Trade Policy Flexibility and Enforcement in the WTO – Reform Agenda Towards an Efficient “Breach” Contract

D I S S E R T A T I O N
of the University of St. Gallen,
Graduate School of Business Administration,
Economics, Law and Social Sciences (HSG)
to obtain the title of
Doctor of Economics

submitted by

Simon A.B. Schropp

from

Germany

Approved on the application of

Prof. Dr. Heinz Hauser

and

Prof. Dr. Petros C. Mavroidis

Dissertation no. 3435
Imprimerie Minute, Genève, 2008
The University of St. Gallen, Graduate School of Business Administration, Economics, Law and Social Sciences (HSG) hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, November 22, 2007

The President:

Prof. Ernst Mohr, PhD
To my parents,
for their love and unquestioning support
and

to my niece Johanna Tilly Moll (* June 26 2007),
in the hope that one day she will read this book and
find its content outdated
Acknowledgments

Writing a doctoral thesis by nature is a solitary and lonesome task. However, I doubt that this book would have seen the light of day without the help of the following individuals and institutions:

Professor Hauser was a “Doktorvater” in the truest sense. He was patient yet demanding, stern yet just, supportive yet always straightforward. I could not have hoped for a better supervision of my thesis. My co-advisor Petros Mavroidis has been a never-ending source of inspiration. His curiosity and his enthusiasm for the World Trade Organization have infected me. He always had the time to discuss and argue, worked “like a dog” on my earlier drafts, and invariably got me out of gloomy moods and phases of self-doubt.

Many thanks go to my “family connection”: My father, Peter Schropp, proved to be a remarkable text editor, and exhibited sheer endless energy when my batteries were running low during the home stretch near the completion of the thesis. Andreas Moll, my brother-in-law, thankfully took a break from his fatherly duties to help me with layout and formatting. Nicolas Tyabji, my flat mate, corrected my English and introduced me to many a linguistic finesse.

Cédric Dupont, Manfred Elsig, Henrik Horn, Alexander Keck, Patrick Low and Kornel Mahlstein provided academic input and moral support. Financial contributions from NCCR-Democracy, Schweizerische Akademie der Geistes- und Sozialwissenschaften, the Swiss Institute for International Economics and Applied Economic Research (SIAW HSG), and the Graduate Institute of International Studies, Geneva (HEI) are thankfully acknowledged.

Geneva, January 2008

Simon Schropp
## Summary of contents

### Abstract

1 Introduction and summary: A roadmap for theorizing about trade policy flexibility in the WTO

1.1 Trade policy flexibility in the WTO: Vice or virtue?

1.2 Objective of the study: Connecting issues of breach, remedies, and commitment level in incomplete contracts

1.3 Overview of the study and summary of findings

1.4 A brief survey of the literature on trade policy flexibility and enforcement in the WTO

1.5 Methodology and research approach

1.6 Contribution to scholarship

### Part 1 An introduction to incomplete contracting

2 Complete contracts, and the contracting ideal

2.1 Contracts: Enforceable commitment over time

2.2 Basics of contracting: Creating rules

2.3 Types of contracts, and alternatives

2.4 The contracting ideal: Pareto-efficient complete contingent contract

### Part 2 Theorizing about the WTO as an incomplete contract

4 Adding context: The WTO as an incomplete contract

4.1 Players, preferences, and contractual intent

4.2 Primary rules of contracting: Basic entitlements in the WTO

4.3 Establishing the WTO as an incomplete contract

4.4 Conclusion: The WTO – an incomplete contract based on market access externalities and minimum standards
5 Analyzing the system of non-performance in the WTO 204
  5.1 Trade policy flexibility and protection of the market access entitlement 205
  5.2 De iure and de facto protection of non-trade entitlements 218
  5.3 Rules of enforcement 220
  5.4 Does the current system of trade policy flexibility and entitlement protection make sense? 223

6 The compliance-vs.-rebalancing debate: A missed opportunity 242
  6.1 The origin of the compliance/rebalancing debate: The objective of WTO enforcement 243
  6.2 Why the compliance/rebalancing debate is misguided: Research questions reframed 255
  6.3 A critique of the compliance and rebalancing view 258
  6.4 Lessons learned from the compliance/rebalancing controversy 274

Part 3 Flexibility and enforcement in the WTO: Towards an agenda for reform 277
7 Theorizing about the “WTO*” as an efficient “breach” contract 278
  7.1 The “trade game” 281
  7.2 Organizing protection of the market access entitlement 285
  7.3 Organizing the protection of multilateral entitlements 308
  7.4 A two-tier system of enforcement 314
  7.5 The WTO* as an efficient “breach” contract: A better trade agreement? 318
  7.6 Post scriptum: Compliance vs. rebalancing – redux 325
8 Towards an efficient “breach” contract: An agenda for reform 327
  8.1 The shortlist of reform 328
  8.2 Long-term reform proposals 336
  8.3 Final remarks and future research 340

Bibliography 344
Curriculum vitae 369
Table of contents

Table of charts ix
List of abbreviations x
Abstract xiii

1 Introduction and summary: A roadmap for theorizing about trade policy flexibility in the WTO 1

1.1 Trade policy flexibility in the WTO: Vice or virtue? 1

1.2 Objective of the study: Connecting issues of breach, remedies, and commitment level in incomplete contracts 5

1.2.1 The system of non-performance: Breach and remedies in incomplete contracts 6
1.2.2 Knowing what is “right” in order to assess what is “wrong” 10
1.2.3 Research questions 10

1.3 Overview of the study and summary of findings 12

1.4 A brief survey of the literature on trade policy flexibility and enforcement in the WTO 17

1.5 Methodology and research approach 23

1.6 Contribution to scholarship 25

Part 1 An introduction to incomplete contracting 28

2 Complete contracts, and the contracting ideal 29

2.1 Contracts: Enforceable commitment over time 29

2.1.1 Timing 30
2.1.2 Commitment: Cooperative intent and assurance 30
2.1.3 Effective enforcement and the link between commitment and enforcement 32
2.1.4 Concluding remarks on the definition of contracts 40

2.2 Basics of contracting: Creating rules 42

2.2.1 Primary rules of contracting – exchange of entitlements 43
2.2.2 Secondary rules of contracting – entitlement protection 45
2.2.3 Tertiary rules of contracting – enforcement of entitlements 48
2.2.4 Mixed regimes of entitlement and entitlement protection 49

2.2.4.1 Multi-entitlement contracts and divided entitlements 49
2.2.4.2 Rules of divided entitlement protection 50
2.3 Types of contracts, and alternatives
2.3.1 Collaboration vs. coordination
2.3.2 Complexity of contracts, and alternatives to contracting
2.4 The contracting ideal: Pareto-efficient complete contingent contract

3 Incomplete contracting, and the essence of flexibility
3.1 A categorization of contractual incompleteness
3.1.1 What makes contracts incomplete? Transaction costs and bounded rationality
3.1.1.1 Transaction costs
3.1.1.2 Bounded rationality
3.1.2 What types of “uncertainty” can be distinguished? The concept of incertitude
3.1.3 Contractual incompleteness: A taxonomy
3.1.4 Do we have the technical toolkit for dealing with contractual incompleteness?
3.1.5 The effects of incompleteness on contracting behavior
3.1.5.1 Victims’ willingness to cooperate
3.1.5.2 Injurers’ willingness to cooperate
3.1.6 Summarizing the findings on contractual incompleteness
3.2 How to deal with contractual incompleteness: Strategies of gap-filling
3.2.1 Circumnavigating incompleteness: Comprehensive contracting
3.2.2 Seizing regret: Drafting flexibility mechanisms
3.2.2.1 Contingency measures
3.2.2.2 Default rules
3.2.3 Minimizing regret: The principle of precaution
3.2.3.1 Diligence and preamble language
3.2.3.2 Relational contracting
3.2.4 Delegating responsibility: Using courts as gap-fillers
3.2.5 Summary: The art of dealing with contractual incompleteness
3.2.5.1 Finding out the accurate type of incompleteness
3.2.5.2 The importance of default rules of flexibility
3.3 Crafting rules of flexibility: Inalienability, specific performance, or liability?
3.3.1 Inalienability or efficient non-performance?
3.3.2 Liability or property rule?
3.3.2.1 Which remedy best complements a liability rule?
3.3.2.2 Property rule or liability rule? A question of transaction costs
3.3.3 Additional modalities of default rule design
3.4 A framework for theorizing about incomplete contracts: Revisiting the concept of the efficient “breach” contract

3.4.1 A framework for theorizing about incomplete contracts

3.4.1.1 Phase 1: Analyzing the nature and context of the contract

3.4.1.2 Phase 2: Mapping post-contractual constraints; assessing trade-offs

3.4.1.3 Phase 3: Hypothetical bargain analysis

3.4.2 The efficient “breach” contract as the achievable first-best contract

3.5 A first step towards a general theory of disputes?

Part 2 Theorizing about the WTO as an incomplete contract

4 Adding context: The WTO as an incomplete contract

4.1 Players, preferences, and contractual intent

4.1.1 Players and preferences: Political economy theories of endogenous trade policymaking

4.1.2 Contractual intent: What is the rationale for trade cooperation?

4.1.2.1 An overview of economic approaches to trade agreements

4.1.2.2 The commitment approach to trade agreements

4.1.2.3 The market access externalities approach to trade agreements

4.1.2.4 A brief evaluation of rationales for trade contracts

4.1.2.5 Additional rationales for trade contracts

4.1.3 A tentative conclusion: Trade agreements based on market access externalities and minimum standards

4.1.3.1 The primacy of market access externalities

4.1.3.2 Trade contracts based on market access externalities

4.1.3.3 Minimum standards as a second(ary) rationale for trade agreements

4.2 Primary rules of contracting: Basic entitlements in the WTO

4.2.1 The bilateral market access entitlement

4.2.2 Minimum standard entitlements

4.2.3 Basic auxiliary rules of entitlement

4.2.4 The prominent role of the market access entitlement

4.3 Establishing the WTO as an incomplete contract

4.3.1 Contingencies and incertitude affecting the market access entitlement

4.3.1.1 The nature of market access contingencies: Political support shocks with spillover potential

4.3.1.2 How contingencies affect the entitlement to trade: The nature of incertitude and the resulting type of incompleteness
4.3.2 Contingencies, incertitude and incompleteness affecting minimum standard entitlements and other multilateral entitlements

4.4 Conclusion: The WTO – an incomplete contract based on market access externalities and minimum standards

5 Analyzing the system of non-performance in the WTO

5.1 Trade policy flexibility and protection of the market access entitlement

5.1.1 De iure protection of the market access entitlement
- 5.1.1.1 Negative integration provisions
- 5.1.1.2 Non-violation complaints (NVC)
- 5.1.1.3 Contingency measures
- 5.1.1.4 The default rule

5.1.2 De facto protection of the market access entitlement
- 5.1.2.1 Voluntary export restraints (and ordinary marketing agreements)
- 5.1.2.2 Subsidies
- 5.1.2.3 Non-discriminatory domestic policies
- 5.1.2.4 Antidumping and countervailing duties
- 5.1.2.5 Violation of the Agreement

5.2 De iure and de facto protection of non-trade entitlements

5.2.1 De iure protection of multilateral entitlements

5.2.2 De facto protection of multilateral entitlements

5.3 Rules of enforcement

5.4 Does the current system of trade policy flexibility and entitlement protection make sense?

5.4.1 The flawed protection of the market access entitlement
- 5.4.1.1 Examining flaws in the de iure system of entitlement protection
- 5.4.1.2 Examining flaws in the de facto system of entitlement protection
- 5.4.1.3 Examining flaws in the system of enforcement

5.4.2 The flawed protection of multilateral non-trade entitlements

5.4.3 Conclusion and consequences
- 5.4.3.1 Summary of the main flaws of the current WTO system of non-performance
- 5.4.3.2 Over-“breach” and under-commitment
7.3 **Organizing the protection of multilateral entitlements**  
7.3.1 Focusing on default rules  
7.3.2 The optimal design of default rules protecting multilateral entitlements  
7.3.2.1 If *ex post* discretion is impermissible: An inalienability rule of default  
7.3.2.2 If *ex post* discretion causes minor damage: A liquidated damages rule of default  
7.3.2.3 If *ex post* discretion is permissible: A property rule of default  
7.3.3 Conclusion: Mixed default rules of protection for multilateral entitlements  
7.4 A two-tier system of enforcement  
7.5 **The WTO* as an efficient “breach” contract: A better trade agreement?**  
7.5.1 How do the WTO and the WTO* differ?  
7.5.2 Efficiency edge of the WTO* over the WTO  
7.5.3 The WTO*: A “better” contract?  
7.6 Post scriptum: Compliance vs. rebalancing – redux  
8 Towards an efficient “breach” contract: An agenda for reform  
8.1 The shortlist of reform  
8.1.1 Establish a revised Art. XIX GATT  
8.1.2 Add Art. Xbis to the WTO Agreement  
8.1.3 Revise Art. 22 DSU  
8.2 Long-term reform proposals  
8.2.1 Reforming the protection of the market access entitlement  
8.2.2 Reforming the protection of multilateral entitlements  
8.2.3 Reforming the WTO enforcement regime  
8.3 Final remarks and future research  

**Bibliography**  

**Curriculum vitae**
# Table of charts

<table>
<thead>
<tr>
<th>Chart Number</th>
<th>Chart Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Non-performance in incomplete contracts</td>
<td>6</td>
</tr>
<tr>
<td>1.2</td>
<td>Commitment, breach, and trade policy flexibility in incomplete contracts</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>Locating the existent WTO literature on trade policy flexibility</td>
<td>18</td>
</tr>
<tr>
<td>2.1</td>
<td>Enforcement constraint in contracts</td>
<td>34</td>
</tr>
<tr>
<td>2.2</td>
<td>The importance of enforcement capacity and enforceability in contracts</td>
<td>37</td>
</tr>
<tr>
<td>2.3</td>
<td>Divided entitlement protection – points along a continuum</td>
<td>51</td>
</tr>
<tr>
<td>3.1</td>
<td>Incertitude in risk management – risk, uncertainty, ignorance, and ambiguity</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Incertitude in contract theory – a novel taxonomy</td>
<td>65</td>
</tr>
<tr>
<td>3.3</td>
<td>A taxonomy of contractual incompleteness</td>
<td>68</td>
</tr>
<tr>
<td>3.4</td>
<td>The inadequacy of the economic toolkit in dealing with incompleteness</td>
<td>77</td>
</tr>
<tr>
<td>3.5</td>
<td>The impact of contractual incompleteness on victims’ commitment levels</td>
<td>82</td>
</tr>
<tr>
<td>3.6</td>
<td>The impact of contractual rigidity on injurers’ commitment levels</td>
<td>84</td>
</tr>
<tr>
<td>3.7</td>
<td>Overview of gap-filling strategies in incomplete contract situations</td>
<td>105</td>
</tr>
<tr>
<td>3.8</td>
<td>Gap-filling strategies broken down by types of incompleteness</td>
<td>108</td>
</tr>
<tr>
<td>3.9</td>
<td>Designing contractual default rules</td>
<td>112</td>
</tr>
<tr>
<td>3.10</td>
<td>A stylized version of a contracting game under incompleteness</td>
<td>136</td>
</tr>
<tr>
<td>4.1</td>
<td>Overview of economic rationales for trade agreements</td>
<td>159</td>
</tr>
<tr>
<td>4.2</td>
<td>Economic and non-economic rationales for trade agreements</td>
<td>173</td>
</tr>
<tr>
<td>4.3</td>
<td>Overview of primary entitlements in the WTO</td>
<td>188</td>
</tr>
<tr>
<td>4.4</td>
<td>The nature of interaction in the WTO</td>
<td>192</td>
</tr>
<tr>
<td>4.5</td>
<td>The nature of incompleteness affecting the market access entitlement</td>
<td>200</td>
</tr>
<tr>
<td>6.1</td>
<td>Rephrasing the compliance/rebalancing debate in WTO scholarship</td>
<td>257</td>
</tr>
<tr>
<td>6.2</td>
<td>Extending the compliance/rebalancing debate</td>
<td>259</td>
</tr>
<tr>
<td>7.1</td>
<td>Trade-offs and constraints in the WTO “trade game”</td>
<td>281</td>
</tr>
<tr>
<td>7.2</td>
<td>The breakdown condition for the simple tariff-setting game</td>
<td>288</td>
</tr>
<tr>
<td>7.3</td>
<td>Stability and breakdown in the escape-clause game after Rosendorff (2005)</td>
<td>291</td>
</tr>
<tr>
<td>7.4</td>
<td>Stability in the escape-clause game after Herzing (2005)</td>
<td>293</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>AJIL</td>
<td>American Journal of International Law</td>
<td></td>
</tr>
<tr>
<td>AoA</td>
<td>Agreement on Agriculture</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>Appellate Body</td>
<td></td>
</tr>
<tr>
<td>ABM</td>
<td>Agreement on Anti-Ballistic Missiles</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>Antidumping (action)</td>
<td></td>
</tr>
<tr>
<td>ADA</td>
<td>Antidumping Agreement</td>
<td></td>
</tr>
<tr>
<td>AER</td>
<td>American Economic Review</td>
<td></td>
</tr>
<tr>
<td>Art./Arts.</td>
<td>Article(s)</td>
<td></td>
</tr>
<tr>
<td>BoP</td>
<td>Balance-of-payment (crisis)</td>
<td></td>
</tr>
<tr>
<td>CCC</td>
<td>Pareto-efficient complete contingent contract</td>
<td></td>
</tr>
<tr>
<td>CvD</td>
<td>Countervailing duty (action)</td>
<td></td>
</tr>
<tr>
<td>DDA</td>
<td>Doha Development Agenda</td>
<td></td>
</tr>
<tr>
<td>DG</td>
<td>Directorate General</td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>Default rule(s)</td>
<td></td>
</tr>
<tr>
<td>DS</td>
<td>Dispute settlement</td>
<td></td>
</tr>
<tr>
<td>DSB</td>
<td>Dispute Settlement Body</td>
<td></td>
</tr>
<tr>
<td>DSM</td>
<td>Dispute Settlement Mechanism</td>
<td></td>
</tr>
<tr>
<td>DSU</td>
<td>Dispute Settlement Understanding</td>
<td></td>
</tr>
<tr>
<td>EBC</td>
<td>Efficient “breach” contract</td>
<td></td>
</tr>
<tr>
<td>EC</td>
<td>European Communities</td>
<td></td>
</tr>
<tr>
<td>EEC</td>
<td>European Economic Communities</td>
<td></td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
<td></td>
</tr>
<tr>
<td>GATS</td>
<td>General Agreement on Trade in Services</td>
<td></td>
</tr>
<tr>
<td>GATT</td>
<td>General Agreement on Tariffs and Trade</td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td>Agreement on Government Procurement</td>
<td></td>
</tr>
<tr>
<td>GSP</td>
<td>Generalized system of preferences</td>
<td></td>
</tr>
<tr>
<td>ILC</td>
<td>International Law Commission</td>
<td></td>
</tr>
<tr>
<td>ILO</td>
<td>International Labour Organisation</td>
<td></td>
</tr>
<tr>
<td>ILP</td>
<td>Agreement on Import Licensing Procedures</td>
<td></td>
</tr>
<tr>
<td>IMF</td>
<td>International Monetary Fund</td>
<td></td>
</tr>
<tr>
<td>IO</td>
<td>Industrial organization</td>
<td></td>
</tr>
<tr>
<td>IP</td>
<td>Intellectual property</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>International relations</td>
<td></td>
</tr>
<tr>
<td>Acronym</td>
<td>Abbreviation/Explanation</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>Information technology</td>
<td></td>
</tr>
<tr>
<td>ITO</td>
<td>International Trade Organisation</td>
<td></td>
</tr>
<tr>
<td>L&amp;E</td>
<td>Law &amp; Economics</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>Liability rule(s)</td>
<td></td>
</tr>
<tr>
<td>LDCs</td>
<td>Least developed countries</td>
<td></td>
</tr>
<tr>
<td>MFN</td>
<td>Most-favored nation (clause)</td>
<td></td>
</tr>
<tr>
<td>n.a.</td>
<td>not available</td>
<td></td>
</tr>
<tr>
<td>NGOs</td>
<td>Non-governmental organzations</td>
<td></td>
</tr>
<tr>
<td>NVC</td>
<td>Non-violation complaints</td>
<td></td>
</tr>
<tr>
<td>OMAs</td>
<td>Ordinary marketing agreements</td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td>Prisoners’ dilemma</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>Property rule(s)</td>
<td></td>
</tr>
<tr>
<td>PROF</td>
<td>Politically realistic objective function(s)</td>
<td></td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research and development</td>
<td></td>
</tr>
<tr>
<td>RPT</td>
<td>Reasonable period of time</td>
<td></td>
</tr>
<tr>
<td>ROO</td>
<td>Agreement on Rules of Origin</td>
<td></td>
</tr>
<tr>
<td>SALT</td>
<td>Strategic Arms Limitations Treaty</td>
<td></td>
</tr>
<tr>
<td>SCM</td>
<td>Agreement on Safeguards and Countervailing Measures</td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>Safeguard (action)</td>
<td></td>
</tr>
<tr>
<td>SGA</td>
<td>Agreement on Safeguards</td>
<td></td>
</tr>
<tr>
<td>SIGs</td>
<td>Special interest groups (lobbies)</td>
<td></td>
</tr>
<tr>
<td>SPS</td>
<td>Agreement on Sanitary and Phytosanitary Measures</td>
<td></td>
</tr>
<tr>
<td>TBT</td>
<td>Agreement on Technical Barriers to Trade</td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>Transaction costs</td>
<td></td>
</tr>
<tr>
<td>TOT</td>
<td>Terms of trade</td>
<td></td>
</tr>
<tr>
<td>TPRM</td>
<td>Trade Policy Review Mechanism</td>
<td></td>
</tr>
<tr>
<td>TRIMs</td>
<td>Agreement on Trade-Related Investment Measures</td>
<td></td>
</tr>
<tr>
<td>TRIPS</td>
<td>Agreement on Trade-Related Intellectual Property Rights</td>
<td></td>
</tr>
<tr>
<td>U.S./US</td>
<td>United States of America</td>
<td></td>
</tr>
<tr>
<td>U.C.C.</td>
<td>United States Uniform Commercial Code</td>
<td></td>
</tr>
<tr>
<td>UR</td>
<td>Uruguay Round</td>
<td></td>
</tr>
<tr>
<td>USTR</td>
<td>United States Trade Representative</td>
<td></td>
</tr>
<tr>
<td>VCLT</td>
<td>Vienna Convention on the Law of Treaties</td>
<td></td>
</tr>
<tr>
<td>VER/VERs</td>
<td>Voluntary export restraint(s)</td>
<td></td>
</tr>
<tr>
<td>WTO</td>
<td>World Trade Organization</td>
<td></td>
</tr>
</tbody>
</table>
Abstract

