting Guidelines for 2015 Readiness/Progress Reports: Public Administration Reform“. [Internal document included in Annex]

European Commission (2015b): „Draft Reporting Guidelines for 2015 Readiness/Progress Reports: Justice Reform“. [Internal document included in Annex]

European Commission (2015c): „*REPORTING GUIDELINES – ORGANISED CRIME*“. [Internal document included in Annex]

European Commission (2015d): „*Draft Reporting Guidelines for 2015 Readiness/Progress Reports: Freedom of Expression*“. [Internal document included in Annex]

European Commission (2015e): „*Enlargement Package 2015 Guidance Note*“. [Internal document included in Annex]

European Commission (2014): „*Press Release: European Commission Presents a Framework to Safeguard the Rule of Law in the European Union*“. Brussels. Available: http://europa.eu/rapid/press-release_IP-14-237_en.htm

European Commission (2013): „COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL Enlargement Strategy and Main Challenges 2013-2014“. Brussels. Available: http://ec.europa.eu/enlargement/pdf/key_documents/2013/package/strategy_paper_2013_en.pdf

European Commission (2012a): „*Comprehensive Monitoring Report on Croatia (2012)*“. Brussels. Available: http://ec.europa.eu/enlargement/pdf/key_documents/2012/package/hr_rapport_2012_en.pdf

European Commission (2012b): „*TURKEY 2012 PROGRESS REPORT*“. Brussels. Available: http://ec.europa.eu/enlargement/pdf/key_documents/2012/package/tr_rapport_2012_en.pdf

European Commission (2012c): „COMMISSION STAFF WORKING DOCUMENT COMPREHENSIVE MONITORING REPORT ON CROATIA“. Brussels. Available: http://ec.europa.eu/enlargement/pdf/key_documents/2012/package/hr_rapport_2012_en.pdf

European Commission (2012d): „*Enlargement Package - Autumn 2012 Guidance Note*“. [Internal document included in Annex]

European Commission (2011a): „*Guidance Note*“. [Internal document included in Annex]

European Commission (2011b): „*CROATIA 2011 PROGRESS REPORT*“. Available: http://ec.europa.eu/enlargement/pdf/key_documents/2011/package/hr_rapport_2011_en.pdf.

European Commission (2011c): „*TURKEY 2011 PROGRESS REPORT*“. Brussels. Available:

http://ec.europa.eu/enlargement/pdf/key_documents/2011/package/tr_rapport_2011_en.pdf

European Commission (2010a): „*CROATIA 2010 PROGRESS REPORT*“. Available:
http://ec.europa.eu/enlargement/pdf/key_documents/2010/package/hr_rapport_2010_en.pdf

European Commission (2010b): „*Guidance Note*“. [Internal document included in Annex]

European Commission (2010c): „*TURKEY 2010 PROGRESS REPORT*“. Brussels.
Available:
http://ec.europa.eu/enlargement/pdf/key_documents/2010/package/tr_rapport_2010_en.pdf

European Commission (2009a): „*Guidance Note*“. [Internal document included in Annex]

European Commission (2009b): „*CROATIA 2009 PROGRESS REPORT*“. Available:
http://ec.europa.eu/enlargement/pdf/key_documents/2009/hr_rapport_2009_en.pdf.

European Commission (2009c): „*TURKEY 2009 PROGRESS REPORT*“. Brussels.
Available:
http://ec.europa.eu/enlargement/pdf/key_documents/2009/tr_rapport_2009_en.pdf

European Commission (2008a): „*Guidance Note*“. [Internal document included in Annex]

European Commission (2008b): „*CROATIA 2008 PROGRESS REPORT*“. Available:
http://ec.europa.eu/enlargement/pdf/press_corner/key_documents/Reports_nov_2008/croatia_progress_report_en.pdf.

European Commission (2008c): „*TURKEY 2008 PROGRESS REPORT*“. Brussels.
Available: http://ec.europa.eu/enlargement/pdf/press_corner/key_documents/Reports_nov_2008/turkey_progress_report_en.pdf.

European Commission (2007a): „*Guidance Note*“. [Internal document included in Annex]

European Commission (2007b): „*CROATIA 2007 PROGRESS REPORT*“. Available:
http://ec.europa.eu/enlargement/pdf/key_documents/2007/nov/croatia_progress_Reports_en.pdf.