The World Trade Organization (WTO) is an incomplete contract among sovereign States. Incomplete contracts contain gaps. Ex post, contractual gaps leave gains from trade unrealized, they may create “regret” in signatories when unanticipated contingencies occur. Regret contingencies are dealt with by means of contractual flexibility mechanisms. Trade policy flexibility rules are intra-contractual non-performance arrangements. They are backed up by enforcement instruments which provide for punishment of extra-contractual behavior.

WTO scholarship is somewhat opaque as to how design and choice of trade flexibility mechanisms and enforcement instruments influence countries’ decisions to liberalize (ex ante) and to default (ex post).

This study provides an extensive introduction to incomplete contracts in general and engages in a comprehensive contract-theoretical analysis of the WTO. Establishing the WTO as a necessarily incomplete contract concluded between self-interested and reasonably rational trade policymakers, it analyzes the flaws of the contemporary system of entitlement protection, trade policy flexibility, and enforcement in the international trading order. The contemporary WTO regime of non-performance is found to be remarkably inadequate in addressing the needs of signatories. In order to substantiate our reform proposal, we conduct a hypothetical bargain analysis of what the WTO should look like if it were to be properly re-designed. The study contains a detailed reform agenda of the flexibility design of the contract and suggests explicit treaty language aimed at turning the WTO into an efficient “breach” contract, the achievable first-best contract that WTO signatories can possibly muster.

This thesis is a contribution to the growing theoretical literature on the positive political economy of the World Trade Organization. It elucidates the linkage of contractual incompleteness, intra-contractual trade policy flexibility mechanisms, contract enforcement, and WTO Members’ willingness to commit to trade liberalization. It contributes to the body of WTO scholarship by providing an assessment of the weaknesses of the current regime of non-performance in the WTO and the systemic and dynamic consequences thereof. This reform agenda is concrete, politically realistic, and systemically viable.
1 Introduction and summary: A roadmap for theorizing about trade policy flexibility in the WTO

The World Trade Organization (WTO) is a multilateral trade agreement and as such the international equivalent of a contract. It lies in the nature of a trade agreement that governments make commitments on policies that will affect not only the present time but also the future. However, economic circumstances in the future may evolve in ways that make maintenance of the present policy concessions untenable. The WTO contract provides countries with means to depart from previously agreed obligations under well-defined and circumscribed conditions. Trade policy flexibility tools, or trade contingency measures, address this need. Yet we find fault with the quality of the current regime of trade policy flexibility in the WTO, and lay out a roadmap for theorizing about trade flexibility and enforcement in the multilateral trading system.

1.1 Trade policy flexibility in the WTO: Vice or virtue?

It is widely accepted in contemporary scholarship that the WTO\(^1\) is an incomplete contract.\(^2\) Incomplete contracts contain gaps: Important contingencies (future conditions, or “states of nature”) are not considered in the original contract, and thus are not conclusively and unambiguously specified. Contractual gaps can leave ex post gains from trade unrealized. This creates room for “regret” whenever unanticipated and unforeseen developments occur, such as a protectionist shock within a country. Such a shock may seriously threaten some domestic import-competing or export sectors, and

---

\(^1\) Throughout the course of this study, the terms “WTO” or “the Agreement” will be used interchangeably as shorthand for the bundle of multilateral contracts that are known as the Uruguay Round Agreements. These Agreements include the Marrakech Agreement (“WTO Agreement”, or “WTO Charter”), and all the treaties mentioned in Annexes 1-4.

\(^2\) This statement actually contains two messages. First, the WTO is an international contract. The WTO Appellate Body (AB) in Japan – Alcoholic Beverages expressly stated that “the WTO Agreement is a treaty – the international equivalent of a contract” (WT/DS 8,10,11/AB/R: 16). Second, the contract is incomplete in important areas. This is a widely accepted view in the literature (e.g. Downs and Rocke 1995; Dunoff and Trachtman 1999; Ethier 2001a; Hauser and Roitinger 2004, 2003; Herzing 2005; Horn et al. 2006; Rosendorff 2005; Rosendorff and Milner 2001; Lawrence 2003). Neither assertions will remain unchecked. The nature of the contract, as well as the sources of its incompleteness, will be subject to an extensive and substantiated analysis infra.
therewith welfare and/or employment of certain groups of society, or economic growth and social cohesion at large. For the affected WTO Member, performance as previously (ex ante\(^3\)) agreed upon may then no longer be desirable or mutually efficient.\(^4\)

The framers of the WTO were aware of the presence of contractual gaps and the inevitable uncertainty in the economic environment. In order to seize gains from regret and to deflate the build-up of domestic pressure against trade liberalization, the WTO contract includes certain trade policy flexibility instruments that permit one party (the “injurer”) to (partially) default, i.e. to step back, or withdraw, from contractual performance as previously agreed. The injurer can do so if certain preconditions are met, most notably those of compensating the parties affected by such back-tracking behavior (the “victims”).\(^5\)

The WTO provides for several formal, de iure, trade policy flexibility mechanisms.\(^6\) Examples in the General Agreement on Tariffs and Trade (GATT) are Art. XII (Restrictions to Safeguard the Balance of Payments, applicable only to developed countries), Art. XVIII (infant industry protection and balance of payments crises; applicable to developing countries only), Art. XIX (Emergency Actions on Imports of Particular Products, also known as the “safeguards clause”), Art. XX (General Exceptions), Art. XXI (Security Exceptions), and Art. XXVIII (Modification of Schedules, also known as tariff renegotiation).\(^7\) Common to these de iure flexibility mechanisms is a rather high level of conditionality (enactment preconditions and scope

---

\(^3\) The terms “ex post” and “ex ante” are used in this study with regard to the conclusion of a contract. Ex ante refers to periods in advance of signing an agreement, whereas ex post events denote those after the conclusion of the deal, i.e. during contract performance.

\(^4\) To grasp the concept of ex post regret, consider the simple example of a fixed-price (non-contingent) contract that obliges one party to produce and the other party to buy a product. An earthquake destroys the production facilities and makes delivery as prescribed extremely costly: The producer will prefer not to perform; by means of a side payment to the buyer (exceeding the latter’s personal value of the good) both parties can be made better off by not conducting the transaction (see Shavell 1980 at footnote 4).

\(^5\) No negative or positive connotations are implied in calling the respective parties “injurer” and “victim”. As a convention, the contracting party (WTO Member) experiencing regret shall be called the “injurer”, and the party/parties affected by any actual or planned measures, shall be called “victim”.

\(^6\) Trade policy flexibility tools are sometimes also called “trade contingency measures”, “safety valves”, or “escape clauses”. Later on in the study we will explain why neither of these terms is sufficient in covering the entire realm of trade policy flexibility mechanisms.

\(^7\) Similar examples of trade policy flexibility instruments can be found in other WTO Agreements, such as the General Agreement on Trade in Services (GATS), the Agreement on Technical Barriers to Trade (TBT), or the Agreement on Agriculture (AoA).
of application), as well as relatively modest indemnity payments to the affected victim countries.

In addition to these de iure escape clauses there are various informal, de facto, flexibility tools available to WTO Members. Albeit in contravention of the letter of the law or at least the spirit of the Agreement, trade policy tools such as voluntary export restrictions (VERs), ordinary marketing agreements (OMAs), antidumping (AD) and countervailing duty (CvD) measures, subsidies, or a simple violation of the Agreement are often used by WTO Members as ways to escape initially made trade liberalization commitments. Given that these de facto trade policy flexibility mechanisms happen more or less in the shadow of the law, their use is characterized by lower enactment costs, far-reaching scope of application (especially in the case of violation of the Agreement), and indemnity payments (damages) that are strictly lower than commensurate.

The current system of trade policy flexibility in the WTO raises a string of serious systemic issues. As an example: Why do certain Members prefer the use of AD and CvD measures over the use of the escape clause of Art. XIX GATT, what are the consequences of such behavior, and what can be done to reverse this trend (e.g. Barfield 2001; Barton et al. 2006; Blonigen and Bown 2003; Bown 2001; Finger et al. 2001; Palmeter 1991b; Finger et al. 1982; Messerlin 2000)?

Next, what is the logic of sanctioning legal escape options and contractual defection in the same manner? Note that the WTO applies the same remedy – substantially equivalent damages – to legitimate non-performance (Arts. XIX, XXVIII) as well as to flat-out violation of the Agreement (Art. 22.4 DSU). Further, what is the WTO’s rationale for

---

8 The level of conditionality of a flexibility instrument is composed of two elements, the first being enactment thresholds, or preconditions. Enactment thresholds are contingency-related preconditions that the injurer has to surpass before making use of a flexibility mechanism. Enactment costs are sunk, and compensation payments do not form part of conditionality-related costs. The second element of conditionality is the scope of application, the contractual strings that are attached to the use of a trade policy flexibility mechanism. Safeguards under Art. XIX GATT, for example, can be invoked exclusively in times of economic distress, and only used once. As a rule, the duration of safeguard measures is a period of four years. In addition, safeguards cannot in principle be directed at the country or set of countries that is the source of the injury; they have to be used on a non-discriminatory basis (safeguards must be offered on a “most-favored nations” basis, or “MFN’ed” in WTO parlance). A flexibility instrument’s ease of use is thus clearly a function of the level of conditionality and its scope of application.

9 As will be shown later in more detail, many informal escape mechanisms, such as AD and CvD measures, do not provide for any compensation of victims at all. Even utilizing violation-cum-retaliation as an escape mechanism (i.e. breaching the Agreement, losing a trade litigation, and withering retaliatory measures enacted by the victim) does not add up to commensurate damages due to the way dispute panels have interpreted Art. 22.4 of the Understanding on Rules and Procedures Governing the Settlement of Disputes (DSU).

10 Many of these issues have been addressed by WTO scholarship; some have already been subject to litigation in
having a whole arsenal of substitutive escape clauses that have overlapping scopes of application? In a given situation, a Member has the choice of resorting to GATT Art. XIX, Art. XXVIII, VERs/OMAs, an AD measure (under Art. VI GATT and the Antidumping Agreement, ADA), or violation of the Agreement. Various flexibility mechanisms only differ in their level of conditionality and the compensation payable to the victims. It is thus evident that an injuring country will always go for the very escape instrument which promises “most mileage”, i.e. the fewest enactment costs, the lowest compensation, and the largest scope of application. As a consequence, instead of engaging in legal contractual escape, Members opt for informal protectionist escape instruments, such as antidumping and countervailing duty actions, and risk losing the ensuing disputes (see for instance Bown 2001, 2002a, 2002b, 2004; Finger 1998; Finger, et al. 2001; Lawrence 2003; Roitinger 2004; Schropp 2005; Sykes 1991).

Another concern is the lack of scope in existing de iure escape clauses. Many scholars argue that de facto breaches of WTO obligations often occur because of the rigidity connected to the enactment of formal escape mechanisms, such as Art. XIX GATT. Mavroidis (2006) states that the more rigid and “expensive” (in terms of remedies) contractual safeguards are, the less they are used. According to Mavroidis, WTO Members are more likely to violate the WTO treaty if rigid safeguards deny them the necessary “breathing space”.

The current WTO safeguards regime allegedly does not address Members’ needs for policy flexibility (see also Horn and Mavroidis 2003; Roitinger 2004; Sykes 2003). As became clear in the course of the EC – Hormones case, the European Communities, for political or health reasons, wished to step back from a previously made commitment. This endeavor, however, is not considered in any formal WTO escape clause. Hence, lacking any official means of withdrawing from existing concessions, the EC saw no alternative to keeping up its violation of the Agreement.

In summary: While it is well-established that contractual escape mechanisms are an indispensable feature of trade agreements, it is the contention of many WTO pundits – trade practitioners, trade lawyers, economists, and international relations scholars alike –

11  The US – Steel case, e.g., patently revealed that Art. XIX safeguards and violation of the Agreement can be used as ready substitutes (Definitive Safeguard Measures on Imports of Certain Steel Products; WT/DS 248, 249, 251, 252, 253, 254, 258, 259).

12  “Measures Affecting Livestock and Meat (Hormones)” (DS 26 and 48), and “Continued Suspension of Obligations in the EC — Hormones Dispute” (DS 320 and DS 321).
that the current system of trade policy flexibility in the WTO does not provide for adequate contractual escape, and therefore is profoundly flawed. In the course of this study we will show that the current system sets the wrong incentives for injurers, and under-compensates victims of escape. This situation may consequently lead, or already have led, to excessive breach, under-commitment (less-than-ideal *ex ante* trade liberalization concessions), and an atmosphere of mistrust. As a result, disgruntled and disillusioned Members have resorted to retaliatory strategies *within* and *outside* the realm of the WTO (e.g. retaliatory antidumping or retaliatory litigation). It could even be argued that the flawed system of trade policy flexibility and enforcement has resulted in a destabilization of the entire multilateral world trading system.

1.2 Objective of the study: Connecting issues of breach, remedies, and commitment level in incomplete contracts

While many commentators remain largely conjectural about the imminence of the WTO’s problems in its system of contractual escape, this study aims to provide a structured, differentiated, and comprehensive approach towards trade policy flexibility in multilateral trade agreements. We want to assess exactly where the WTO system of *ex post* escape is at fault, with what effect, and how it should be improved.

In order to do so, two fundamental aspects must be taken into consideration: First, we can only grasp the full extent of the flexibility debate if it is preceded by a discussion of the intricate connection between trade policy flexibility, contractual enforcement, and *ex ante* commitments. Second, the question “What is wrong?” can only be addressed in a sensible manner once we have a solid conception of “What is right, what is feasible?” in the first place. Without a clear conception of the realistic (attainable) optimum, any reform proposal would be of dubious quality. The realistic optimum flexibility design can then serve as a positive yardstick against which to measure the current system of trade policy flexibility.

We explain both of these aspects in turn (subsections 1.2.1 and 1.2.2) before developing the central research questions of this study (subsection 1.2.3).
1 Introduction and summary

1.2.1 The system of non-performance: Breach and remedies in incomplete contracts

Chart 1.1 captures our general understanding of breach and remedies in trade contracts. It also illustrates the important interlinkage between trade policy flexibility, or contractual escape, and enforcement in incomplete contracts.

Chart 1.1 Non-performance in incomplete contracts

<table>
<thead>
<tr>
<th>Breach (ex post non-performance)</th>
<th>Remedies, damage measures (undo a situation of concern)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intra-contractual breach</strong> (legal): default, escape</td>
<td><strong>Trade Policy Flexibility</strong> e.g. Art. XIX GATT</td>
</tr>
<tr>
<td><strong>Extra-contractual breach</strong> (illegal): defection, violation</td>
<td><strong>Enforcement</strong> e.g. Art. XXIII GATT, DSU</td>
</tr>
<tr>
<td><strong>Intra-contractual remedies</strong>: compensation, indemnity</td>
<td><strong>Extra-contractual remedies</strong>: punishment, sanctions</td>
</tr>
</tbody>
</table>

Source: author
Notes: This chart depicts the relationship between ex post non-performance (breach) and remedies. Depending on whether the breach is intra-contractual (legal), or extra-contractual (illegal), a breach-cum-remedy combination is either called a “trade policy flexibility mechanism” or an “enforcement instrument”.

Every contract – no matter how trivial the underlying transaction or how well it specifies the rights and obligations of the signatories – must have implicit or explicit rules of non-performance. Whilst the definitional terms are not entirely satisfactory, breach and remedies shall used in a broad sense so as to delineate any form of ex post contractual non-performance, and any behavior towards undoing a situation of concern, respectively.13, 14

13 The term breach is somewhat misleading, since in everyday terminology it bears the connotation of extra-contractual, illegal behavior. However, in contract theory, breach is often used to describe lawful opt-out clauses, or liability rules, which allow the injurer to unilaterally decide on contractual performance and non-performance at his discretion. In order to avoid confusion, we will use breach as a generic term for any kind of non-performance. Whenever the word is used as an intra-contractual sense (such as in “efficient breach”), we will put it in quotation marks (“breach”).

14 Following standard contract-theoretical terminology, the term remedy is used in a comprehensive sense, so as to cover any action aimed at undoing unanticipated behavior by one contracting party. It is the generic term encompassing intra-contractual remedies (compensation, indemnity) and extra-contractual sanctions (punishment). Our understanding of the term remedy is notably different from the customary extra-contractual
Ex post non-performance, or breach, of previously agreed contractual commitments can happen in two ways: Firstly, breach is contractually specified and therefore legitimate (“breach”). Whenever parties agree on the permissibility of “breach”, this arrangement – called escape, default, or excuse from obligations – forms an integral part of the contract. Non-performance as agreed upon then represents intra-contractual, permissible behavior, not a violation of the terms of the accord. Generally, escape rules can be organized as opt-out mechanisms, or as renegotiation clauses. A second form of ex post non-performance constitutes extra-contractual behavior and as such is illegal. As a convention, we call this breach behavior defection, or violation of the contract (other terms would be infringement, reneging, deviation, or contractual misdemeanor).

Every act of post-contractual non-performance is necessarily connected to a remedy rule, or a rule of damage. As Chart 1.1. demonstrates, there are intra- and extra-contractual remedies payable to the victim of a violating measure. Those remedies in connection with legitimate escape clauses shall be called compensation or indemnity. Extra-contractual remedies shall be termed punishment or sanctions. In general, remedies are placed on a continuum ranging from zero to infinitely high, or coercive, damages.

In the context of a multilateral trade agreement, a combination of a rule of intra-contractual non-performance and the accompanying remedy procedure together constitute a trade policy flexibility mechanism or instrument (see Chart 1.1). A trade connotation it bears in the WTO literature – or, for that matter, in public international law in general, as spelled out in the ILC Draft on State Responsibility (see Grané 2001; Mavroidis 2000; Vazquez and Jackson 2002): “Remedies” in the WTO context is either used in a broad sense as covering any solution between (two or more) WTO Members once a matter has been formally raised, i.e. consultations have been officially requested under Art. 4 DSU. This broad notion comprehends the type of legal claim by the complaining party (violation, non-violation-, or situation complaint), as well as all ensuing procedural alternatives, including withdrawal of the measure, enactment of alternative measures in compliance with the rules, bilateral settlement, retaliation, and tariff compensation offers (Mavroidis 2005). Usually, however, the notation “remedies” is used in a narrow sense in WTO parlance as legally sanctioned responses pursuant to noncompliance by the injuring WTO Member whose practices have been multilaterally condemned. DSU-remedies, narrowly defined, are comprised of the WTO-legal “countermeasures”, namely retaliation and tariff compensation (Mavroidis 2000, p. 800). Finally, note that our conception of remedies is strictly different from so-called “trade remedies”, a WTO expression that covers the trade contingency measures AD, CvD, and safeguard measures.

As will be explicated in more detail later, escape clauses are usually organized as “liability rules” that grant the initiative and the decision of “opting out” to the injurer. Alternatively, “breach” may happen via “property rules”, which require the potential injurer to renegotiate the terms of the agreement with the victim, and to “buy off” the latter’s right to performance as promised. Some contracts prohibit any sort of ex post discretion, and instead institute a rule of unconditional specific performance (or “inalienability”) that is binding for all signatories at all times.

It should be noted that we use the words punishment and sanction in their customary contract-theoretical, objective, connotation. Neither term is part of the official WTO vocabulary. The DSU speaks of “suspension of concessions or other obligations”, and “damages”, or “trade effects”, respectively. However, in WTO matters, the term sanctions has evolved into a colloquialism for the countermeasure of retaliation. This is not how we will use this expression.
flexibility tool is to be defined as *any* intra-contractual, legal provision legitimizing a departure from performance as promised. The use of trade policy flexibility shall be understood as legitimate *ex post* discretion, structured defection (Rosendorff 2005), selective disengagement (Rodrik 1997, chapter 5), or as a constructive safety valve.

*Extra*-contractual breach behavior and the subsequent punishment shall be bundled together in the term *enforcement*. The WTO deals with issues of enforcement mainly in Arts. XXIII GATT/GATS and the DSU, although some Agreements feature their own dispute settlement clauses (e.g. the Agreement on Safeguards and Countervailing Measures, SCM; or the Agreement on Agriculture).

With this definitional groundwork in place, we can now move on to a discussion of the close interlinkage between mechanisms of flexibility and enforcement, as well as that between the contractual system of non-performance and *ex ante* commitment. Chart 1.2 illustrates this interrelationship:

**Chart 1.2 Commitment, breach, and trade policy flexibility in incomplete contracts**

<table>
<thead>
<tr>
<th>Trade Policy Flexibility Mechanisms</th>
<th>(Self-)Enforcement Instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>(dealing with <em>intra</em>-contractual escape)</td>
<td>(dealing with <em>extra</em>-contractual escape)</td>
</tr>
<tr>
<td><strong>System of non-performance (ex post)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Commitment (ex ante)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Trade liberalization level</strong></td>
<td></td>
</tr>
</tbody>
</table>

Source: author
Notes: This chart shows how trade policy flexibility mechanisms, enforcement instruments and *ex ante* trade liberalization commitments in incomplete contracts are linked together: A proper enforcement scheme encourages the use of *de iure* flexibility mechanisms in situations of *ex post* regret on part of an affected signatory. In anticipation of a functioning system of non-performance, all contracting parties are well-inclined to cooperate and thus willing to undergo extensive upfront commitments. Whenever the system of trade policy flexibility of enforcement is flawed, signatories can be assumed to cut down their pre-contractual concessions.

---

17 We use a broad notion of “trade policy flexibility”. Our understanding of the term differs from some conventional definitions that depict trade policy flexibility as “the ability of governments to decide *unilaterally* when to introduce new *temporary import restrictions* after an international trade agreement has been concluded” (Roitinger 2004, p. 1, emphases added). The difference is thus threefold: First, in this study, trade policy flexibility mechanisms are not reduced to liability rules, i.e. to those instruments assigning the discretion to injurers. Second, we do not discriminate between temporary and permanent flexibility. Third, non-performance is not limited to *ex post import restrictions*, but more generally to *all* agreed-upon contractual behavior (e.g. retreat from a non-reciprocated obligation, such as a notification requirement).
The intuition of the intricate relationship between escape tools and enforcement provisions in a trade agreement is rather straightforward: The more incomplete a contract, the more important is the careful design of viable escape mechanisms.\textsuperscript{18} The availability and quality of the negotiated flexibility mechanism(s) have an immediate impact on extra-contractual breach behavior by potential injurers: Whenever permissible intra-contractual behavior is mis-specified,\textsuperscript{19} injurers under pressure may look for legal loopholes, and resort to extra-contractual, illegal actions – either hoping not to get caught, or because the punishment faced in the aftermath will be smaller than the expected gains from non-performance. Alternatively, if escape mechanisms are not available when needed, injurers may exit the agreement (or rather refrain from signing it in the first place). Another real problem occurs when intra-contractual remedies are such that they over- or under-compensate the victim. Over-compensatory escape clauses are “under”-enacted, under-compensatory ones are used too often compared to what a hypothetical first-best contract would stipulate.

The design of the available enforcement provisions also crucially determines the incentive structure of potential injurers: Whenever enforcement remedies are too lenient (under-compensatory), injurers may decide to violate the agreement, instead of choosing intra-contractual default – as they should.

As a result, three general dangers relating to the system of trade flexibility and enforcement loom large: (i) that of opportunism (inefficient redistribution) on part of the injurer, (ii) that of regret contingencies not seized by the injurer, and (iii) that of insufficient compensation paid to the victim. All those potential pitfalls can be expected to have serious repercussions on the ex ante commitment that signatories are willing to make.

Applied to the WTO context, the extent to which a country agrees to liberalize trade ex ante is a direct reaction to the quality and design of the contractual system of non-

\textsuperscript{18} If – hypothetically – a contract were complete, that is, specified in detail all possible contingencies and prescribed comprehensive plans of actions, flexibility mechanisms would be superfluous. Every ex post non-performance then would by definition be extra-contractual, i.e. deviating, behavior that must to be punished.

\textsuperscript{19} Escape clauses can be said to be mis-specified or ill-defined whenever they are too lax or too restrictive. Flexibility instruments are too rigid if they do not allow signatories to seize regret contingencies, are too expensive to enact, too restrictive in application scope, display ambiguous language, or fail to anticipate certain contingencies completely. They are too lax if they permit injurers to opt out inefficiently often, i.e. more frequently than a hypothetical complete contingent contract would permit.
performance.\footnote{In the case of a trade agreement, \textit{ex ante} commitment can be defined as the scale and scope of trade liberalization concessions or, more generally, as the composition and level of international trade cooperation. \textit{Ex ante} commitments define the gains to be had from international trade cooperation.} Intuitively, if a Member country is not allowed to react to unforeseen developments in a certain industry or sector, it may not be willing to liberalize that sector in the first place. Similarly, if a WTO Member expects to be compensated inadequately for suffering from another Member’s protectionist backtracking, the former will be hesitant to liberalize in the first place.

Prior to the conclusion of the WTO contract, countries do not possess full knowledge of the nature and impact of future contingencies, or of the possible trade policies and instruments that its partners might concoct. Nor can they anticipate whether these contingencies will make them victim or injurer. Therefore, the best a country can do is to assess the efficiency, feasibility, and credibility of the negotiated system of trade policy flexibility and enforcement. It will shape its trade liberalization commitments accordingly.

\subsection*{1.2.2 Knowing what is “right” in order to assess what is “wrong”}

Any contractual flexibility regime must be considered in its context. If we ever want to comprehend and assess what is wrong with breach and remedies in the WTO today, we have to take a step back and ask: “What is the general nature of a multilateral trading agreement such as the WTO; what would the appropriate institutional design of flexibility and enforcement ideally look like?”. In other words, we must assess what system of non-performance we can expect rational trade negotiators to draft in the first place – given the initial negotiation context with all its actors, preferences, trade-offs and constraints.\footnote{Sykes (2000, p. 348) confirms: “[F]rom the perspective of academics interested in the positive political theory of the WTO and of international relations in general, it is important to understand what [institutional framework] WTO members have fashioned for themselves. If we are to theorize successfully about the rules of the game, we must understand the \textit{nature} of those rules at the outset.”}

\subsection*{1.2.3 Research questions}

This study intends to provide a systematic and comprehensive examination of trade policy flexibility and enforcement mechanisms in the WTO. To that end, the following string of questions form the core of our research:
1. What is the optimal design of trade policy flexibility and enforcement in an incomplete multilateral trade agreement? Issues that arise in this context are: How should flexibility in a multilateral trade agreement (such as the WTO) best be organized? Is it practical to have multiple escape mechanisms with overlapping scopes of application? Should Members allow for temporary deviation from all previously agreed commitments? Is trade policy flexibility in the WTO best organized as a ready-to-use escape mechanism that allows any injuring party to (partially) “opt out” of its contractual performance obligation, or as ex post renegotiations between injurer and victim (“buy-out”)? Should a high level of conditionality (preconditions and mandated scope of application) be in place for the enactment of trade flexibility tools? In case ex post discretion is apposite: Which intra-contractual remedies are to be awarded to the victim of such a backtracking measure? Should the victim be put in as good a position as if the injurer had performed? Should the status quo ante the breach be re-established, or rather the status quo ante the contract?22 Alternatively, should the victim instead receive a fair share of the efficiency gains incurred by the non-performance of the injurer? Finally, what kind of enforcement mechanisms, accompanied by which extra-contractual remedies, should be in place to protect those flexibility rules of the game?

As will be shown, any well-crafted system of trade policy flexibility must solve what Ethier (2001a) refers to as the “reciprocal-conflict problem”.23 Ideally, ex post discretion is organized such that it manages to fulfill three crucial criteria: First, trade policy flexibility must allow injurers to seize post-contractual regret and consequently reap all available efficiency gains from non-performance. Second, at the same time injuring Members must compensate the victim countries such that they agree to maximize the scale and scope of their ex ante commitment. Third, the enforcement provisions flanking

22 We will demonstrate infra that these options correspond to the expectation-, the reliance- and the restitution-damage measure, respectively.
23 Ethier (2001, p. 5) describes the reciprocal-conflict problem as follows:

“Each country is aware, ex ante, that it may find itself, ex post, harmed by a policy that some trading partner wishes to make. So the former will want a recognized punishment procedure as a deterrent. But that country will also be aware, ex ante, that it might find, ex post, itself in a position where it would be costly not to take some policy action that would harm a partner. This is the reciprocal-conflict problem: Every country knows that it might turn out to be either the accuser or the accused. Thus it is in no country's best interest, ex ante, to agree that, ex post, either the accuser should be unconstrained in its ability to punish or the accused should be unconstrained in its ability to proceed without punishment” (emphasis in original).
this system must protect the existing system of trade policy flexibility from abuse. In effect, enforcement must deter opportunism in the form of an inefficient breach.24

2. What is the contractual logic – and what are the flaws – of the existing regime of trade policy flexibility and enforcement in the WTO today? How does the contemporary system of contractual non-performance fare compared to the hypothetical benchmark? Is the regime for escape, as currently designed, incentive compatible with injurers’ strategies and victims’ compensation needs? What kind of loopholes encumber the system? Is the regime of enforcement stringent and effective as a protection belt against opportunistic deviation? What are the expected dynamic ex ante effects on signatories’ trade liberalization commitments that the current system of ex post flexibility and enforcement entails?

3. After having assessed what an optimally shaped system of contractual non-performance should look like and how the contemporary WTO system compares to this benchmark: Which reform steps should be taken towards improving trade policy flexibility and enforcement in the WTO?

1.3 Overview of the study and summary of findings

The study is divided into three parts. Part 1, entitled “An introduction to incomplete contracting”, consists of Chapters 2 and 3. It is a basic coverage of contract theory, incomplete contracts, and flexibility mechanisms in general. A novel approach to categorizing incompleteness in contracts will be developed. Part 2 (“Theorizing about the WTO as an incomplete contract”) is comprised of Chapters 4, 5 and 6. Part 2 is a comprehensive contract-theoretical analysis of the WTO. In particular, the current system of trade policy flexibility and enforcement in the WTO is scrutinized for flaws and inconsistencies. Part 3 (Chapters 7 and 8) of the study bears the title “Flexibility and enforcement in the WTO: Towards an agenda for reform”. This final part is a hypothetical bargain analysis of the WTO contract. We speculate on which institutional system of non-performance a group of reasonably rational trade negotiators could be expected to design at the outset of trade cooperation negotiations. In laying out a positive

---

24 A system of trade policy flexibility and enforcement in any incomplete contract must strive to mimic the outcome of the Pareto-efficient complete contingent contract, the unachievable contracting ideal. The ultimate goal of flexibility is thereby to provide for contractual escapes in exactly those instances where the hypothetical complete contract would mandate non-performance (see section 2.4 infra).
benchmark for trade policy flexibility and enforcement, we recommend a detailed reform agenda of the current WTO system.

**Chapter 2** presents an introduction to the nature of contracts in general. We blend approaches from economic contract theory and law and economics (L&E) theories of contracting. A contract is defined as “enforceable commitment over time”. We review reasons for contracting and highlight the intricate connection between signatories’ willingness to cooperate and contract enforcement.

Special consideration is given to the issue of contract design (entitlement choice, entitlement protection rules and enforcement of entitlements). We find that the process of contracting necessarily consists of three steps: Setting out the level of ambition and defining the entitlements to be traded (primary rules of contracting), defining *intra-contractual* rules of entitlement protection and flexibility (secondary rules), and defining rules of enforcement (tertiary rules). Secondary rules of entitlement lay out the scope for *ex post* discretion.

**Chapter 3** is entitled “Incomplete contracting, and the essence of flexibility”. We review sources of contractual incompleteness and transactors’ strategies of overcoming them. In contrast to many other studies, which basically assume some form of incompleteness, we analyze contractual incompleteness, its nature and consequences. We establish a novel categorization of the concept, and point to the inadequate toolkit that standard contract theory has to offer when dealing with contractual incompleteness.

Contractual incompleteness creates room for regret contingencies that signatories need to address. We review various strategies for overcoming contractual incompleteness that reasonably rational actors may choose. Two basic strategic trajectories lend themselves to contracting parties: One towards completion of the contract, and the other to embracing contractual incompleteness. Flexibility mechanisms are designed with the intention to embrace contractual incompleteness and to efficiently seize gains from *ex post* regret. Flexibility provisions are secondary rules of contracting, i.e. rules of entitlement protection. We show that there are two kinds of contractual flexibility mechanisms: Contingency measures and default rules. Contingency measures are thereby special cases of default rules. Default rules are central, because they apply to all previously unspecified (unanticipated, unforeseen, unforeseeable) situations of
contractual regret. Any flexibility tool must be complemented by an *intra*-contractual remedy. We pay special attention to the design of contractual remedies, since they determine the victim’s willingness to cooperate *ex ante*.

Chapter 3 ends with a characterization of the *achievable* first-best flexibility design of any incomplete contract: An efficient “breach” contract is able to replicate exactly the outcome of a complete contingent contract – it is the best *ersatz* contract signatories can possibly craft.

**Chapter 4** is a contract-theoretical examination of the WTO Agreement. We discuss the essence of the WTO contract (reason for contracting, exchanged entitlements, rules of entitlement protection and contractual system of enforcement) and demonstrate why the WTO is a *necessarily* incomplete contract.

We start by reviewing what trade scholarship deems the most important rationales for the conclusion of multilateral trade agreements. We concur with mainstream trade economics that the WTO was accomplished with the paramount aim of overcoming international market access externalities. However, we establish a second(ary) motivation for entering into a trade contract: Countries wish to make the trade in goods and services more efficient by setting regulatory minimum standards and basic rules of conduct.

The WTO is far more than the tariff-liberalization treaty that many scholars, economists in particular, like to characterize it as. Rather, the WTO is introduced as a multi-issue, multi-entitlement contract of a complexity that to the best of our knowledge has not yet been captured adequately by research. The treaty’s multi-dimensionality is imperative for understanding contractual rules of entitlement protection and enforcement. By far the most important entitlement exchanged in the WTO contract is the right to reciprocal trade, or “market access entitlement”, in which countries commit to granting each other mutual market access. Other WTO commitments traded in the WTO contract are subsumed under the term “multilateral entitlements”, i.e. basic auxiliary entitlements and minimum standard entitlements.

---

25 Default rules (also known as backstop-, fallback-, gap-filling-, supplementary-, or background rules) are imperatives “that define the parties’ obligation in the absence of any explicit agreement to the contrary” (Craswell 1999, p. 1).
**Chapter 5** looks at the contemporary system of trade policy flexibility and enforcement in the WTO and discusses its flaws and problems. The verdict is not enthusiastic. Currently, the WTO does not adequately address Members’ needs for contractual escape. The *de iure* system of escape and enforcement is rather dubious: The available escape clauses display too many preconditions, and an insufficient scope of application. The contractual default rules are underdeveloped. DSU enforcement remedies are systemically under-compensatory, and too weak to deter violation of the Agreement. Injuring Members can afford to seek legal loopholes to satisfy their (oftentimes opportunistic) escapist ambitions. We find that the informal WTO trade policy flexibility regime practically annihilates the *de iure* rules, and thereby defies much of what contract theory has to say about efficient entitlement protection. *Violation-cum-retaliation* is the *de facto* default rule for all WTO entitlements. This can be expected to not only crowd out *ex ante* trade liberalization commitments and create significant discontent among WTO Members, but also risks destabilizing the WTO in the long run.

**Chapter 6** is a parenthesis of sorts. Against the backdrop of the insights gained in the two previous chapters, we wish to analyze why the academic “compliance-vs.-rebalancing” debate so far must be seen as a lost opportunity to adequately address the problems which impede the WTO system of non-performance. Reviewing the origin and essence of the debate, we find that scholars of both camps should reframe their controversy, which in reality is not about the object and purpose of WTO enforcement (as they like to think), but is essentially much broader. Based on the previous analysis conducted in Chapters 4 and 5, we suggest refocusing the debate and embedding it in a discussion of trade flexibility and enforcement. We thus find that the compliance/rebalancing debate is actually concerned with the nature of the WTO contract and its system of non-performance. Reframing the questions that should have been at the core of the debate – namely: What are the primary, secondary and tertiary entitlements traded in the WTO contract? – we find that neither the compliance- nor the rebalancing school of thought manages to present a coherent understanding of the WTO contract and its system of non-performance. The rebalancing camp, on the one hand, solely refers to market access-related trade policy flexibility mechanisms, while completely neglecting other WTO entitlements as well as the issue of enforcement. The compliance camp, on the other hand, is excessively focused on issues of enforcement. It fails to distinguish between any entitlements exchanged in the WTO, and neglects fundamental systemic issues connected to trade policy flexibility.
Chapter 7 pieces together the lessons learned from Parts 1 and 2 and conducts a hypothetical bargain analysis of the WTO contract. Following the systematic assessment of why the non-performance regime in the current-day WTO is flawed (Chapter 5), Chapter 7 theorizes about the organization and design of an efficient flexibility and enforcement regime that reasonably rational trade negotiators could be expected to negotiate. How should the various entitlements traded in the WTO ideally be protected—and at what costs for the injurer? Much of the discussion focuses on the optimal trade policy flexibility regime of the salient market-access entitlement. We find evidence that an unconditional liability rule of default backed by expectation damages clearly Pareto-dominates both a rule of inalienability and a renegotiation requirement. Chapter 7 also deals with the relationship between trade policy flexibility mechanisms and enforcement provisions: Escape and enforcement are not rationally designed as mutual substitutes, but as strategic complements. In an incomplete contract, enforcement is the second line of defense of entitlement protection.