European Commission (2007c): „*TURKEY 2007 PROGRESS REPORT*“. Brussels.
Available:
http://ec.europa.eu/enlargement/pdf/key_documents/2007/nov/turkey_progress_Reports_en.pdf

European Commission (2006a): „*Autumn Package 2006 Progress Reports Guidance Note*“. [Internal document included in Annex]

- European Commission (2006b): „CROATIA 2006 PROGRESS REPORT“. Available: http://ec.europa.eu/enlargement/pdf/key_documents/2006/nov/hr_sec_1385_en.pdf.
- European Commission (2006c): „TURKEY 2006 PROGRESS REPORT“. Brussels. Available: http://ec.europa.eu/enlargement/pdf/key_documents/2006/nov/tr_sec_1390_en.pdf
- European Commission (2005a): „2005 Regular Reports Guidance Note“. [Internal document included in Annex]
- European Commission (2005b): „CROATIA 2005 PROGRESS REPORT“. Available: http://ec.europa.eu/enlargement/archives/pdf/key_documents/2005/package/sec_1424_final_progress_report_hr_en.pdf
- European Commission (2005c): „TURKEY 2005 PROGRESS REPORT“. Available: http://ec.europa.eu/enlargement/archives/pdf/key_documents/2005/package/sec_1426_final_progress_report_tr_en.pdf
- European Commission (2005d): „GUIDE TO THE MAIN ADMINISTRATIVE STRUCTURES REQUIRED FOR IMPLEMENTING THE ACQUIS“. [Internal document included in Annex]
- European Commission (1997): „AGENDA 2000 For a Stronger and Wider Union“. Brussels. Available: http://ec.europa.eu/agriculture/cap-history/agenda-2000/com97-2000_en.pdf
- European Parliament (2012): „Hungary: The EP should make sure there are no double standards“. European Parliament. Accessed 17. 01. 2015. Available: <http://www.europarl.europa.eu/news/en/news-room/content/20120928STO52499/html/Hungary-The-EP-should-make-sure-there-are-no-double-standards>
- Europäischer Rat (1993): “Europäischer Rat Kopenhagen: Schlussfolgerungen Des Vorsitzes.” Kopenhagen. Available: <http://www.pascal-hector.de/Vertiefung-Erweiterung-Dateien/06-Kopenhagen-Kriterien.pdf>
- European Union: The Council (1998): „1998 Draft Conclusions of the Council on Common Principles for the Rule of Law in Criminal Matters“. Brussels. Available: <http://database.statewatch.org/e-library/10116.en98-rule-of-law.pdf>.
- Bergling, P, et al. (2008): „Rule of Law in Public Administration: Problems and Ways Ahead in Peace Building and Development“. Folke Bernadotte Academy Publications. Available: https://fba.se/globalassets/rule_of_law_in_public_administration.pdf
- Freedom House (2014): „Freedom in the World 2014 Methodology“. Available: <https://freedomhouse.org/report/freedom-world-2014/methodology>
- Freedom House (2012): „Freedom in the World – Croatia“. Available: <https://freedomhouse.org/report/freedom-world/2005/croatia>.

Freedom House (2012): „Freedom in the World – Turkey“. Available:
<https://freedomhouse.org/report/freedom-world/2012/turkey>.

Freedom House (2011): „Freedom in the World – Croatia“. Available:
<https://freedomhouse.org/report/freedom-world/2011/croatia>.

Freedom House (2011): „Freedom in the World – Turkey“. Available:
<https://freedomhouse.org/report/freedom-world/2011/turkey>.

Freedom House (2010): „Freedom in the World – Croatia“. Available:
<https://freedomhouse.org/report/freedom-world/2010/croatia>.

Freedom House (2010): „Freedom in the World – Turkey“. Available:
<https://freedomhouse.org/report/freedom-world/2010/turkey>.

Freedom House (2009): „Freedom in the World – Croatia“. Available:
<https://freedomhouse.org/report/freedom-world/2009/croatia>.

Freedom House (2009): „Freedom in the World – Turkey“. Available:
<https://freedomhouse.org/report/freedom-world/2009/turkey>.

Freedom House (2008): „Freedom in the World – Croatia“. Available:
<https://freedomhouse.org/report/freedom-world/2008/croatia>.

Freedom House (2008): „Freedom in the World – Turkey“. Available:
<https://freedomhouse.org/report/freedom-world/2008/turkey>.

Freedom House (2007): „Freedom in the World – Croatia“. Available:
<https://freedomhouse.org/report/freedom-world/2007/croatia>.

Freedom House (2007): „Freedom in the World – Turkey“. Available:
<https://freedomhouse.org/report/freedom-world/2007/turkey>.