On the basis of this comprehensive analysis of the WTO as an incomplete contract, Chapter 8 lays out an agenda for reform which would precipitate a more efficient and viable system of flexibility and enforcement in the WTO. Briefly, the WTO should evolve into an efficient “breach” contract, which would involve three major changes in the current Agreement:

- First, institute a liability rule of default for the market access entitlement. This could be achieved by turning the safeguards clause of Art. XIX GATT (and Art. X GATS) into a simple, non-contingent liability rule of flexibility. That way, WTO Members could react to unforeseen contingencies by unilaterally opting out of previously made trade concessions. For the liability-rule regime to work, WTO Members need binding third-party arbitration, the procedures of which have to be contractually specified. The intra-contractual remedy tied to the unconditional liability rule must amount to the expectation damage measure and is payable to the victim(s) in the form of tariff compensation.

- Second, introduce an unambiguous default rule for all other entitlements, for example by adding an Art. Xbis to the WTO Charter. This article would demand a specific

26 Only the contractual remedy of the expectation damage measure is apt to satisfy the strict efficient “breach” criterion. Expectation damages place the victim in as good a position as it would have been had the injurer performed, and as such constitutes the replacement value of the deal.
performance duty (a property rule), a rule of inalienability, or a rule of liquidated damages, depending on the nature of the multilateral entitlement.

- Third, reorganize the WTO enforcement regime regulated by Art. 22 DSU. This article could be remodeled so as to establish a two-tier system of enforcement: Tier one, an inner protective belt of contractual entitlements, is aimed at dealing with welfare-enhancing good-faith trade disputes (that emerge due to contractual ambiguity, interpretative problems, or unintentional contract infringements), and at solving them in an amicable manner. Remedies at this stage are strictly commensurate to the damage caused. Tier two, the outer layer of protection, mandates punitive and collective punishment. After all, contract enforcement must protect against extra-contractual behavior, not invite it. Given that there is always an efficient safety valve in place for benevolent injurers thanks to the presence of default rules and the first tier of enforcement, WTO enforcement must protect WTO Members against contractual misdemeanor by means of effective penalties.

1.4 A brief survey of the literature on trade policy flexibility and enforcement in the WTO

This study seeks to provide a comprehensive examination of non-performance in the WTO. In this regard, there is a sizeable gap in the literature. Chart 1.3 illustrates schematically the way in which this paper approaches trade policy flexibility more encompassing than many contributions to WTO research on the topic: Established strands of literature (numbered from 1 to 7) have each highlighted some aspects of the triangular relationship between trade policy flexibility, enforcement and ex ante trade liberalization, but have not addressed the inherent interrelationships in a comprehensive manner. The systemic links between trade policy flexibility mechanisms and enforcement instruments, as well as their effect on the pre-contractual trade liberalization commitment of signatories, should be clearly demonstrated.
Introduction and summary

Chart 1.3 Locating the existent WTO literature on trade policy flexibility

<table>
<thead>
<tr>
<th>Trade Policy Flexibility Mechanism(s)</th>
<th>Commitment (trade liberalization level)</th>
<th>(Self-) Enforcement Instrument(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Source: author
Notes: Chart 1.3 revisits Chart 1.2. Various strands of WTO literature have dealt with issues of trade policy flexibility, self-enforcement, and *ex ante* commitment. Most approaches are selective in that they either focus exclusively on one issue (1, 2, 3, 7), or on a bilateral relationship (4, 5, 6). The numbers hereby stand for strands of literature (explained below).

**Strand 1: Descriptive assessments of trade policy flexibility**
Kleen (1989), Finger (1998) and Roitinger (2004), among others, have described and compared over twenty WTO provisions that Members can use as escape mechanisms. However, these sources (with the notable exception of Roitinger 2004) are mainly descriptive in nature. They tell us little about the deeper rationale of trade policy flexibility, or how (and on what basis) to compare and assess the merit of various different mechanisms in the contract. The studies are elusive in explaining how these mechanisms relate to WTO enforcement rules, and how they affect signatories’ *ex ante* willingness to cooperate. Without these systemic links and without any notable theory of trade policy flexibility, purely descriptive assessments of trade policy flexibility tools should be seen as insufficient, and ought to be complemented.

**Strand 2: Literature on single trade policy flexibility tools**
Various contributions have focused exclusively on description and analysis of a single escape tool. They bring forth far-reaching reform suggestions (e.g. Lee and Mah (1998), Sykes (2003), and Finger (2002) on safeguards; Sykes (1989) on countervailing duties; Palmeter (1991a), Messerlin (2000), Lindsey and Ikenson (2003), and Barfield (2005) on antidumping). For our purpose, however, these studies are of limited added value, since there is presumably a high degree of substitutability between individual *de iure* and *de facto* flexibility instruments. It makes little sense to base reform proposals for a particular escape tool on an isolated analysis, since any effect intended by such proposals risks being undone by a rational injurer through an evasion maneuver towards another escape mechanism.
Strand 3: Literature on enforcement and WTO dispute settlement
There is an abundance of literature on systemic flaws in and challenges to WTO enforcement and dispute settlement.\footnote{See Chapter II.D.3.e of the WTO World Trade Report 2007 (WTO 2007), which contains an exhaustive review of the relevant literature. We will take up some of the concerns again in (sub-)sections 5.4.3 and 6.1.} Scholars have found fault with various aspects of the current system of WTO dispute settlement, such as:

- **Participation** (Are developing countries targeted excessively by industrialized countries? Why are least developed countries under-represented in WTO dispute settlements?);
- **Adjudication** (Have dispute settlement panels exceeded their mandate in WTO litigation? Has the dispute settlement system become too legalized? Is there a lack of transparency in dispute settlement?);
- **Implementation** (What are the systemic flaws of the current retaliation and compensation regimes? Are there alternative enforcement remedies?);
- **Arbitration** (Have DSB arbitrators correctly calculated damage awards when interpreting the level of nullification and impairment?).

This literature does not usually take a systemic view, but isolates relevant enforcement concerns from those of trade policy flexibility. One could also argue that these sources often presuppose (treat as implicit) a clear conception of what constitutes right and wrong behavior, and that any contractual gaps should always be filled by dispute panels according to those normative guidelines. Reform demands in this tradition are frequently shaped by the respective author’s normative judgment, rather than a coherent view of the nature of the contract or of signatories’ preferences.

Strand 4: Theories of trade cooperation and self-enforcement
There is unanimity in the literature that in the absence of a supra-national authority any contract among sovereign nations must necessarily rely on self-enforcement. In trying to explain the rationale for trade cooperation, early work on the GATT and subsequently the WTO has solely dealt with self-enforcement. Issues of rules and the organization of trade agreements were neglected (Staiger 1995). Scholars modeled the GATT/WTO as a standard infinitely-repeated prisoners’ dilemma in which cooperation is sustained by the deterrent threat of the grim-trigger strategy of retaliation.\footnote{See Chapter II.D.3.e of the WTO World Trade Report 2007 (WTO 2007), which contains an exhaustive review of the relevant literature. We will take up some of the concerns again in (sub-)sections 5.4.3 and 6.1.} Although these formal economic models accommodate the intricate relationship between enforcement and ex ante commitments (e.g. Bagwell and Staiger 2002b, chapter 6), they nevertheless...
abstracted from reality when assuming a stationary (static) environment. As a result, these early models presented trade agreements as fully efficient and complete contracts that never needed renegotiation or modification. Consequently, trade agreements in formal models of this kind never witnessed deviation, and were free of any kind of disputes (cf. Keck and Schropp 2007, section B).29

Starting with the works of economists like Dixit (1987), Copeland (1990), or Kovenock and Thursby (1992), and international relations (IR) scholars like Downs and Rocke (1995), the WTO literature added realism by moving away from the assumption of a stationary environment. The aforementioned authors integrated the notion of uncertainty concerning future events, and consequently introduced *contractual incompleteness* (the logical nexus between uncertainty and incompleteness will be shown *infra* in Chapter 3 of this study). This prompted trade scholars to examine the role of trade policy flexibility mechanisms in trade agreements.30 However, the focus of this strand of scholarship was on the stability-enhancing role of – exogenously given – trade policy flexibility mechanisms. Authors conjectured and/or proved that escape-clause-induced negotiation equilibria Pareto-dominate other arrangements without contractual safety valves. Flexibility mechanisms were thus proven to successfully decrease the breakdown risk of the trade game and to add to the stability of the world trading system (see e.g. Goldstein et al. 2000; Herzing 2005; Koremenos et al. 2001; Rosendorff 2005; Rosendorff and Milner 2001).

While this literature is important in establishing the rationale for trade policy flexibility mechanisms in trade agreements, it fails on four accounts for the purpose of this study: First, it remains largely elusive about the impact of trade policy flexibility on *ex ante* commitments (with the notable exception of Herzing 2005). Second, the formal models assume the presence of one (exogenously given) single escape mechanism, namely an unconditional liability-type opt-out. Third, they take the WTO contract to consist uniquely of reciprocal *tariff*-reduction commitments. Lastly, these approaches fail to differentiate between an *intra*-contractual escape clause, and violation of the WTO Agreement. In short, they take violation of the contract to be just another opt-out

---


29 To be more precise, in these models disputes are equivalent to the breakdown of the system: Under a fully efficient trade agreement any deviation from the specified terms is necessarily opportunistic and thus automatically provokes the grim-trigger response.

30 Confer subsection 7.2.2.1 for modeling details.
mechanism. These contributions to WTO scholarship thus fail to discuss the crucial issues of the nature of the contract, the escape clause design, and how trade policy flexibility relates to enforcement and links back to \textit{ex ante} commitments.

**Strand 5: Linking trade flexibility and commitment level**

Scholars of this category have examined the systemic link between trade policy flexibility and commitment level, or level of trade liberalization. Whereas scholars in strand 4 of the literature have studied the stability-enhancing qualities of trade policy flexibility, this literature takes a different route. It assesses the beneficial character of escape clauses in capturing contractual regret in incomplete contracts, and how these provisions consequently shape the \textit{ex ante} willingness of governments to liberalize trade in the first place.\footnote{See, for example, Sykes (1991, p. 279) who states: “[I]n the absence of an escape clause, trade negotiators may decline to make certain reciprocal concessions for fear of adverse political consequences in the future. But with an escape clause in place the negotiators will agree on a greater number of reciprocal concessions, knowing that those concessions can be avoided later if political conditions so dictate.”} Examples of this line of research include Sykes (1991), Ethier (2001a; 2002), Schwartz and Sykes (2002b), Roitinger (2004), Mahlstein and Schropp (2007), and Herzing (2005). These studies discuss the incentive-compatibility constraints posed by \textit{intra}-contractual remedies. In other words, these contributions bring in the concerns of \textit{victims} of flexibility mechanisms by raising attention to the fact that adequate compensation for the victim is important.

Sykes (1991), Ethier (2001a), Mahlstein and Schropp (2007), and Herzing (2005, chapter 3) are examples that formally calculate the optimal remedies payable to a victim of \textit{ex post} contractual escape. These authors go on to discuss the dynamic relationship between \textit{ex ante} trade liberalization commitments and \textit{ex post} flexibility.\footnote{All of these approaches, however, consider only one flexibility mechanism, namely Art. XIX GATT or violation of the Agreement, respectively. Also, all models reduce complexity by representing the WTO as a reciprocal \textit{tariff-reduction} contract.} Schwartz and Sykes (2002) assess the impact of trade policy flexibility design on \textit{ex ante} trade liberalization commitments, but their analysis is too narrow in that it only broaches issues in connection with DSU enforcement (i.e. violation of the Agreement as informal opt-out). Roitinger (2004) offers a thorough review of trade flexibility in the WTO, although the author’s criteria for comparing the effectiveness of different escape mechanisms (\textit{ibid.}, chapter 2.2) seem somewhat arbitrary. In addition, his examination of the connection between \textit{ex ante} trade liberalization and \textit{ex post} flexibility remains largely suggestive. None of the papers mentioned analyze the relationship between enforcement and trade flexibility in sufficient detail.
Strand 6: Theories linking trade policy flexibility to enforcement

WTO scholars have linked the issue of trade policy flexibility to that of DSU enforcement. Approaches assessing the political economy of protectionism in the WTO seek to explain the trade-off by escapist Member governments as that between enacting some trade policy flexibility mechanism, and a blatant violation of the Agreement – given the constraint set presented by the contemporary WTO system (examples include Anderson 2002; Barton, et al. 2006; Bown 2002b; Rosendorff 1996; Schropp 2005; Tharakan 1995; Finger 1991; Finger, et al. 1982; Sykes 1989). This line of research is paralleled by more formal work in international economics which endogenizes governments’ choices between one specific escape mechanism and defecting from the Agreement. Models in this line of research explain players’ protectionist incentive structures when deciding whether to utilize AD measures or to violate the Agreement (see e.g. of Bloningen and Bown (2003), Bown (2001), and Martin and Vergote (2004)), or whether to enact safeguard action or opt for violation-cum-retaliation (see e.g. Bown 2002a).

This literature sensitizes readers to the systemic link between trade policy flexibility and enforcement, but must be supplemented for the following reasons: First, it presupposes a notion of the contract: It explains governments’ choices within the existing system without discussing why the system is the way it is, and how things can be improved. Second, some of the formal work may just as well be termed models of protectionism, because the authors allege that injurers wish to seize legal loopholes only in order to opportunistically opt out of the system. The welfare-enhancing value of non-performance (seizing regret), as dealt with in Strand 5, is hardly ever addressed. Finally, all approaches remain conjectural at best, when it comes to the intricate systemic effects of protectionist opt-out on the expected commitment level of WTO signatories.

Strand 7: The compliance-vs.-rebalancing debate

The so-called compliance/rebalancing debate involved some of the most prominent WTO scholars from the fields of international economics, L&E, international relations, and international trade law. Disagreement concerning the objective of WTO enforcement and the general “bindingness” of WTO dispute panels formed the core of the debate. Yet the opportunity to discuss the WTO’s system of non-performance at large lapsed un-seized, and the debate hit a dead end. Chapter 6 will document in detail why this discussion was not able to resolve the underlying issues of trade policy flexibility, enforcement and trade liberalization concessions.
1.5 Methodology and research approach

Influential scholars of the L&E discipline have suggested that there is potential for a rich research agenda in approaching international contracts (treaties) from a contract-theoretical perspective (Bhandari and Sykes 1998; Dunoff and Trachtman 1999; Posner 1988). This potential, however, seems to be currently underutilized. To date, a thorough and systematic contractarian exploration of the world trading system and the WTO is absent from WTO scholarship. The point of departure of this study is the assumption that the WTO is a contract of sorts. We embrace this assumption to take the WTO treaty as the international equivalent of a contract: a multilateral, multi-issue, long-term, repeated-interaction, incomplete contract.

Two methodological building blocks substantiate our contractarian approach to the WTO: (i) A theory of decision-making, and (ii) a theory of the locus of decision-making. When discussing the WTO contract, we are guided by the fundamental assumptions that all relevant agents are self-interested, utility-maximizing individuals, and that the locus of decision-making resides with trade policymakers, because as the legal representatives of sovereign countries it is they who conduct international trade negotiations.

(i) Methodological individualism forms the very basis of a range of disciplines, including economics and contract theory, and as such also of this study. All actors, be they policymakers, consumers, or producers, are homines oeconomici, that is, they behave in a rational, self-interested, and reasonably far-sighted manner (e.g. Hausman and McPherson 1996; Kirchgässner 2000). When these agents interact, there is a natural drive towards efficiency, and the Pareto-criterion becomes of crucial importance.

(ii) The second key methodological ingredient is the locus of decision-making: When examining the contracting process in the WTO, it is necessarily to have a concept of who takes the decisions to cooperate, how to structure this trade cooperation, and when to breach, renegotiate or terminate this contractual relationship. As such, a

---

33 Sykes (1999 at p. 1127) remarks: “Much like other subject areas under the rubric of international law, law and economics has only begun to make a dent in the set of potential topics in the trade area.”

34 Trachtman and Dunoff (1999, p. 27) correctly contend that “methodological individualism actually requires a contractarian approach”.

35 We do not assume that individuals are rational supercomputers: Although they engage in careful cost-benefit considerations and continuous learning, actors are not always perfectly informed, do not operate flawlessly, and lack the power (or will) to anticipate all future events. To that end, we use the term of “reasonably rational” individuals (see subsection 3.1.4).
theory of domestic trade policymaking is necessitated. We thereby rely on the disciplines of public choice and political economy. Political economy scholars have brought forth the now-standard paradigm that signatories to a trade agreement are represented by rational, self-interested policymakers whose policy decisions are directly influenced by various domestic stakeholders and international externalities. In short, we follow the lead of Sykes (1991; 2000) and Schwartz and Sykes (1996; 2002b), who suggest that the WTO, “like its predecessor the GATT, is a multiparty contract designed to secure joint political gains to its signatories” (Sykes 2000, p.254), whereby these authors, too, take “signatories” to be individual self-interested policymakers.

Based on the methodological proposition that the WTO Agreement is in effect a contract aimed at fulfilling the selfish needs of those trade policymakers who sign the treaty, we feel ready to lay out our research approach. This study is purely theoretical in nature. It is a contribution to the “growing theoretical literature on the positive political economy of the World Trade Organization” (Schwartz and Sykes 2002a, p. 2). Without an adequate theoretical framework any discussion of WTO reform risks being vacuous. It is a central task of this study to establish a solid framework for theorizing about incomplete contracts in general, and the WTO in particular.

This requires two essential components: First, we must provide an understanding of the nature of the contract at hand. In order to grasp the nature of the WTO contract, this study conducts a stock-taking exercise with a twist: We survey the relevant contract-theoretic literature in the fields of L&E, economic contract theory and IR, and adapt it to the special contractual context of the WTO as appropriate. The defined outcome is a complete theory of the WTO from a contractual point of view. We thereby have to avoid the pitfall of a-contextualism that the argumentation of the compliance/rebalancing

36 Political economy is the concept of gain derived from mutual exchange between individuals, applied to the realm of political decision-making. In trade policy, political economy approaches may be considered the state of the art methodology (see Grossman and Helpman 2001; Rodrik 1995; Staiger 1995 for excellent literature overviews).

37 This statement bears two far-reaching implications: The first one is the now-established standard assumption of “politically realistic” government utility functions (Baldwin 1987), in which politically motivated policymakers have particular domestic distribution concerns. Different groups of society bring to bear unequal influence in the domestic trade-policymaking process. Second, trade policymakers care about international spillovers. The presence of international externalities gives rise to the mutual motivation of concluding trade agreements so as to effectively secure market access commitments from co-signatories (see section 4.1).

38 Sykes submitted as early as 1991 (p. 274) that seeing the GATT as a mutually advantageous contract among self-interested political officials provides for convincing answers concerning the nature and design of the trade accord.
debate is beset with (see Chapter 6). Some effort is thus spent on introducing the nature of the WTO contract, and on deliberating to what extent outcomes of standard (economic) contract theory can be adopted.

The second key ingredient to a framework of theorizing about the WTO is to conduct a “hypothetical bargain analysis” (Scott 1990, p. 598) of trade policy flexibility and enforcement. We examine what kind of trade agreement rational trade negotiators can be expected to discuss. We establish a positive benchmark against which the reality of non-performance in the current WTO accord can be measured. This benchmark then functions as an archetype of a governance structure which rational trade negotiators can be expected to agree upon, given the context negotiations are embedded in. This research approach to the nature of the WTO contract, its rules and procedures is somewhat hybrid – neither purely prescriptive, nor purely positive. Yet the analysis necessarily contains vital positive elements, namely in the form of the paradigm of methodological individualism underlying all economic logic.