Freedom House (2006): „Freedom in the World – Croatia“. Available:
<https://freedomhouse.org/report/freedom-world/2006/croatia>.

Freedom House (2006): „Freedom in the World – Turkey“. Available:
<https://freedomhouse.org/report/freedom-world/2006/turkey>.

Freedom House (2005): „Freedom in the World – Croatia“. Available:
<https://freedomhouse.org/report/freedom-world/2005/croatia>.

Freedom House (2005): „Freedom in the World – Turkey“. Available:
<https://freedomhouse.org/report/freedom-world/2005/turkey>.

Fuchsberg, J (2011): “The Copenhagen Criteria: Are They Helping or Hurting the European Union?” *Touro International Law Review* 14(2): 390–411.

- Gateva, E (2010): „Post-Accession Conditionality Support Instrument for Continuous Pressure?“ 1. Aufl. Berlin: KFG.
- Giannone, D (2010): “Political and Ideological Aspects in the Measurement of Democracy: The Freedom House Case.” *Democratization* 17(1): 68–97.
<http://www.tandfonline.com/doi/abs/10.1080/13510340903453716> (January 18, 2015).
- Gibler, D (2008): “The Costs of Reneging: Reputation and Alliance Formation All Use Subject to JSTOR Terms and Conditions The Costs of Reneging Formation.” *Journal of Conflict Resolution* 52(3): 426–54.
- Gino, F; Ayal, S; Ariely, D (2013): „Self-Serving Altruism? The Lure of Unethical Actions that Benefit Others“. Duke University. Accessed 14. 02. 2015. Available:
<http://people.duke.edu/~dandan/Papers/Cheating/contagion.pdf>
- Glüpker, G (2013): “Effectiveness of Eu Conditionality in the Western Balkans: Minority Rights and the Fight against Corruption in Croatia and Macedonia.” *Journal of Contemporary European Research* 9(2): 220–361.
- Gray, J, Hicks, R (2014): “Reputations, Perceptions, and International Economic Agreements.” *International Interactions* 40(3): 325–49.
<http://www.tandfonline.com/doi/abs/10.1080/03050629.2014.899227> (July 20, 2015).
- Guarnieri, C, Piana, D (2009): “Independence and the Rule of Law: Lessons from Europe.” In *World Congress on the International Political Science Association*, Santiago del Chile, 1–24.
- Hartleb, F, Grabow, K (2013): „Exposing the Demagogues, Publications, Konrad-Adenauer-Stiftung“. Konrad-Adenauer-Stiftung. Accessed 16. 02. 2016. Available:
<http://www.kas.de/wf/en/33.35420/>
- Hendersen, D (2016): „Rent Seeking“. Econlibrary. Accessed 19. 04. 2016. Available:
<http://www.econlib.org/library/Enc/RentSeeking.html>
- Herrmann, R et al (1997): “Images in International Relations : An Experimental Test of Cognitive Schemata.” *International Studies* 41(3): 403–33.
- Hoff, K, Stiglitz, J (2016): “American Economic Association After the Big Bang ? Obstacles to the Emergence of the Rule of Law in Post-Communist Societies.” *The American Economic Review* 94(3): 753–63.
- Hsieh, H; Shannon, S (2005): “Three approaches to qualitative content analysis“. *Qualitative health research*, 15(9), pp.1277-1288.
- Inkeles, A, ed. (1991): „On measuring democracy: Its consequences and concomitants“. Transaction Publishers

- Inkeles, A (1990): "Introduction: On Measuring Democracy." *Studies in Comparative International Development* 25(1): 3–6.
- Jensen, C (2007): "Double Standards and Backsliding: The Double Edged Sword of Supermajoritarian Voting in the EU." In *Annual Meetings of the European Union Studies Association*
- Joslyn, S, et al. (2011): „Reducing probabilistic weather forecasts to the worst-case scenario: anchoring effects. - PubMed - NCBI“. Ncbi.nlm.nih.gov. Accessed 24. 11. 2015. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22039767>
- Jović, D (2006): "Croatia and the European Union: A Long Delayed Journey." *Journal of Southern Europe & the Balkans* 8(1): 85–103.
<http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=aph&AN=20573515&site=ehost-live&scope=site>.
- Keohane, R (1999): "International Relations and International Law: Interests , Reputation , Institutions." *American Society of International Law* 93: 375–79.
- Kimmel, E (2005): „Grundzüge des Marshallplans“. Bundeszentrale für Politische Bildung. Accessed 31. 05. 2014. Available: <http://www.bpb.de/geschichte/deutsche-geschichte/marshallplan/40034/einleitung?p=0#bio0>
- Kleinfeld, R (2005): "COMPETING DEFINITIONS OF THE RULE OF LAW Implications for Practitioners." *Carnegie Papers* (55).
<http://www.carnegieendowment.org/files/CP55.Belton.FINAL.pdf>.
- Kochenov, D (2006): "Why the Promotion of the Acquis Is Not the Same as the Promotion of Democracy and What Can Be Done in Order to Also Promote Democracy instead of Just Promoting the Acquis." *Hanse Law Review* 2(2): 171–95.
<http://www.hanselawreview.org/pdf4/Vol2No2Art02.pdf>.
- Kochenov, D (2009): "THE EU RULE OF LAW : CUTTING PATHS THROUGH CONFUSION." *Erasmus Law REVIEW* 2(1): 6–24.
http://www.erasmuslawreview.nl/files/ELR_issue6_002.pdf.
- Kochenov, D (2004): "Behind the Copenhagen Façade. The Meaning and Structure of the Copenhagen Political Criterion of Democracy and the Rule of Law." *European Integration online Papers (EIoP)* 8(10).
- Lindberg, S, et al. (2014): „V-Dem: A new way to measure democracy“. *Journal of Democracy* 25 (3), pp.159-169.
- Lipset, S (1959): "Some Social Requisites of Democracy". *American Political Science Review* 53(1), pp. 69-105.
- Magen, A (2009): "The Rule of Law and Its Promotion Abroad: Three Problems of Scope." *Stanford Journal of International Law* 45: 51–115.

- Mathes, R (1992): „Hermeneutisch-klassifikatorische Inhaltsanalyse von Leitfadengesprächen: über das Verhältnis von quantitativen und qualitativen Verfahren der Textanalyse und die Möglichkeit ihrer Kombination“. In: Hoffmeyer- Zlotnik, Jürgen H. P.(Ed.) (1992): „Analyse verbaler Daten : über den Umgang mit qualitativen Daten“. Opladen : Westdt. Verl.,
- Mayer, S (2008): “Demokratieindizes Im Vergleich Theorie Und Empirie Anhand Vier Ausgewählter Indizes.” Universität Wien.
- Mayring, P (2000): „Qualitative Content Analysis“. *Qualitative Research*. Accessed 08. 07. 2014. Available: <http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385>
- Menon, S; Sinha, A; Sreekantan, B.V. eds. (2014): „Interdisciplinary perspectives on consciousness and the self“. Springer.
- Mineshima, D (2002): “THE RULE OF LAW AND EU EXPANSION.” *Liverpool Law Review* (24): 73–87.
- Müftüler-Baç, M (2016): “The Pandora’s Box: Democratization and Rule of Law in Turkey.” *Asia Europe Journal* (14): 61–77. <http://link.springer.com/10.1007/s10308-015-0435-9>.
- Müller, T, Pickel, S (2007): “Wie Lässt Sich Demokratie Am Besten Messen? Zur Konzeptqualität von Demokratie-Indizes*.” *Politische Vierteljahresschrift* 3(2000): 511–39.
- Munck, G, Verkuilen, J (2002): „Conceptualizing and Measuring Democracy: Evaluating Alternative Indices“. *Comparative Political Studies* 35(1): 5-34. Available: <http://cps.sagepub.com/cgi/doi/10.1177/001041400203500101>
- Nazli Inal, N; Yegenoglu, D (2005): „German and French Leaders’ Views on Turkey’s EU Membership“. Washington Institute. Accessed 04. 11. 2015. Available: <http://www.washingtoninstitute.org/policy-analysis/view/german-and-french-leaders-views-on-turkeys-eu-membership>
- Negrine, R, et al. (2008): “Turkey and the European Union. An Analysis of How the Press in Four Countries Covered Turkey’s Bid for Accession in 2004.” *Turkey and the European Union: Facing New Challenges and Opportunities* 23(1): 47–68. Available: <http://ejc.sagepub.com/content/23/1/47.full.pdf+html>.
- Noutcheva, G., Aydin-Düzgit, S (2012): “Lost in Europeanisation: The Western Balkans and Turkey.” *West European Politics* 35(1): 59–78.
- O'Brien, B (2009): „Recipe for Bias: An Empirical Look at the Interplay between Institutional Incentives and Bounded Rationality in Prosecutorial Decision Making“, *A., Mo. L. Rev.* 74 pp. 999.