1.6 Contribution to scholarship

Are the central questions to the study concerning the permissibility and organization of contractual non-performance in the WTO merely an academic exercise devoid of any practical significance to the global trading system? We think not.

First, the meaning and the implications of the WTO as an “efficient ‘breach’ contract” deserves further attention: Whereas scholars tend to set efficient “breach” equal to a contractual rule of unconditional liability across the board, our analysis results in a
more nuanced picture of what it really takes to install a truly efficient system of non-performance in an incomplete contract such as the WTO. The efficient “breach” condition is somewhat more complex than some WTO scholars like to believe: Its efficiency lies in the property that all gains from trade in a non-stationary world are being exhausted – not in the fact that the contract gets “breached” efficiently often. This requires the presence of a system of trade policy flexibility mechanisms which mandates the injurer to internalize the costs of default.\(^5\) This system of flexibility – which may or may not be provided through a liability rule of flexibility – needs to be backed up or protected by a system of enforcement that prevents opportunistic abuse of the flexibility system. Lastly, any efficient system of non-performance must be incentive compatible to both victims and injurers, that is, it must secure optimal \textit{ex ante}, or ongoing, commitments. Any system that fails to do so cannot be called efficient (more in subsection 3.4.1).

Second, our study may help understand the dynamics between ongoing trade liberalization and the existing system of non-performance in the WTO: Obviously, the success of the current trade round (the Doha Development Agenda or DDA) is to some degree influenced by systemic flaws in the system.\(^6\) Without overstating the case, we point out that the critical deadlock the DDA is currently facing may have something to do with its insufficient system of rule enforcement and contractual escape.\(^7\) However, this important aspect can only be understood and appreciated properly if the intellectual systemic links between trade policy flexibility, enforcement, and commitment level and scope are established.

breaching party to pay damages to redress the harm (loss of profit, etc.) without performing its obligations under the contract. In most cases, the nonbreaching party will be made ‘whole’ and, in some cases, even better-off. Thus, the breaching party has the option of refusing further performance if its compensation fully protects the nonbreaching party’s reasonable economic expectations from performance of the contract” (Jackson 2004, p. 122).

\(^5\) Put differently, the contractual system of flexibility must guarantee the victim the receipt of enough indemnity to place it in as good a position as if the injurer had performed (Posner 1988, pp. 110) – anything less would result in Pareto inferiority.

\(^6\) Bagwell and Staiger (2002b, chapter 6) call the intricate connection between trade round negotiations and \textit{ex post} non-performance the “intertemporal balance of concessions”. Intuitively, policymakers cannot be expected to liberalize freely, when they know that the current system of non-performance is flawed.

\(^7\) As Roitinger (2004, p. 204) contends: “[O]ne is tempted to ask if the main obstacle for a smooth process of progressive liberalization is not rather of a \textit{structural} nature […T]he key problem might not exist in the diverging interests among members over which sectors to liberalise first. Instead, the problem could be hidden in the current institutional design of the world trading order” (emphasis in original).
Third, our study may bring WTO scholarship a step closer towards a coherent theory of disputes. One crucial shortcoming of WTO research, whether originating from the disciplines of trade law, international economics, political economy, or international relations, lies in the absence of a viable theory of disputes – how they emerge, and how to avoid and resolve them (Mavroidis 2007, chapter 5.4.9 at p. 590). This study is geared towards designing efficient trade policy flexibility and enforcement provisions. Ideally, efficient escape clauses are such that they can separate good-faith clashes (Pareto-efficient non-performance) from opportunistic bad-faith clashes. Although reality adds contractual ambiguity, ambivalence, opposite interpretation, and misunderstandings – and thereby blurs the lines between good-faith and bad-faith clashes considerably – a well-crafted system of trade policy flexibility and enforcement should weed out many of the WTO disputes that are currently overloading the dispute settlement process of the WTO.
Part 1

An introduction to incomplete contracting

Part 1 of the study addresses the basics of complete and incomplete contracting. Chapter 2 provides an introduction to the nature of contracts in general. We will discuss the nature of complete contracts and the contracting ideal, the Pareto-efficient complete contingent contract. Chapter 3 assesses the sources of contractual incompleteness, and present signatories’ available strategies for overcoming the inevitable incompleteness that besets all real-life contracts. Since the aim of this section is to explicate the whole range of contracts, and to carve out central characteristics and properties of contracts at large, the topic is treated in a fairly generalist manner.

We will demonstrate that a contract is determined by its context, including the nature of contingencies, of entitlements traded, entitlement-specific *ex post* transaction costs, the number of players, the time-horizon of a contract and the frequency of transaction. Signatories must adapt the contractual design to the context of the situation. To judge whether they did a good, mediocre, or bad job at that, it is vital to understand the nature of the underlying contract in the first place.
2 Complete contracts, and the contracting ideal

In order to comprehend the intricacies of *incomplete* contracting, we should first get a grasp of *complete* contracting and the overall nature of contracts. This chapter represents a general introduction to the economic theory of *complete* contracts and of contract design.48 We will deal with contracts in the most general way so as to include all sorts and types of contract – from simple handshake barters to national constitutions or sophisticated covenants with thousands of pages and hundreds of signatories. It should be acknowledged that owing to the diversity of real-life contracts, a general introduction to complete contracting is a daunting task, and one that necessarily implies painting with a broad brush.

The difficulties of generalization notwithstanding, all contracts share some fundamental properties. It is these commonalities we intend to concentrate on here to distill the essence of contracts, i.e. to filter out the common characteristics, motivations and principles of contractual relationships.49

We will examine what constitutes a contract, why parties enter into contracts, and what determines signatories’ *ex ante* willingness to cooperate in a self-enforcing contract (section 2.1). Next, we will assess the basic design of contracts (section 2.2), and distinguish different kinds of contracts (section 2.3). Finally, the contracting ideal of the Pareto-efficient complete contingent contract as the unachievable first-best will be introduced in section 2.4.

2.1 Contracts: Enforceable commitment over time

Among the many definitions of a contract in the L&E and economics literature, the essence of a contractual exchange is best captured as “an enforceable mutual commitment over time” (Craswell 1999, p. 18; Dunoff and Trachtman 1999, p. 30). This definition integrates the three features that are essential to understanding contracts: The

48 This is not a study in the discipline of law. We will not provide a legal introduction to contracts in this chapter or the next.

49 Note that the following introduction is but one view on the principles of contractual logic and the inherent trade-offs connected with forming contracts. Other authors may put the emphasis elsewhere. But this introduction suffices to prepare the ground for later chapters.
time dimension, commitment (or cooperation) as the currency of the agreement and credible (effective) enforcement.

2.1.1 Timing

In contrast to an instant-barter, swap, or a spot-transaction where a good or service is handed over instantaneously in return for another good, service or money (a concept economists usually refer to as “transaction”), the points in time between agreement/consent and delivery in a contract necessarily diverge. By default the subject of exchange (the transacted “item” or “good”) in a contract is not a commodity, merchandise, natural resource, or service, but a promise of future delivery: a commitment.50

2.1.2 Commitment: Cooperative intent and assurance

The essence of a contract, hence, is not the exchange of goods and services per se, but rather the promise thereof. Credible commitments are the currency of a contract. A commitment is constituted by two salient components: the motivational/cooperative aspect and the assurance/enforcement aspect.

The motivational aspect can be called the “contractual intent”, or what Schwartz (1992, p. 284) calls “substantive goals” of an agreement. Every contract is concluded for a purpose, and this purpose is cooperation in some issue area or another. Cooperation is geared towards increasing the welfare of not just one, but of all the signatory parties, by means of reaping the ensuing gains from trade.51 Contracting parties usually engage in a

---

50 Craswell (1999, p. 18) submits: “There is an important difference between permitting free exchange and permitting binding promises. An exchange can take place instantaneously, but a promise necessarily involves a commitment to act in a certain way at some time in the future. Once this temporal element is recognized, it can be seen that enforcing promises does not simply transfer existing goods from one owner to another. Instead; the enforcement of promises creates a new good” (emphases in original).

51 Throughout this study (except when expressly noted otherwise), we endorse the tenet of “rational”, instead of “constrained”, choice. This is to say, we take a strictly Paretian view of contracting: Signatories are believed to strive for maximization of their own, private welfare. Mutual interaction then leads to a maximization of the absolute gains of interaction. The (bounded) rationality of all actors prevents contracting parties from entering into agreements that are harmful, abusive, or coercive to an individual transactor in the short or long term. This Paretian view of rational choice stands in contrast to concepts of constrained choice: “Realpolitik” (realist or neorealist) theories of IR have argued that cooperation (especially between States) is geared towards achieving Kaldor-Hicks (distributive) efficiency gains between the players, but not Pareto efficiency. Under this realist conception of efficiency, contracting parties care about relative gains and largely ignore (or take for granted) absolute gains. To illustrate, a non-Paretian view of the WTO contract will argue that countries conclude trade agreements (or refrain from doing so) in order to keep trade rivals in check (e.g. Gowa 1989; Gowa and Mansfield 1993; Grieco 1990; Grieco and Ikenberry 2003; but see Baldwin 1993 for an empirical critique). Also, some neorealism IR scholars contend that powerful countries have regularly coerced smaller countries into joining trade agreements (e.g. Gruber 2000; Jawara and Kwa 2004). Interesting, or challenging, as
cooperative relationship with other transactors with one or a combination of the following objectives:

1. **Minimizing transaction costs.** Contracts significantly reduce *pre-contractual* transaction costs.\(^{52}\) By concluding a contract, signatories can reap efficiencies through avoiding *ex ante* sorting, searching, information gathering and bargaining costs in contexts where additional information serves merely to redistribute rather than to expand the available surplus (Kenney and Klein 1983; Masten 1999). In a repeated-interaction setting, for example, considerable transaction costs may be saved through the conclusion of a contract, since parties dispense themselves from having to re-bargain the terms of a contractual exchange over and over again.

2. **Engaging in risk transfer.** In risk-transfer transactions, the contractual objective is to shift risk to the less risk-averse transactor or low-cost risk bearer (cf. Masten 1999, p. 27). Throughout this paper we will assume *risk-neutral* actors. Therefore, issues of risk transfer will be neglected in all future considerations.

3. **Reaping interpersonal gains from trade.** Probably the most important reason for concluding contracts is the aim of achieving direct transaction efficiencies created by the exchange (handing over) of goods and services. Transaction efficiencies are generated *ex post* through the conversion of the promise (the delivery of a good, a service, money, a concession, risk-transfer, etc.) into a transaction. As in an on-the-spot barter, *ex post* transaction efficiencies are reaped as long as each party values the item/good received more than the item/good foregone (Mahoney 1999, p. 117).

The second component defining a credible commitment is *assurance*. Since in a contract the moments of consent and delivery by definition diverge, assurance against other actors’ defection is an important factor for contracting parties. Before entering into the contract and making a binding promise containing far-reaching commitments, each party will assess the risk of the other party’s contractual non-performance, and want to take the

---

\(^{52}\) For the purpose at hand, “transaction costs” (TC) is a general term for all those real-life costs that signatories must incur when cooperating. Such costs include *pre*-contractual sorting and searching-, information gathering- and processing costs, costs of bargaining, as well as *post*-contractual litigation-, enforcement- and policing costs. Early work on transaction costs dates back to Coase (1937) and Williamson (1985; 1979). More on TC *infra* in Chapter 3.
necessary precautions against such incidents.\textsuperscript{53} Most importantly, deliberate violation of the letter or spirit of the agreement terms, or \textit{ex post} opportunism, has to be contractually anticipated and forestalled.\textsuperscript{54} To this end, rational and far-sighted parties often give the contract \textit{additional structure} over and above a mere specification of their envisioned terms of trade. Schwartz (1992, p. 284) refers to those contractual clauses that specify the “parties’ desire to achieve substantive goals in the best way” as “contracting” clauses. Contracting clauses hence only have an indirect, secondary significance for the covenant.

2.1.3 Effective enforcement and the link between commitment and enforcement\textsuperscript{55}

Enforcement is the third key feature of contracts, next to timing and commitment. Every contract needs to be enforceable – whether it is a quasi-instant barter, such as the purchase of a candy bar (where only a few seconds lie between consent and conversion of the promise), a more complicated employment contract, or a complex long-term, repeated-interaction treaty (such as a multilateral trade agreement). Enforcement is the \textit{sine qua non} of contracts, giving effect to the mutual commitment and deterring defection. Without enforcement the contract is likely to break down, or, more probable, is not concluded in the first place.

Enforcement is a function of “enforcement capacity” and “enforceability” (see WTO 2007, section II.C.5). Enforcement capacity is the ability to reciprocate credibly against violation of any terms of the contract; it is the capability to sanction contractual misdemeanors. Punishment can thereby be exercised by the affected party itself (self-enforcement), by a neutral third party,\textsuperscript{56} by society at large, or through collective enforcement enacted by a circle of affected or concerned parties (such as the membership of a multilateral contract). Enforcement instruments can vary from physical (incarceration), to economic (penalty fees) or emotional (reputation loss, withdrawal of affection) measures.

\textsuperscript{53} Masten (1999, p. 26) points out: “Without some form of assurance that others will, when the time comes, uphold their end of the bargain, individuals will justifiably be reluctant to make investments, forgo opportunities, or take other actions necessary to realize the full value of exchange.”

\textsuperscript{54} Opportunism has been defined as “self-interest seeking with a guile” by Williamson (1985 at p. 47; see also Williamson 1979, p. 234 at footnote 3). A more precise definition is “inefficient redistribution” (Cohen 1999, p. 90): In contrast to a lump-sum transfer, opportunism is always welfare-depreciating.

\textsuperscript{55} This section draws heavily on Keck and Schropp (2007).

\textsuperscript{56} Third-party enforcement may be called “court-and-copper” enforcement, since constitutional States require that a judiciary determines a legal infringement, and an executive (police) enforces the law.
Enforceability of a contract is another vital determinant of enforcement. Typically, issues of observability, verifiability and quantifiability feature prominently when it comes to enforceability: Observability is the property that contract infringements can be detected in the first place – either by the affected party itself, or by a third party (say, an attorney or prosecutor). Verifiability is concerned with the issue of whether the contract can actually be enforced as written or agreed upon: A violation or infringement is verifiable if the affected party can point to a clause in the contract and prove its contravention. This presupposes that such a clause is actually mentioned explicitly in the contract and can be interpreted unambiguously (i.e. that the contract does not contain a gap). Also it presupposes that the violation can be conclusively proven to a neutral third party.\textsuperscript{57} Quantifiability is the ability of the aggrieved party (or a court\textsuperscript{58}) to calculate or quantify the harm suffered as a result of the contract breach. In a way, quantifiability is verifiability of damage incurred by the victim of a contract violation.

Another – maybe more intuitive – way of thinking about the two dimensions enforceability and enforcement capacity is the following: Contractual enforcement is composed of two phases – a dispute or litigation phase, and a punishment or remediation phase. The litigation phase features issues of enforceability, while the remediation phase deals with issues of enforcement capacity.

An important aspect of contracts is the intricate link between the quality of enforcement on the one hand, and the level of ex ante commitment, or cooperation, on the other. Cooperation rarely is a binary issue, but rather a matter of degrees.\textsuperscript{59} For illustrative purposes consider Chart 2.1. The vertical axis shows various degrees of cooperation ($C$),\textsuperscript{60} where ($C_{\text{max}}$) is the point of full cooperation and ($C_N$) is the point where no contract exists.\textsuperscript{61} The vertical axis depicts the excess (dis-) utility generated by the contract. Along this axis we display gains reaped over and above a non-cooperative

\begin{itemize}
\item As to the difference between observability and verifiability (Schwartz 1992, p. 279): “[I]nformation is observable when it is worthwhile for the parties to know it, but the cost of proving it to a third party exceed the gains; information is verifiable when it is both observable and worth proving to outsiders.”
\item Throughout this study, we will use the word “court” as shorthand for an impartial, independent decision-maker or a group of decision-makers. Thus, by the actions of a court (be it adjudication, arbitration, or monitoring), we mean to include that proceedings are conducted by a jury, an arbitrator, an expert agency, a dispute panel, and so forth.
\item Think about a commercial bank assessing how much credit it should give to a client: As much as the client asks for, less than that, or nothing at all?
\item We define cooperation loosely here so as to encompass many forms of inter-personal collaboration and across many issue areas.
\item Note that point $C_N$ represents the Nash-level of cooperation, which is the degree of cooperation with no contract in place. The cooperation at the Nash-level may well be positive and different from a zero transaction-level.
\end{itemize}
2 Complete contracts

situation, and costs of being punished for having defected. Point $U^N$ in the origin corresponds to the Nash-cooperation point $C^N$ in absence of a contract. Suppose a contract between two players.\(^{62}\) The contractual intent, i.e. reason for contracting, is of no consequence here. Assume, however, that we are dealing with a long-term contract of repeated interaction (such as an employment contract).

**Chart 2.1 Enforcement constraint in contracts**

![Chart 2.1 Enforcement constraint in contracts](chart.png)

Source: Keck and Schropp (2007, Figure 1).

Notes: This chart shows trade-offs and constraints from the point of view of an injuring party. *H&R* stands for “hit-and-run”. The *H&R*-curve represents the discounted benefits of defecting from the terms of the initial contract. *S-E* stands for self-enforcement. The *S-E*-curve represents the expected costs (disutility) of defection in a self-enforcing agreement. Under a grim trigger enforcement strategy, those costs are tantamount to the foregone discounted value from future cooperation. *C&C* stands for external “court-and-copper”-enforcement. The *C&C*-line is the expected disutility from defection (here, a liquidated damage clause) in a contract enforceable by a third party. Utility level $U^{opt}$ corresponds to the optimal level of cooperation ($C^{opt}$).

Each contracting party has an immediate incentive to cheat on a contract by deviating optimally from the initial terms of the agreement. The short-term benefit from defection (*hit-and-run advantage*) is the additional one-shot utility the injurer enjoys from

---

\(^{62}\) For simplicity, parties are assumed to be symmetrical: Only one contractor needs to be examined; incentives and actions by the other party are identical. This is without loss of generality. In a model with multiple actors the enforcement can be represented as a two-player game, namely between a player “$X$” and a player “rest of the world”.
defecting over and above the gain from cooperating as promised. It is a short-term benefit only, since it merely stretches from the moment of defection to the moment this violation is detected or enforced.\textsuperscript{63} The hit-and-run benefit (if seized by the potential defector) is by definition an opportunistic, hence inefficient, redistribution of welfare to the detriment of the potential victim (not pictured). This raises the central concern of contracting parties as they design a mutual agreement: How can rule-obedience, that is, continuous commitment to cooperation, be safeguarded? Issues of enforcement come into play.

Two enforcement mechanisms are depicted in Chart 2.1: First, in the absence of a central enforcement instance (such as a court) to deal with the case, and if parties interact repeatedly over time, the best a victim can do to enforce the contract is to engage in self-enforcement: The victim retaliates by exiting the agreement and returning to the non-cooperative past (threat of grim trigger).\textsuperscript{64} The curve \( S-E \) represents the injurer’s opportunity costs of reprisal, that is, the expected discounted value of future cooperation that the injurer foregoes by destroying the cooperation game.\textsuperscript{65} The discounted value of cooperation is the sum of all future per-period gains from having concluded a contract (as opposed to having no contract at all).\textsuperscript{66}

The potential defector balances the short-term incentive to cheat against the long-term cost incurred by such cheating, once the victim retaliates by suspending cooperation indefinitely: Gains and losses approach zero if the contract stipulates little cooperation.

\textsuperscript{63} The convex curvature of line \( H&R \) is intuitive: The curve is flat and equal to zero at \( C^N \), where the contracted cooperation is equivalent to the situation without agreement. The higher the agreed-upon level of cooperation (i.e. the higher the \textit{ex ante} commitment), the more a one-time defection pays off: Increasing marginal returns from defection. The convexity of the curve is not necessary. Bagwell and Staiger (2002b, p. 102) supply some arguments in favor of this curvature for the case of a trade agreement/contract.

\textsuperscript{64} Alternative – less drastic – punishment rules are of course possible, for example the infamous “tit-for-tat” strategy of enforcement, where the victim of a defection retaliates for exactly one round and then returns to cooperative behavior. It is sometimes argued that tit-for-tat retaliation produces more stable outcomes than the grim trigger strategy (e.g. Axelrod 1984; Axelrod and Keohane 1986; Oye 1986). For the situation at hand, however, the retaliation strategy chosen by the victim, or agreed to by both signatories, is inconsequential, since it may be assumed that the nature of resumed cooperation is linked to the history of players’ choices. On the latter issue, see for instance Ludema (2001).

\textsuperscript{65} Indeed, the defector’s expected disutility from defection is identical to the sum of forfeited future benefits from cooperation that the injurer forsakes by having defected once and therefore having prompted the grim trigger response.

\textsuperscript{66} The concave curvature of the \( S-E \) curve is intuitive: More cooperation in every round is beneficial up to some optimal point \( C^p \). After that point additional commitments in the form of more cooperation display declining (possibly even negative) returns. This is so, for example, because of a loss of freedom and sovereignty caused by overly zealous levels of cooperation. Bagwell and Staiger (2002b, p. 102) give the intuition for concavity of long-term cooperation gains in the case of trade agreements. The \( S-E \) curve also goes through the origin: The more the negotiated commitment approaches the no-contract level, the smaller are the future gains from cooperation.
since this practically replicates the non-cooperative Nash outcome. If the contract specifies greater cooperation (levels above $C^N$), short-term defection and long-term opportunity costs begin to rise. As the agreed-upon cooperation increases, the costs of reprisal first exceed the hit-and-run gains from one-time defection. The two curves intersect at $C^{SE}$, which can be defined as the most-cooperative level of concessions that can be sustained through self-enforcement. Beyond this point, the gains from one-time infringement of the agreement exceed all the compiled future gains of cooperation. It is thus evident that it would be foolish (irrational) for the injuring party to comply beyond a cooperation level of $C^{SE}$ due to its binding incentive constraint. Anticipating the injurer’s behavior, it is equally irrational for the victim to agree to a more cooperative deal, even though its welfare-optimal level of cooperation would equally be $C^{opt}$ (due to symmetry of the players). Hence, without a central enforcer only the range between $C^N$ and $C^{SE}$ is self-enforceable, yielding an additional utility in the range between $U^N$ and $U^{SE}$. The levels of cooperation and the according utility range of the contract that are achievable under self-enforcement are represented by the grey area.

To understand a second mechanism of enforcement, take a look at the line C&C in Chart 2.1: If there is a central enforcer – a Court to detect/verify an infringement, as well as Coppers to enforce the court’s verdict – that is capable of and willing to implement the contract under pain of penalty, contracting parties conclude their agreement in the “shadow of the law”. Under these circumstances parties can agree to a more far-reaching cooperation (between $C^N$ and $C^{law}$), because any defection will immediately be discovered and punished. For the injurer punishment results in a utility amounting to the difference between the $H&R$ gain and the C&C punishment, which is negative everywhere below the cut-off point $C^{law}$. Hence, third-party enforcement can yield a higher mutual utility than the self-enforcement mechanism ($U^{law}$ instead of $U^{SE}$).

Now consider Chart 2.2:

---

67 Other kinds of enforcement than the grim trigger and external punishment are not considered here. They are just variants or combinations of the two mechanisms explained.

68 Think of line C&C in Chart 2.1 as a liquidated damages clause stipulating that the victim is entitled to a flat-rate amount of $X$ units of money whenever the other party reneges from its contract obligations. Alternatively, suppose a judge applies a law protecting the victim party and is assisted by a policeman that gives effect to the verdict. The shadow of the law effectively stretches from $C^N$ to $C^{law}$, i.e. over the entire contracting space. Beyond $C^{law}$, however, no contract is probable, since the gains of defection are more attractive than taking the losses of legal punishment.

69 However this does not mean that the shadow of the law is always able to safeguard an optimal mutual cooperation $C^{opt}$, possibly because the law cannot enforce every little detail of the contract.
Chart 2.2 The importance of enforcement capacity and enforceability in contracts

Source: Keck and Schropp (2007, Figure 2)
Notes: The general set-up of this chart is analogous to Chart 2.1. The graph illustrates the impact of weak enforcement capabilities on players’ willingness to enter into contractual obligations. Spurious enforceability skews the H&R curve to the left. Inadequate enforcement capacity on the part of the victim skews the self-enforcement curve S-E and the C&C curve down. Insufficient enforcement (intersection of H&R’ and S-E’ curves) results in less self-enforceable contracts (lightly shaded, instead of darkly shaded area) and deficient third-party enforcement (intersection of H&R’ and C&C’ instead of H&R and C&C).

It demonstrates that the efficient enforcement range is substantially dependant on both enforcement capacity and enforceability. Under a self-enforcement regime a victim’s enforcement capacity is directly dependant on the injurer’s time-value, the victim’s enactment costs of retaliation, as well as the latter’s general ability to cause (political, economic, social, emotional) harm to the injurer. Weak enforcement factors may skew the injurer’s opportunity-cost curve downwards (shown as the move from the dotted S-E to the solid S-E’ in Chart 2.2).

---

70 The injurer’s time value, or discount factor, describes how much the injurer is interested in the current period as opposed to future periods. In the extreme case in which the injurer is only interested in today’s utility and not at all in the future, a grim strategy has no deterrent effect. Therefore, no agreement would be signed.

71 Any costs connected with enacting a sanction mitigate the victim’s power (and willingness) to retaliate. It is sometimes argued that a grim trigger strategy is costly to apply for the victim, since it is not only the defecting party that forgoes future benefits of cooperation, but also the punishing party (Dixit 1987; Downs and Rocke 1995; Klimenko et al. 2002).
Enforceability plays a role in shaping the short-term hit-and-run curve: Insufficient observability, verifiability, and quantifiability make one-time defection more attractive (the H&R curve gets skewed to the left, as illustrated in the move from the dotted line to the solid H&R' curve). For every level of cooperation, defection pays off more due to a higher defection utility.

The result of inefficient enforcement depicted in Chart 2.2 is a smaller self-enforcement range. Therefore less cooperative contracts are possible in the first place: Whereas the intersection of the dotted H&R and S-E curves produces a larger enforcement range (the darkly shaded area), the scenario with insufficient enforcement (intersection of H&R' and S-E' curves) results in much less enforceable contracts (lightly shaded area). Moving away from the self-enforcement case, enforcement capacity and enforceability are equally crucial for contracts concluded in the shadow of the law: When "court-and-copper" enforcement capacity is weak, the dotted C&C line shifts down the solid C&C' line. Imperfect detection and conviction of contractual deviation causes a leftward shift of the H&R curve. Hence, under weak third-party enforcement, the possible contract range shifts to the left from $C^{law}$ in Chart 2.1 to $C^{law'}$.

Charts 2.1 and 2.2 are meant to illustrate three important points about contractual enforcement: First, the level of commitment or cooperation that a contracting party is willing to undergo is crucially dependant on its ability to enforce the terms of the agreement. A party that rationally anticipates the difficulties of punishing defective behavior by its contracting partner is not willing to make excessive ex ante cooperation commitments – even though it ideally would want to do so.

Second, it is the most reluctant transactor, i.e. the party with the lowest enforcement capabilities that will set the terms of the deal (such as the level of mutual trade liberalization, cf. Ethier 2001b). Any mutual deal (if not coercive) is decided by the party least willing to give concessions. The most reluctant liberalizer (victim or injurer) can effectively decide on the common level of exchange. Neither will a weak victim concede to “deeper” cooperation than it can sustain, nor will its stronger contract partners give additional concessions for free.72

---

72 Another way of saying this is that a mutual deal is struck at that level of cooperation where the first contracting party hits its participation constraint (Bagwell and Staiger 2002a, chapter 4). This constraint is then binding for all parties, since any deviation from that cooperation level (even if other signatories would favor it) necessarily brings Pareto-deterioration for at least one of the signatories. If parties are uncertain about their future roles and the frequency of being victim or injurer, they must assess the probability of becoming victim or injurer ex ante.
Third, the schematic representation above shows that (self-)enforcement is a *process* much more than it is an *act* of punishment. It is a lengthy dispute activity consisting of a litigation/adjudication phase and a punishment/remediation phase. Enforcement is a function of both enforceability and enforcement capacity. Whenever either aspect is deficient, contracting parties react by scaling down their ambitions; their willingness to cooperate shrinks. Having all the enforcement capacity in the world may not help the victim much if the injurer cannot be proven guilty of his defection from the letter (or the spirit) of the contract.

The fact that enforceability is an equally important ingredient of enforcement, as is enforcement capacity, has unfortunately been neglected by contract theory. Standard economic contract theory usually focuses on three issues of self-enforcement: i) repeated games of strategy; ii) soft-law enforcement in relational contracts; iii) non-verifiability.\(^\text{73}\)

Thereby, the self-enforcement literature usually conveniently presupposes a complete or at least comprehensive contract,\(^\text{74}\) one where every signatory knows what constitutes right and what is wrong behavior, and is able to immediately observe and punish defection. Difficulties in observing defection, the presence of a violation investigation, and issues of litigation are notably neglected.\(^\text{75}\) Those models therefore reduce contract enforcement to punishment and remediation – the *H&R*-curve in Charts 2.1 and 2.2 is

---

\(^{73}\) In a nutshell, models of self-enforcement assess situations, with no central enforcer to solve a dispute and execute the judgment. Since repeated interaction offers interesting sanction possibilities, one strand of the literature concerning *repeated games of strategy* examines endogenous cooperation, folk theorems and repeated strategic (non-cooperative) games. Game theorists and economists have put effort into finding deciding factors (such as a player’s reputation) and punishment strategies that can best sustain cooperation between equally powerful and asymmetrical players. Hvid (1999), and Downs and Rocke (1995) offer solid literature surveys. Another branch of literature studies *soft-law issues of punishment in relational contracts* (long-term, repeated-interaction, incomplete agreements). The focus is on *alternative enforcement instruments*: Taking human rights and environmental agreements as an example, a number of scholars contend that compliance in relational contracts is best promoted through more harmonious and less confrontational processes that rely on transparency and accessibility of information, increased reporting, monitoring, implementation review procedures, outside involvement and capacity building. This notion of “managing” the contractual compliance process relies on the argument that behavior is more easily affected through suasion and a “normative pull”, a culture of compliance, than through coercion. In order to counter excessive and unfriendly litigation, greater use of mediation, conciliation and arbitration techniques is proposed (Yarbrough and Yarbrough 1987; Guzman 2002b; Chayes and Chayes 1993a, 1993b, 1990; Holmes et al. 2003).

\(^{74}\) On the distinction between complete and comprehensive contracts, cf. section 3.2.1 below.

\(^{75}\) This is due to the rigid set of assumptions that usually underlie the models of cooperation, namely a stationary environment and zero transaction costs. Economic theories of contracts largely single out one, and ignore (i.e. model away) all other real-life imperfections that may occur during the contracting phase (i.e. before the
assumed to be fixed. Traditional WTO scholarship is no exception to this standard contract theory.\textsuperscript{76} The omission of dealing with issues of enforceability is consequential for questions of contractual incompleteness in general, as we will show in the next chapter.

\subsection*{2.1.4 Concluding remarks on the definition of contracts}

To conclude this section on the nature of contracts, four important observations can be derived from the above definition of a contract:

- Firstly, contracts differ from simple coordination activities, since contracts solve a problem. Whereas in some coordination games parties can easily join forces so as to reap mutual transaction efficiencies, or synergies, contracting parties have partially conflicting interests, namely those of cheating the other party for their own benefit.\textsuperscript{77} Contracts are thus largely concluded with the objective of hedging against misdemeanor, and to forestall (or limit) the opportunistic behavior of signatories.

- Secondly, commitment is valuable (Hviid 1999, p. 48). The \textit{ex ante} effect of contracts is often neglected in the literature, but is of immense significance: Commitment effectuates a present transfer – not of goods and services, but of rights and duties. Commitment can be assumed to have a value over and above the pure spot contract- or transaction value, and is hence socially useful. A transfer of rights and duties produces efficiency gains of its own right, independently of whether it is efficient to actually carry out the promise \textit{ex post} or not (Craswell 1999, pp.18). The more

\footnotesize

contract is signed) and/or during the performance phase of the trade agreement. This renders unnecessary any treatment of disputes, enforceability and the need for courts.

\textsuperscript{76} As we will show in more detail in subsection 4.1.2 \textit{infra}, externality-based theories of trade agreements (terms-of-trade school and political externalities school) presuppose a complete contract, and reduce WTO enforcement to remediation; the presence of a violation investigation and issues of litigation are notably neglected. This is due to the rigid set of assumptions that underlie the externality-based models of trade cooperation (see Keck and Schropp 2007, section B): In their basic form these models largely assume away those imperfections that may occur during the contracting phase and during the performance phase of the trade agreement. That way, these models can afford to focus on enforcement capacity (the shadow of the grim trigger), while disregarding issues of enforceability – verifiability, observability and quantifiability. Therefore, albeit helpful for carving out a rationale for trade cooperation, those models of trade agreements suppress all real-life problems occurring during and after the conclusion of a multilateral contract (cf. WTO 2007 section II.C.1).

\textsuperscript{77} A commonly utilized distinction in the structure of strategic games is that between collaboration problems and coordination problems (Martin 1993; Stein 1983). Briefly, in collaboration games (such as the prisoners’ dilemma), parties have partially conflicting interests, namely those of acting opportunistically by cheating the other party. Opportunism is not a problem in the simplest versions of coordination games (such as the “battle of sexes” game). Coordinating parties join forces so as to reap mutual transaction efficiencies or synergies, and no party gains from defecting. We will have more to say on the collaboration/coordination distinction in subsection 2.1.3 \textit{infra}. 

- 40 -
effectively opportunism is contained and transaction efficiency reaped, the bigger is the exchanged “good”, i.e. the higher is each party’s ex ante willingness to cooperate and to engage in an exchange in the first place.\textsuperscript{78} The scale and scope of each party’s ex ante commitments – i.e. the size of the promise – are thus a direct function of the other party’s inclination and possibilities to defect, which is equivalent to saying that it is a function of the enforceability and enforcement capacity aspects surrounding the deal.

• Thirdly, issues of assurance and enforcement are responsible for a good deal of the complexity and bargaining costs of contracts: Whereas the motivation for cooperating is usually quite trivial (building a house, performing a series of duties as a manager, conducting R&D activities), it is much more difficult and time-consuming to craft contractual obligations so as to curb potential opportunism and to maximize mutual enforceability/enforcement capacity. Depending on the special situation at hand, a contract may be shaped more by assurance or by motivational issues: When the risk of opportunism is small (such as in the contract: “I will buy a bushel of wheat from my cousin John Doe for 3$ tomorrow”), the assurance aspect of the contract is less salient. Increased attention can be focused on the substance of the contractual exchange (here: the quality of the wheat, or specifics of the delivery). Those kinds of contracts then resemble coordination- rather than collaboration accords.

In other contractual relations, however, the underlying transaction is trivial, but the assurance aspect is central. Correspondingly, enforcement provisions assume a prominent role in the contract. Examples are incentive alignment contracts that economic principal-agent theory frequently deals with. Pure incentive-alignment contracts are nearly exclusively concerned with specifying contracting goals of the agreement, i.e. establishing the rules of the game so as to achieve maximum mutual assurance for the contracting parties.

• Fourthly, the logic of joint welfare maximization of contracting parties is at the core of every contract. All rational choice contracts obey a strict Paretian logic (but see footnote 51 above). A contract is necessarily a Pareto-superior transaction (Krauss

\textsuperscript{78} For example, if I know that the defection risk of my contracting partner is close to zero, I will be willing to trade more bushels of wheat at an agreed price and a specified point in time. I might even add bushels of oats to the deal. However, if I must assume that my contracting partner will probably defect at my expense, I will be hesitant to trade my entire harvest with him. The lower the risk of being cheated, the larger will be the size of the ex ante commitment and the more valuable the promise.
2 Complete contracts

1999, p. 783), in which each signatory maximizes his/her expected value of cooperation, given the present contextual constraints. Rational actors can safely be assumed to enter into a contractual relationship only, if doing so results in an improvement of their expected short-, medium- or long-term utility. Since all sides of the deal are rational utility maximizers, it is in the interest of all contracting parties to maximize the joint gains from trade, that is, to enable them to attain the “Pareto frontier” (Schwartz and Sykes 2002b, 1996; Scott 1987). A contract on the Pareto frontier has the property that no alternative contract can increase the utility for one signatory without decreasing any other contracting party’s utility.

2.2 Basics of contracting: Creating rules

Every contract has four key stages: contract formation, contract design (“rule-making” or “contracting”), contract performance, and contract enforcement (cf. Masten 1999). It is in the contract design phase that signatories determine how their substantive goals be achieved in the most effective and efficient way – they concoct the best possible governance structure.

Pursuant to the seminal 1972 article by Calabresi and Melamed, scholars realized that contracting parties are “making their own law” by assigning themselves residual ownership rights, or “entitlements”: Signatories mold mutual commitments into a set of rights and obligations that capture – as unambiguously as possible – the nature, extent, and limits of the agreed-upon entitlements. Therefore, contracting parties generally undertake three distinct phases of contract design in which they (i) define the entitlements-to-be, (ii) define entitlement protection rules and remedies, and (iii) lay down enforcement rules in case of contractual defection by one party or the other (Calabresi and Melamed 1972; Pauwelyn 2006; Trachtman 2006).

79 Contract formation is the stage at which two or more parties convene to assess their motivation, scope for cooperation, level of ambition and the objective to be pursued by the mutual interaction. Contractors lay out their basic contractual intentions by loosely formalizing their substantive goals.

80 The contract performance phase is the stage at which the agreed-to transaction occurs, and the commitments are effectuated/converted into the promised transactions. Alternatively, it is the phase where intra- or extra-contractual defection occurs. Illegitimate behavior may then trigger the enforcement phase, which can include dispute settlement/litigation and punishment as sub-stages.
2.2.1 Primary rules of contracting – exchange of entitlements

Contracting parties first define the “primary rules” of their contractual exchange. Signatories assign residual rights of ownership, or “entitlements” to each other.\(^{81}\) Entitlements are the essence of a contract and are formulated in the form of one or a bundle of mutual rights and obligations. Some issues around contractual entitlements are especially noteworthy:

**Substantive vs. auxiliary entitlements.** Substantive entitlements qualify the contractual intent, or the cooperative gains from trade pursued by the contract.\(^{82}\) Substantive entitlements circumscribe the level of cooperation which each party is willing to concede to.\(^{83}\)

Auxiliary entitlements specify ways in which the substantive goals can be ultimately achieved in the most efficient and effective manner (Macaulay 1985). Auxiliary entitlements are non-substantive rules in that they do not follow any independent contractual intent. They safeguard, facilitate and ameliorate the contractual interaction, such as procedural requirements and timelines, codes of conduct (anti-circumvention rules, transparency guidelines, reporting obligations, rules of voting, exit, or entry to the contract), exceptions or amendments to substantive entitlements, and other organizational details.\(^{84}\) Auxiliary entitlements are geared towards increasing transaction efficiencies, reducing transaction costs, and improving observability, verifiability and quantifiability.

**Bilateral vs. multilateral entitlements.** Not all entitlements concern all signatories. In a contract with many signatories, some entitlements are owed reciprocally (bilaterally) while others are owed to the entire membership, or *erga omnes partes*. For reciprocal entitlements the rights of one signatory constitute the obligations of the other. A multilateral entitlement charges all of the contracting parties with an obligation to

\(^{81}\) Residual ownership- or property rights are referred to by L&E and economic scholars as the “individual’s ability to directly consume the services of an asset, or to consume it indirectly through exchange” (Barzel 1997 at p. 3).

\(^{82}\) Substantive entitlements define the *substance* of the transaction (prices, quantities, dates, and other transaction details, cf. Masten 1999, p. 33). Take a simple fixed-price contract for example: Both parties agree to grant the buyer an entitlement to own a product and the seller an entitlement to a certain amount of money.