- (o. J.): „Validity and Reliability - How to Know if the Research is Correct?“. Explorable.com. Accessed 21. 03. 2015. Available: <https://explorable.com/validity-and-reliability>
- Pauly, C, Schult, C (2012): „Plans For Tougher Rule Enforcement: EU At Risk of Remaining a Toothless Tiger - SPIEGEL ONLINE“. SPIEGEL ONLINE. Accessed 23. 04. 2015. Available: <http://www.spiegel.de/international/europe/eu-attempts-to-get-tougher-on-rule-enforcement-at-risk-of-failing-a-860038.html>
- Pavelic, B (2013): „Slovenia Pressured Not to Delay Croatia's EU Accession“. Balkan Insight. Accessed 22. 11. 2015. Available: <http://www.balkaninsight.com/en/article/eu-expects-slovenia-not-to-block-croatia>
- Payne, D. (2010): “Turkey’s Path to Europe: Defense of Human Rights and Respect for International Obligations.” *Mediterranean Quarterly* 21(3): 8–15. Available: <http://mq.dukejournals.org/cgi/doi/10.1215/10474552-2010-012>.
- Pelham, N (2000): „Morocco's quest to be European“. BBC news online. Accessed 31. 05. 2015. Available: <http://news.bbc.co.uk/2/hi/africa/699760.stm>
- Piana, D (2010): “Beyond Judicial Independence: Rule of Law and Judicial Accountabilities in Assessing Democratic Quality.” *Comparative Sociology* 9(1): 40–64. Available: <http://booksandjournals.brillonline.com/content/journals/10.1163/156913210x12535202814351> (May 13, 2014).
- Pierson, P (2000): “Increasing Returns, Path Dependence, and the Study of Politics.” *The American Political Science Review* 94(2): 251–67. Available: http://www.uazuay.edu.ec/estudios/com_exterior/tamara/Pierson-Path_Dependency.pdf.
- Pinder, J (1991): “The European Community , the Rule of Law and Representative Government : The Significance of the Intergovernmental Conferences.” *Government and Opposition* 26(2): 199–214. Available: <http://onlinelibrary.wiley.com/doi/10.1111/j.1477-7053.1991.tb01133.x/pdf>.
- Ponsioen, S (2014): „Contagious business: when we copy unethical behavior“. Groningen University. Accessed 28. 07. 2015. Available: [http://www.rug.nl/research/portal/publications/pub\(328a0064-cfd1-4626-b82d-e54fc94e3764\).html!null](http://www.rug.nl/research/portal/publications/pub(328a0064-cfd1-4626-b82d-e54fc94e3764).html!null)
- Rankin, J (2016): „Turkey's long road to EU membership just got longer“. The Guardian. Accessed 28. 07. 2016. Available: <https://www.theguardian.com/world/2016/jul/20/turkeys-long-road-to-eu-membership-just-got-longer>
- Redmond, J (2007): “Turkey and the European Union: Troubled European or European Trouble ?” *International Affairs* 83(2): 305–17.

- Reiter, D (2012): „Democratic Peace Theory“. Oxford Bibliographies. Accessed 04. 05. 2015. Available: <http://www.oxfordbibliographies.com/view/document/obo-9780199756223/obo-9780199756223-0014.xml>
- Rios-Figueroa; J; Staton, J (2009): “Unpacking the Rule of Law: A Review of Judicial Independence Measures.” In *CELS 2009 4th Annual Conference on Empirical Legal Studies*, Available: <http://people.bu.edu/jgerring/Conference/MeasuringDemocracy/documents/RiosStaton2009.pdf> (February 2, 2015).
- Rogers, P; Senturia, T (2013): „Decision Insights: How group dynamics affect decisions“. BAIN Company. Accessed 24. 05. 2015. Available: <http://www.bain.com/publications/articles/how-group-dynamics-affect-decisions.aspx>
- Roos, S (2007): “The ‘Rule of Law’ as a Requirement for Accession to the European Union.” In *INTERNATIONAL LAW SYMPOSIUM IN BUCHAREST*, Bucharest: Konrad Adenauer Stiftung. Available: http://www.kas.de/wf/doc/kas_12154-1522-2-30.pdf?071026144542.
- Schimmelfennig, F (2008): “EU Political Accession Conditionality after the 2004 Enlargement: Consistency and Effectiveness.” *Journal of European Public Policy* 15(6): 918–37. Available: <http://www.tandfonline.com/doi/abs/10.1080/13501760802196861> (December 20, 2014).
- Schimmelfennig, F; Sedelmeier, U (2004): “Governance by Conditionality: EU Rule Transfer to the Candidate Countries of Central and Eastern Europe.” *Journal of European Public Policy* 11(4): 661–79. Available: <http://www.tandfonline.com/doi/abs/10.1080/1350176042000248089> (November 6, 2014).
- Schmidt, M (1994): „Measuring democracy and autocracy“. Handbook of Political Science Methods, London: Routledge (forthcoming)
- Schweickert; Hammermann (2005). „EU Enlargement and Institutional Development“. *Kiel Working Papers*, Kiel Institute for the World Economy, Available: <http://econpapers.repec.org/paper/kielkieliw/1261.htm>
- Shirrefs, M (2014): „Democracy distorted in Hungary“. Radio National. Accessed 13. 08. 2015. Available: <http://www.abc.net.au/radionational/programs/lawreport/5315660>
- Simmons, P (2011): “Review Commentary: The State of the Art in the EU Democracy Promotion Literature.” *Journal of Contemporary European Research* 7(1): 129–42.
- Skaaning, S (2010): “Measuring the Rule of Law.” *Political Research Quarterly* 63(2): 449–60.