\(^{83}\) As pointed out before (cf. footnote 59), commitment is usually not a binary issue, but one of degrees. Take the example of an employment contract: Although the employee obliges himself to work to the benefit of the employer, he does not wish to put all his time and effort into the contract. Also, he will promise his cooperation in some fields, but not in those, say, that involve illegal or immoral conduct.

\(^{84}\) In the international context, such organizational entitlement would be the establishment of an international organization that supports the contract and lends credence to the parties’ contractual endeavors.
behave in a certain way, whereby the entitlement is not exchanged bilaterally, but owned by the entire membership.

Dependant vs. general auxiliary entitlements. Some auxiliary entitlements have a limited application scope; they are second-order obligations which are “pegged” to or dependant on a single substantive entitlement (we shall call this type “dependant auxiliary entitlements”). Other auxiliary entitlements apply to multiple substantive entitlements (“basic auxiliary entitlements”). More complex settings are conceivable in which there are auxiliary entitlements of tertiary and lower order (e.g. an exception to a procedural rule).

Level of detail. Both substantive and auxiliary entitlements can come in various levels of detail. For example, parties may grant entitlements to pollute, or to be free from pollution. They may, however, craft more meticulous covenants which define under what circumstances (and with what magnitude and effect) pollution shall be allowed (Pauwelyn 2006, p. 7). In this case the entitlement “environmental pollution”, owned by one party or the other, is broken down into more sensible portions. Thereby, the ownership to the entitlement may change.85

Balance of substantive and auxiliary entitlements. It is probably fair to say that the more complete a contract, the more and finer grained substantive entitlements it provides. Also, the ratio of substantive to auxiliary entitlements can be expected to be higher, the more complete a contract proves to be. Incomplete contracts may often consist of a set of very basic entitlements and many auxiliary entitlements that lend additional structure to the contractual exchange. The less detail with which an entitlement is formulated, the less complete the contract can be assumed to be (more on that infra).

Multiple entitlements. Few, if any, contracts assign just one entitlement per signatory. Indeed, perhaps all real-life contracts encompass multiple entitlements (implicitly or explicitly), even if they only exchange a single set of substantive entitlements.86 More complicated contracts affect more than one specific entitlement, and hence comprise a

---

85 A more fine-tuned contract may read: “If polluter Y stays under the legal minimum of 50 mg/m³ then he has the entitlement to pollute. If Y pollutes between 50 and 100 mg/m³ then he may do so, but must pay compensation to X for the damage caused. If Y wishes to exceed 100 mg/m³, he must buy this entitlement from X. Any amount exceeding 250 mg/m³ is prohibited.” Thus, the entitlement to pollution switches from the polluter Y to the affected party X with increasing intensity of contamination.
2 Complete contracts

A multitude of issue areas, tasks, and responsibilities.\(^{87}\) In fact, the best way to understand any contract is to view it as a *bundle* of rights and obligations mutually consented to by all contracting parties. As we will see later on, it is important for the researcher examining a certain contract to detangle this bundle of entitlements, and to split up the contract into its constituent parts.

### 2.2.2 Secondary rules of contracting – entitlement protection

The second concern of contracting parties (“secondary rules”) is to allocate *residual decision rights* that organize or prohibit *ex post* behavior during the contract performance stage. In laying down the secondary rules, parties define how each entitlement should be protected, and agree upon the adequate *intra*-contractual remedy if entitlements are allowed to be transferred *ex post*. Residual decision rights are different from initial entitlements, since they lay down how, and how strongly, the initial entitlement choice is to be protected from *ex post* discretion in the course of the contractual relationship. In short, the secondary rules are rules of flexibility, escape, and remedies.

Three generic types of entitlement protection are noted in the literature: inalienability, liability and property rules of protection.

If entitlement transfer (to take, sell, or trade residual rights) is considered inefficient or immoral, a rule of *inalienability* (also termed “mandatory specific performance”) is agreed upon. The protection of the initial entitlement allocation is then called absolute, and a taking of the entitlement is strictly prohibited.\(^{88}\) Rules of inalienability mandate unconditional specific performance of the contractual rules.

---

\(^{86}\) A single-issue contract touches upon one entitlement, and leaves all the other entitlements of each transactor unaffected.

\(^{87}\) Think of a simple employment contract between a shop-clerk and an owner. Principal and agent exchange much more than the promises “work effort” for “money”. Issues such as perquisites, social insurance, leisure, health benefits, continuing education, responsibility, job rotation, etc. are all additional promises on the employer’s side. The employee, on the other hand, makes commitments over and above his job description, such as filling in for colleagues, working overtime, handling goods with accuracy and diligence, embracing responsibility, being honest, training new colleagues, planning the yearly Christmas party, etc. In short, many more rights and obligations are of relevance in a real-life contractual relationship than one might think, many of which are not even put down in writing. The point is that there are more than a single pair of entitlements (work effort and salary) affected in a seemingly simple contract.

\(^{88}\) Political rights are an example of inalienable rights that cannot be traded: It is not legally possible to sell off your vote in a ballot.
If contracting parties generally consent to the possibility of trading or reallocating entitlements *ex post*, they determine whether to protect initial entitlements by means of a *liability rule* (LR) or a *property rule* (PR) of entitlement protection (or a mix thereof). Under a pure LR, one party (the taker or injurer\(^89\): he) has the option to take away parts of the other party’s (the owner or victim: she) entitlement unilaterally, i.e. without the owner’s prior assent. The taker can engage in unilateral appropriation (which is an expropriation of the holder of the entitlement) under the condition that he compensates the owner for damages suffered in some fashion or other – usually by paying a previously specified exercise price.

Under a property rule of protection both parties are under a “specific performance duty”, i.e. a strict obligation to respect the initial entitlement distribution. In this case, the taker is directed to perform, and a failure to do so is punished so severely that he would never prefer violating the order to complying with his obligation. However, a potential taker can buy off the owner’s entitlement through *renegotiations*. He can still avoid his commitments by securing permission from the owner, usually by paying for it.\(^90\) Whenever the parties come to an agreement, the owner cedes her entitlement and sells it to the taker – the transfer is thus *bilateral*.

A rule of entitlement protection must in all cases be accompanied by a corresponding remedy rule. Parties thereby agree on how “costly” *intra*-contractual non-performance by the injurer should be. There is a continuum of remedies ranging from a zero-damage payment to coercive (infinitely high) damages (Mahoney 1999). In between lie the most important damage remedies: The *restitution* remedy re-establishes the *status quo ante* the *contract*. The *reliance* measure obliges the injurer to re-establish the victim’s *status-quo ante* the “breach”.\(^91\) *Expectation damages* place the victim in as good a position as it would have been had the injurer performed. The expectation remedy represents the

\(^{89}\) As was mentioned in Chapter 1, we define “victim” and “injurer” as *roles*, not as *conditions*. This is to say that any contracting party experiencing some form of contractual regret or doubt automatically assumes the role of the “injurer” independently of whether it acts on that regret (e.g. by exercising a liability-type opt-out, engaging in renegotiations, or violating the respective agreement) or not. Equivalently, the role of the “victim” is that of the actor affected by the other party’s regret or doubt. Both terms are non-judgmental and are free of welfare implications for the players (a victim is *not* necessarily worse off by assuming that role).

\(^{90}\) The bargaining space under a PR thereby leads to efficient *ex post* non-performance: “Since the promisor [injurer] will pay no more than the value to it of the breach, and the promisee [victim] will accept no less than the value of the harm [she] expects to suffer from the breach, an agreement that permits the promisor to breach can be reached only when the benefit to the injurer of the breach exceeds the harm to the victim resulting from the breach, that is, when breach is efficient” (Schwartz and Sykes 2002a, p. 5).
replacement value of the deal and acts as a complete insurance policy to the victim: *Ex post*, expectation damages make the victim indifferent between the injurer’s performance and default. Next, *efficiency* damages are expectation damages plus the efficiency gains from non-performance (should there be any). *Negotiated* remedies can in principle lie anywhere between zero and coercive remedies. Their size, however, crucially depends on the underlying rule of entitlement protection.\(^{92}\) Not all combinations of escape and remedy rules make sense: Inalienability rules are best adhered to by coercive penalties (prison, forfeit, liquidated damages, etc.). Property-rule protection is logically accompanied by bilaterally negotiated remedies. Liability-rule protection is usually accompanied by expectation damages, a combination sometimes – unfortunately – referred to as “efficient breach”.\(^{93}\)

Technically, secondary rules of contracting are formulated as *dependant* auxiliary entitlements, i.e. as norms that are pegged to a specific substantive or auxiliary entitlement. A single basic entitlement can be protected by various rules of entitlement protection, whereby different protection rules come into effect under different circumstances, contingencies, or actions (see the example in footnote 85 above).

As will be shown in the next chapter, signatories’ choice for entitlement protection becomes an interesting topic in incomplete contracts, that is, in those situations where the original contract contains gaps, and where *ex post* non-performance can be welfare-enhancing to both (or all) signatories to the contract. The more difficult the initial contracting context (number of traded entitlements, level of entitlement refinement, number of signatories, etc.), and the more dynamic the environment surrounding a

---

\(^{91}\) Reliance is presumably a larger amount for the victim than the restitution rule of damage, since prior transaction efficiencies have to be accounted for. The longer the contractual relationship has been going on, the more restitution and reliance measures differ.

\(^{92}\) Under an LR of entitlement protection that allows the injurer to opt out whenever he desires, negotiated remedies can potentially be close to zero. Under a PR, however, negotiated damages are strictly in between the expectation and the efficiency measures. The rational victim will not settle at a loss (and expectation damages put her in a position where she is exactly indifferent between the injurer’s performance and his non-performance), and ideally would want to capture all the injurer’s efficiency gains from non-performance for its willingness to “let go” of its contractual rights.

\(^{93}\) It is evident by now that the nomenclature “efficient breach” for an LR-cum-expectation damages arrangement of protection is doubly misleading: First, as mentioned, if a contract specifically permits non-performance to occur, it is imprecise to speak of a “breach” of rules. Second, the term does not resolve the more vital question of how to organize non-performance efficiently – as a liability rule or a property rule. Efficient adjustment to changing circumstances and unforeseen contingencies is notably possible under both LR and PR: “To be sure, it is possible for efficient breach to occur with other remedial options [than the LR]. In particular, if specific performance is the remedy for breach a party wishing to breach can always approach the other party and attempt to negotiate a release of performance”, as Sykes (2000, p. 353) points out.

- 47 -
contract, the more important – and difficult – the choice of entitlement protection rules proves to be.

2.2.3 Tertiary rules of contracting – enforcement of entitlements

After having delineated intra-contractual, permissible behavior, transactors then have to decide how to sanction extra-contractual uncooperative behavior should it nevertheless occur. These tertiary rules concern what can be called backup entitlement protection – enforcement mechanisms and procedures taken in response to the illegitimate taking or destruction of an entitlement. Hereby, the parties consent to how the previously agreed-upon assignment and the rules for transfer of entitlements can or should be protected against unilateral defection. The task for the contracting parties is to make sure that a protection of rights is enforceable, and that punishment is not an empty threat. Enforcement capacity and enforceability clearly play a large role: Depending on how easily the contractual rules can be thwarted through opportunistic behavior, parties are more or less willing to cooperate to the full extent. If the original primary commitments cannot be enforced, parties will scale down their original level of cooperation to a level that can be sustained by the enforcement tools at hand. As Pauwelyn (2006) points out, the question of contract enforcement is especially relevant in the international realm, where usually no supra-national enforcer exists and abidance with the rules of the game cannot be taken for granted.

In summary: The outflow of the contract design phase is a set of enforceable mutual commitments in the form of a bundle of traded entitlements and entitlement protection rules. Each party knows its rights and its obligations and has given its consent to sanctions for deviating behavior. The contract design phase is the most important contracting stage, since the quality of the negotiated governance structure sets the level of ambition ex ante, and determines signatories’ behavioral patterns in the subsequent phases, namely the contract performance and enforcement phases: All sorts of ex post problems and challenges have to be anticipated and the governance structure adapted accordingly. This implies that the originally intended level of cooperation may have to

---

94 Eventually, in the course of the contract design phase one or more contracting parties may scale back their initial level of ambition. This is precisely the application of some of the last subsection’s insights: If the governance structure is faulty (due to, say, imprecise or ambiguous language), or if one party is aware that it cannot safeguard a certain level of enforcement, it will have to reduce the level of cooperation it is willing to muster.

95 Commitments are frequently implicit and not written down. Nevertheless they have been given the consent of both parties.
be modified depending on how feasible or risky the attainment of the transaction goals are in light of the difficulties in later stages of contracting.

Depending on the contractual context (number of players, complexity of the exchange, longevity, market imperfections, transaction costs, information asymmetries, etc.), parties regularly need to engage in trade-offs between what is desirable and what is feasible. In particular, parties have to anticipate and adapt to environmental eventualities (“contingencies”) and ensuing opportunistic party actions – and how such misdemeanor shall/can be sanctioned. As we will show \textit{infra} in Chapter 3.2, contractual design efforts geared towards preventing opportunistic behavior can take place either by explicit wording or by contractual default rules. Since rule-making is a costly enterprise, rational parties have to engage in another trade-off: between the costs of writing contractual contingencies and the expected gain generated therefrom (Battigalli and Maggi 2002; Horn, et al. 2006; Horn et al. 2005).

2.2.4 Mixed regimes of entitlement and entitlement protection

An important aspect of contracting is that of mixed regimes. We have so far stated that contract entitlements “belong” to one or the other contracting party (the injurer or the victim), and that entitlement protection is organized as a liability, property, or inalienability rule. We did so for expositional convenience only. In real life we often see mixed entitlements, and systems of mixed entitlement protection.

2.2.4.1 Multi-entitlement contracts and divided entitlements

Real-life contracts oftentimes feature regulations of entitlements divided between contracting parties (so-called “co-ownership”, cf. Kaplow and Shavell 1996b, p. 749): Entitlements are not fully owned by one transactor or the other, but are in fact shared. Think of the aforementioned example of environmental pollution: Victim $Y$ may think she has an entitlement to be “free from pollution”. In reality, however, firms are allowed to pollute the environment within prescribed limits. Otherwise they would not pursue their production activity. The prescribed pollution limits, or technological requirements, are expected to balance the harmfulness of pollution against the costs of its prevention.

What Kaplow and Shavell refer to as co-ownership or divided entitlement is in fact little more than a more detailed (and more accurate) definition of mutual rights and obligations. To fix the mutual balance of rights and obligations, and to carve out entitlements more finely, clear and unambiguous entitlement language has to be agreed
on by contracting parties. The finer the determination of mutual entitlements, the more detailed, nuanced, and accurate the entitlement protection rules have to be.\footnote{As we stated above (footnote 85), finer-grained entitlement language may bring more nuanced ownership rights.}

As ready alternatives to explicit entitlement definitions, contracts often provide for conditionality clauses, consisting of enactment thresholds and limitations of scope. The level of conditionality is often determined in the form of a set of prerequisites that have to be fulfilled before the contracting parties may engage in \textit{ex post} discretion.\footnote{A pertinent example of a hybrid entitlement protection somewhere in between a simple liability rule and a property rule is Art. XIX of the GATT. Under the safeguards clause a potential injurer is faced with a substantial threshold of application: For a safeguard measure to be imposed, an enacting country must show that “i) as a result of unforeseen development; ii) imports in increased quantities; iii) have caused or threatened to cause; iv) serious injury to the domestic industry producing the v) like product” (see Howse and MAVROIDIS 2003 at p. 686; Roitinger 2004 at p. 102).} This high level of prerequisites bestows significant negotiation power, and dispute sway, on potential victims. Alternatively (or additionally) to a high enactment threshold, the scope of application of \textit{ex post} discretion may be restricted: A measure may only be enacted once, only on Fridays, only for a limited amount of time, or only vis-à-vis a certain subset of signatories, etc.\footnote{High levels of prerequisites and restrictions on application scope may, however, entail non-trivial problems, as will be detailed below in the discussion on contingency measures of flexibility (Chapter 3.2.2.1).}

\section*{2.2.4.2 Rules of divided entitlement protection}

One-entitlement contracts are a rare animal. Realistically, multiple entitlements are traded in real-life contracts. Finding one-size-fits-all solutions to protect all traded entitlements with the same entitlement protection rule is highly unlikely. Just as the nature, context, and level of detail of entitlements vary, so will the appropriate mechanisms of protection. Signatories are hence better off protecting each exchanged entitlement by a (set of) unique rule(s) of entitlement protection. Therefore, a mixed system of entitlement protection seems the logical consequence of multi-entitlement contracts. Thus, we can expect bundles of entitlements that together constitute a contract to be protected by a mixture of the possible default rules, namely alienability-, property-, or liability rule.

Closely connected to this point and the previous issue of co-ownership of entitlements is another important aspect of real-life contracting: divided entitlement protection. Many entitlements are not protected against \textit{intra}-contractual non-performance by one simple liability rule whereby the injurer is permitted to cause harm anytime, provided he
compensates the victim for the damage done (or a court’s best estimate of it). Kaplow and Shavell (1996b), and Ayres and Talley (1995b) have shown that liability and inalienability are two points along a spectrum. In fact, inalienability can be understood as a special case of liability that uses fixed *ex ante* estimates and prohibitive damages. A property rule, in turn, is nothing but a special case of an LR with varying damage awards. Consequently, there is a considerable grey area of divided entitlement protection. To see this consider Chart 2.3 below.

**Chart 2.3 Divided entitlement protection – points along a continuum**

![Diagram of divided entitlement protection](chart23.png)

Source: author

Notes: The horizontal axis represents the power of *ex post* discretion, which can lie with the victim, or the injurer, or with neither of them. *PR* stands for “property rule”, which can be granted to the injurer (*PR_{inj}*); or to the victim (*PR_{vic}*); *LR* stands for “liability rule”; *IR* stands for “inalienability rule”. The vertical axis maps the size of remedies, which can range from zero to infinitely high. The curvature of the graphs is of no importance, and as such three variants are shown.

The horizontal axis depicts which party enjoys the power of *ex post* discretion, or in other words, who “owns” a contractual gap (not: who owns the entitlement). Post-contractual discretion can thereby reside with the injurer, with the victim, with both players (divided entitlement protection), or with none (in case of inalienability). The vertical axis represents the size of contractual remedies payable for the exercise of *ex post* non-performance, ranging from zero to infinitely high amounts. The curvature and progression of the remedy graph is inconsequential here, because signatories usually

---

99 “[W]ere we to allow damages to be any quantum, then ‘liability’ rules and property rules would no longer be distinct: […] The conventional liability rule that we emphasized is the rule with damages equal to courts’ best estimate of harm […] a liability rule with very high damages is equivalent to property rule protection of victims,”
establish a single point on the graph in the contract, where they define a single entitlement protection rule accompanying a particular contractual entitlement. So, for exposition purposes three possible graphs are shown in the chart. It is important to see is that the two ends of the continuum are a property rule (PR) awarded to the injurer and an inalienability rule, respectively. A PR to the injurer is connected to a zero-damage rule of remedy: The victim actually has to buy off the injurer if it wishes maintain the performance of the contract. An inalienability rule is connected with infinitely high remedies for both parties. A PR to the victim is connected with negotiated remedies: The injurer has to renegotiate his non-performance, or else be obliged to perform as promised. In between the endpoints of the continuum are liability rules (LR) with different amounts of intra-contractual remedies, with the expectation damage measure featuring most prominently here, since it exactly reimburses the victim for her damage (expectation remedies are the “replacement value” of a promise). Hence, we witness contract clauses in which compensation for damages paid by the liable party is systemically short of harm (see the reliance and restitution measures in Chart 2.3). Liability damages can also exceed harm done, in order to add a deterrent effect for future violations (so-called “liquidated damages”, cf. for instance Mahoney 1999 at p. 27).

Examining mixed regimes has relevance for the conceptual analysis and for the understanding of particular contractual rules. When considering how property and liability rules are actually applied in real-life contracts, we realize that mixed regimes of entitlement and entitlement protection are much more prevalent than the conventional extremes.