- Spreeuw, M (2001): „Double Standards in the Application of the EU Values and Principles: Member States versus Potential Candidate and Candidate Countries“. Available: https://www.pravo.unizg.hr/_download/repository/Paper_M._Spreeuw.pdf
- Strübing, J (2004): „Grounded Theory: Zur sozialtheoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung“. VS Verlag für Sozialwissenschaften, Wiesbaden
- Szolucha, A (2010): “The EU and Enlargement Fatigue : Why Has the European Union Not Been Able to Counter Enlargement Fatigue ?” *Journal of Contemporary European Research* 6(1).
- Tamanaha, B (2004): “THE DARK SIDE OF THE RELATIONSHIP BETWEEN THE RULE OF LAW AND LIBERALISM.” *NYZ Journal of Law & Liberty* 139: 516–47. Available: http://www.law.nyu.edu/sites/default/files/ECM_PRO_060975.pdf.
- Thomas, K (2013): „Ungarn droht eine zügellose Demokratie“. *Süddeutsche Zeitung*. Accessed 23. 05. 2015. Available: <http://www.sueddeutsche.de/politik/verfassungsänderung-in-ungarn-ungarn-droht-eine-zuegellose-demokratie-1.1622980>
- Touchton, M (2010): “Book Review: Munck, G. (2009). *Measuring Democracy: A Bridge Between Scholarship and Politics*.” *Comparative Political Studies* 43(10): 1313–17. Available: <http://cps.sagepub.com/cgi/doi/10.1177/0010414010371183> (February 1, 2015).
- Tretbar, C (2013): „Nun ist Karlsruhe am Zug“. *Tagesspiegel*. Accessed 21. 11. 2015. Available: <http://www.tagesspiegel.de/politik/npd-verbot-nun-ist-karlsruhe-ist-am-zug/9159536.html>
- Vanhanen, T (2000): “A New Dataset for Measuring Democracy, 1810–1998*.” *Journal of Peace Research* 37(2): 251–65.
- Veebel, V (2011): “Relevance of Copenhagen Criteria in Actual Accession: Principles, Methods and Shortcomings of EU Pre-Accession Evaluation.” *Studies of Transition States and Societies* 3(3): 3–23.
- Yin, R (2009): „Essential guide to qualitative methods in organizational research“ In: „Case Study Research: Design and Methods“. eds. Leonard Bickman and Debra J Rog. Sage Publications. Available: <http://books.google.com/books?id=FzawIAdiHkC&pgis=1>
- Youngs, R (2014): „The EU and International Conflicts - Central African Republic“. *Carnegie Europe*. Accessed 31. 12. 2015. Available: <http://carnegieeurope.eu/strategieurope/?fa=55341>
- Youngs, R (2002): „The European Union and the Promotion of Democracy“. Oxford University Press.

Zalewski, P (2004): „Sticks , Carrots and Great Expectations : Human Rights Conditionality and Turkey ' S Path towards Membership of the European Union“. Available: [http://edoc.bibliothek.uni-halle.de/servlets/MCRFileNodeServlet/HALCoRe_derivate_00002977/Sticks, carrots and great expectations.pdf](http://edoc.bibliothek.uni-halle.de/servlets/MCRFileNodeServlet/HALCoRe_derivate_00002977/Sticks,_carrots_and_great_expectations.pdf).