To summarize this section on the basics of contracting: Whereas in the last subsection (2.1) the “level of cooperation” was treated in the abstract, this section helps to elucidate what actually constitutes the phenomenon of cooperation: A contract consists of a bundle of entitlements, or mutual rules and obligations. These various entitlements combined determine the overall level of contractual commitment, or cooperation. Ex ante commitment in an entitlement will vary with the quality of its respective system of entitlement protection and enforcement. The protection of all entitlements is called the “governance structure” of a contract. The more sophisticated the system of entitlement protection, and the more effective the enforcement measures, the higher the signatories’ overall level of commitment.

and a liability rule with damages of zero is equivalent to property rule protection of injurers” (Kaplow and Shavell 1996, p. 724, 754).
2.3 Types of contracts, and alternatives

Contracts – narrowly defined as signed pieces of paper, or broadly defined as any “enforceable commitment over time” – are a principal foundation of human interaction. Many different types of contracts exist, and almost as many ways to categorize them. For our purpose we want to distinguish contracts along two dimensions: the underlying problem to be solved by the contract, and the complexity of the agreement.

2.3.1 Collaboration vs. coordination

A contract is always created to solve a problem. Without a problem, contracts would be superfluous, and a simple cooperation accord (by handshake or nod) would suffice. Negotiation and “ink costs” would notably be saved.

Generally, in cooperation games that display the collaboration problématique (examples are the prisoners’ dilemma, or the stag hunt game), parties have partially conflicting interests, namely those of acting opportunistically by cheating on the other party. Through the conclusion of a contract, cooperating parties can overcome the problem of mutual defection and welfare-depreciating opportunism.

Opportunism is not a problem in the simplest versions of coordination games (WTO 2007, section II.B.3). The problem in these games is distribution: Consider the infamous battle-of-sexes game: Parties have to choose between two cooperative equilibria (whereby each player prefers a different cooperative outcome), but ultimately no party ever gains from defecting. Players can only reap mutual transaction efficiencies if they cooperate. In coordination games the conclusion of a contract (which is, after all, a costly enterprise) would seem superfluous.

However, coordination games are not generally free from disagreement: Whenever applied to a less clinical and more real-life setting than a 2x2-matrix, coordination over complex issues typically yields a vast amount of self-enforcing equilibria that two or more parties prefer to no agreement at all. But parties are in vivid disagreement on their

---

100 Without a problem, contracts would be superfluous, and a simple cooperation accord (by handshake or nod) would suffice. Negotiation and “ink costs” would notably be saved.

101 See footnote 77 above. There are more nuanced strategic game characterizations (Aggarwal and Dupont 2004, 1999; Ostrom 2003; Sandler 1992), which are of little added value here.

102 The stag hunt is a game which describes a conflict between safety and social cooperation. The game lends its name to a situation described by Jean-Jacques Rousseau in which two riflemen go out on a hunt. Players simultaneously choose whether to hunt a stag or a hare. If an individual hunts down a stag, she must have the cooperation of her partner in order to succeed. An individual can get a hare by herself, but a hare is worth less than a stag. Formally, a stag hunt is a game with two pure strategy Nash equilibria – one being risk dominant, another being payoff dominant (see e.g. Skyrms 2003).
subjective rankings of the mutually preferable agreement candidates. Different equilibria are thereby favored by different players. Parties must choose collectively one of various welfare-superior equilibria. This creates tensions and disagreement among the players. Consequently, transactors tend to argue about the distribution of gains from cooperation.103 A contract can thus help signatories fix the negotiated distribution of expected gains in a way that is acceptable to all players, and verifiable by external third parties, such as a court.

It is often argued, that with regard to surveillance and enforcement, collaboration regimes are typically more formalized and institutionalized than coordination regimes. Fearon (1998, pp. 270, 275-276), however, points out that the collaboration-coordination dichotomy is misleading, because most contractual problems in existence have a common – mixed – strategic structure.104 Fearon contends that most cooperative settings first involve a problem of bargaining about the distribution of future cooperation gains. This bargaining is akin to a coordination game. In a second stage, games of cooperation involve a problem of enforcement and monitoring, which is akin to a prisoners’ dilemma (PD) game of collaboration. Fearon concludes that approaches which treat a PD as the key problem in contractual cooperation tend to under-estimate ex ante bargaining. Bargaining problems are often more important obstacles to cooperation (especially cooperation between sovereign States) than monitoring and enforcement.105 The lesson here for our purpose at hand is that most real-life contracts solve problems of coordination and collaboration at the same time (see also Powell 1994).

---

103 Adding realism by considering dynamism, uncertainty, and asymmetrical information about other parties’ “bottom line” makes it clear that virtually all contexts of international cooperation involve such distributional conflict concerning the terms of cooperation (Fearon 1998).

104 Fearon’s contribution underlines a point made by Snidal that “almost all international cooperation problems mix efficiency and distribution concerns” (Snidal 1997 at p. 485; see also Morrow 1994).

105 While “a long shadow of the future may make enforcing an international agreement easier, it can also give States an incentive to bargain harder, delaying agreement in hopes of getting a better deal” (Fearon 1998, p. 270).
2.3.2 Complexity of contracts, and alternatives to contracting

Real-world contracts can be placed on a “complexity” continuum. Contractual complexity can be characterized by the following dimensions: longevity of the contract, frequency of exchange, dynamism of the environment, contractual incompleteness, number and heterogeneity of players, number of cooperative goals (or “issue areas”), and the “depth” of commitments.

Along the complexity dimension we can now rank contracts from simple quasi-barters of a single good (one-off transactions displaying high contractual explicitness and completeness, as well as high environmental stability), to more complicated multi-issue contracts (like drilling a tunnel, employment contracts, R&D orders), to repeated-interaction relational contracts, to constitutions, and international organizations (displaying a multitude of heterogeneous signatories, longevity, repeated transactions, a high level of incompleteness, and significant environmental volatility). Differences in complexity notwithstanding, common to all types is the contractarian logic outlined above, namely the need for substantive and contracting goals, for assigning entitlements and entitlement protection, the Paretian principle, and the four phases of contracting.

Inter-subjective cooperation is the essence of peaceful human interaction. Formal contracts are very frequent nowadays, and continuously gaining in significance. They are, however, merely a social structuring device by which individuals or groups regulate their interaction. Alternatives to (explicit, written) contracts include spontaneous barters (spot-transactions), reputation-based covenants (Masten 1999, p. 26), and integration into a firm, family, or group.

---

106 Contractual complexity has not been defined conclusively in the literature. We take the liberty to characterize it as noted.

107 Dunoff and Trachtman (1999, p. 17) contend: “Between spot market transaction and the formal organization there exist many types of formal contracts and informal arrangements, and even the formal organization is a nexus of contracts. Thus, the supposed dichotomy [between transactions and institutions] is, in fact, a continuum: the boundary between the transaction and the institution is blurred.” International organizations are institutions founded through treaties, and therefore logically are to be understood as contracts of sorts. Following the “New Institutional Economics” strand of literature (Coase 1937; North 1991, 1990, 2005; Williamson 2000, 1979; Menard 2004), we contend that an institution is a contract, an equilibrium to a game of strategic interaction (North 1990).

108 There is a large literature in the field of industrial organization dealing with the boundaries of the firm and with the question of when it makes sense to integrate processes within the firm instead of “contracting out” at arm’s length (Grossman and Hart 1986 and references therein).
2.4 The contracting ideal: Pareto-efficient complete contingent contract

The quest for the very contract terms that yield optimal outcomes is the subject of a prodigious theoretical literature in economics. The customary starting point is the concept of the “complete contingent claims contract” associated with the work of Arrow and Debreu (e.g. Hart and Holmström 1987). In a seminal contribution to contract theory, Shavell (1980) coined the term “Pareto-efficient complete contingent contract” (hitherto CCC) as the first-best benchmark for every contract – irrespective of contracted content, number of contracting parties, or situational context at hand.

A CCC is the Arrow-Debreu ideal of a contract that completely informed, perfectly rational parties would write in absence of any contracting imperfection (such as negotiation costs, costs of information gathering, or bounded rationality, see Shavell 1980, pp. 466) and in the presence of optimal enforcement capacity. It is an imaginary, hypothetical contract that provides for a complete description of every possible present and future state of the world – no matter how small the probability of the contingency. A CCC assigns rights and ownership between parties in every situation and for every contingency, spelling out exhaustively and in complete detail the exact legal rights and duties of each party including the set of instruments that a signatory may or may not use (Cohen 1999, p. 79). The fully efficient contract thereby exhausts all possible gains from trade – it is the first-best contract between trade partners (cf. also Craswell 1999; Hart and Moore 1988; Posner 1988).

A couple of comments illustrate the nature of the CCC:

- It is the property of efficiency that makes the CCC the “archetype against which to compare all realistic agreements” (Masten 1999, p. 27). The CCC is free of market imperfections, unforeseen developments, and opportunistic behavior. It maximizes joint welfare (Shavell 1980). The CCC not only satisfies the requirements of Pareto superiority, but of Pareto efficiency, too. No other contract can do better than the CCC.

- The CCC takes full care of the assurance aspect of contracts (cf. subsection 2.1.2): Since contractual opportunism is forestalled by perfect foresight, contracting parties
are prepared to maximize their *ex ante* commitments. In other words, the exchanged up-front promises of contracting parties are of maximal size. Parties are willing to offer the optimal scale and scope of the contractual exchange that are needed to maximize the *ex post* gains from trade.

- The CCC is a *complete* contract in the sense that, no matter what happens, the CCC prescribes a course of action. There are *no degrees of freedom* and *ex post* discretion. This brings about three interrelated consequences: First, no mutually beneficial *ex post* modifications to this completely state-dependent contract can be made. A complete and conclusive set of mutual rights and obligations is in place from the beginning, and no decision at a later stage can effectuate any improvement. The contract is “renegotiation-proof” and never needs to be revised or complemented (Holmström and Tirole 1989, p. 68; Mahoney 1999, p. 119). Second, CCC-entitlements are logically protected by a rule of *inalienability* – an unconditional specific performance rule backed up by prohibitively high *extra*-contractual remedies. Indeed, since the CCC is Pareto-optimal, *every* contractual non-performance must by definition be opportunistic and welfare-depreciating. Third, to tie in with what has been said about the primary rules of entitlements in previous subsections: A CCC consists of a comprehensive set of substantive and auxiliary entitlements that assign rights of ownership to signatories in every possible state of nature (“if contingency *x* occurs, party *A* is to do *y* and party *B* is to do *z*). This fine-meshed set of entitlements is protected by a general enforcement rule of inalienability.

- *Intra*-contractual default is very well provided for in the CCC (in the form of carefully crafted primary rules of entitlement). *According to the terms* of the CCC, a party will typically be released from certain obligations under well-specified contingencies (the right of ownership shifts). For example, one party may be excused from having to sell its good if the factory burns down. Therefore, the statement that a party always obeys the terms of a CCC does not mean that the party always meets a named obligation, i.e. takes a particular action (cf. Shavell 1980 at note 2).

---

109 This goes for symmetrical parties. In the case of asymmetrical players, mutual commitment continues, until one player achieves his/her preferred (optimal) level of cooperation. Additional mutual cooperation beyond the binding participation would invariably lead to Pareto-deterioration (cf. footnote 72 above).

110 Shavell (1980, p. 467): “[A] Pareto efficient complete contingent contract is one to which the parties would find it in their mutual interest to be bound to adhere. In particular, they would wish for damages for failure to meet the terms of the contract to be set sufficiently high that the terms would always be obeyed” (emphasis in original).
3 Incomplete contracts

3 Incomplete contracting, and the essence of flexibility

The Pareto-optimal complete contingent contract is a hypothetical construct that does not correspond to any contract that has ever been or will be concluded. Intuitively, writing a CCC is neither practical nor feasible: It would not be practical, since even in a stationary world, i.e. one where environmental circumstances do not change at all, it is prohibitively costly to lay down in detail all permissible (read: joint welfare maximizing) behavior of transactors (Horn, et al. 2005).¹¹¹ Even if signatories were living in a “Coasian world”, a world void of transaction costs and with infinitely rational actors, writing a CCC would be impractical. In such a world any initial allocation leads to an efficient outcome through negotiation (which is an application of the Coase Theorem, cf. Kaplow and Shavell 1996b). Wasting resources on negotiating and writing a contract in a Coasian world must be seen as a futile task (Dunoff and Trachtman 1999, p. 23).

In real life it is not only impractical but infeasible to write a complete contract, since the future is only imperfectly foreseeable for contracting parties (Masten 1999; Shavell 1980; Craswell 1999). The world is not stationary but volatile and ever-changing. A complete contingent contract would have to consider all sorts of present and future environmental contingencies and prescribe in a detailed and unambiguous fashion the admissible welfare-maximizing behavior by all contracting parties. Every contingency, occurring even with the slightest probability at any point of time during the contract duration, would have to be considered and contractually fixed.

Thus, all contracts are necessarily incomplete. They contain gaps, that is, they are insufficiently contingent and – in a strictly technical sense – inaccurately written. This chapter is to provide for a structured and comprehensive introduction to contractual incompleteness. In particular, we will examine the reasons for, propose a novel taxonomy of, and assess the consequences of contractual incompleteness for signatories (sections 3.1.1-3.1.6). We will discuss ways of dealing efficiently with contractual incompleteness and the regret it provokes in signatories (section 3.2). Specific focus will be placed on the design of contractual flexibility mechanisms (section 3.3). The outflow

¹¹¹ Tirole (1994) cites a classic example of the impracticability of writing a complete contract: Suppose a research & development (R&D) contract between a principal (client) and an agent (researcher). A “water-tight”, complete, contract between these two parties would have to specify not only the desired outcome (say, a remedy against cancer), but also the way of achieving it – which, of course, is the whole point of an R&D contract in the first place.
of our discussion of contractual incompleteness will be the introduction of a general framework for assessing incomplete contracts (section 3.4). This will allow us to progress towards a “theory of disputes” (section 3.5).

3.1 A categorization of contractual incompleteness

If we want to operationalize the concept of incompleteness and deal with it in a structured manner, four questions need to be tackled: First, what circumstances make contracts incomplete? (subsection 3.1.1). Second, what varieties, or types, of contractual incompleteness can be distinguished, and along which logical fault lines, or attributes? (subsection 3.1.2-3). Third, do we have the technical means to cope with different types of incompleteness? In other words, are the existing approaches to incompleteness adequate and sufficient to address the complexity of real-life contracts? (subsection 3.1.4). Fourth, what are the consequences of incompleteness for signatories? (subsection 3.1.5).

There is a steadily growing incomplete-contract literature, from both within the L&E and the economics disciplines. However, many contract scholars have contented themselves with stating that contracts are somehow incomplete – without specifying what they mean by that, and what kinds of assumptions underlie this contention. Others give the statement a bit more meaning by citing “uncertainty” as the key reason for contractual incompleteness. But what exactly is implied by the term uncertainty? We can at least distinguish three conceptions in the contract literature: uncertainty over the

---

112 There is no single universally accepted definition of “incomplete contract” (Tirole 1994, p. 743). We will stick to the very rigid definition of Tirole who regards as incomplete every contract that is not a CCC: “A contract is incomplete if it does not exhaust the contracting possibilities envisioned in the complete contract” (ibid.). This definition integrates both the economists’ view of incompleteness (as insufficiently contingent contracts), and the lawyers’ view of missing language. For lawyers, incomplete usually means that the obligations of the parties are not clearly specified – that important terms like price, quantity, time of delivery and quality are not written down in the contract (see Edlin and Reichelstein 1996 at footnote 4; Ayres and Gertner 1989 at footnote 29).

113 Compare four examples of uncertainty from the GATT/WTO literature: Kovenock and Thursby (1992, p. 159) poetically explain random deviations in signatories’ protection structure as “demons” that “temporarily possess countries”. Bown (2002a, pp. 295) defines uncertainty over the future loosely as “unanticipated preference shocks” to the political-economy parameter of governments. Ethier (2001a, p. 10) gives the concept of uncertainty slightly more structure: “I am especially concerned with three types of uncertainty to which [signatories] would be subject ex ante: 1 Uncertainty about what actual policy situation (environmental issues, health or safety concerns, etc., etc.) might give rise to [a contractual escape] action. 2 Uncertainty about the identity of the country in which the situation might emerge. 3 Uncertainty about the extent to which the potential action might be trade-related” (emphases in original). Finally, Milner and Rosendorff take uncertainty to be “ignorance of the configuration of political pressure to policymakers in future periods” (Rosendorff 2005; Rosendorff and Milner 2001), or as “ambiguity regarding preferences of key domestic players” (Milner and Rosendorff 1997).
future; uncertainty over other players’ actions, agent-type, or knowledge (asymmetrical information); and uncertainty over the meaning and scope of existing contractual provisions (e.g. textual ambiguity). While this distinction is a good point of departure, there are still a number of questions unresolved. Take the concept of “uncertainty over the future”: Does this imply that players cannot fathom contingencies at all (“unforeseeability”), or that they just have not bothered to specify them (“unforeseen contingencies”)? And are we referring to uncertainty over future actions, outcomes (situations), probabilities, or the identity of the party affected by future contingencies (roles/types)?

In order to understand better what contract scholars mean by “uncertainty”, to be able to compare contract-theoretical approaches to incompleteness in the literature, and to introduce an additional scientific rigor, we will seek to structure the concepts of uncertainty and incompleteness, and show how the two are interlinked. Thereafter, we can plot the existing approaches to incompleteness into that framework.

3.1.1 What makes contracts incomplete? Transaction costs and bounded rationality

If uncertainty is the reason for the incompleteness of contracts, it makes sense to first assess the origins of uncertainty. Two explanations as to why contractors have to cope with contractual uncertainty (and hence leave gaps in their contracts) are mentioned in the literature: Rational cost-benefit consideration due to the presence of transaction costs, and bounded rationality of signatories.

3.1.1.1 Transaction costs

Transaction costs (TC) serves as a general term for all those real-life costs that signatories must incur when cooperating. TC make collective human interaction cumbersome and contracting costly. When the costs of writing down contingencies exceeds the benefits of doing so, signatories will rationally embrace uncertainty. Although much has been said about transaction cost economics (a strand of the literature linked to the names of Ronald Coase and Oliver Williamson), there is no consistent understanding of what the term really implies. TC seem to be a collective “bulk term”

---

114 In the literature, uncertainty over future contingencies is usually conceived very broadly as the existence of unanticipated political, economic, technological, or natural contingencies (such as demand shocks, technological
for anything that somehow produces imperfections: Pre-contractual sorting and searching, or post-contractual litigation-, enforcement- and policing costs, etc. We suggest a slightly more concise classification, utilizing Cohen’s (1999) distinction between ex ante/ex post, and exogenous/endogenous TC. Consistent with our usage so far, ex ante refers to the time prior to the conclusion of the agreement, while ex post refers to every point in time thereafter.

Exogenous, or “Coasian” TC are real-life impediments to inter-subjective interaction. Exogenous ex ante TC consist of the basket of (opportunity) costs incurred in the run-up to signing an agreement, such as sorting-, searching-, information gathering- and processing costs, and bargaining costs in particular. This includes are negotiation costs, costs of researching probabilities and effects of possible contingencies, legal fees, and “ink costs” (drafting and writing costs). Exogenous ex ante TC are the objective costs of market imperfection.

Exogenous ex post TC consist of renegotiation-, litigation-, policing- and enforcement costs. “Litigation costs” is thereby the catch-all term for information processing-, research-, drafting-, pledging-, court-, lawyer-, and opportunity costs in connection with conducting a litigation. Policing and enforcement costs are expenses incurred for making sure the other party sticks to the terms of the contract, and for taking appropriate action if this turns out not to be the case.

Endogenous transaction costs are original actions, or behavioral responses to exogenous contingencies by contracting parties. Endogenous TC are of major concern in contracting situations: They are subjective, or unilaterally provoked (hence: avoidable) costs. They are also known as strategic behavior. Strategic actions can occur before the contract is signed (ex ante) or during contract performance (ex post). Strategic gamesmanship can
be afforded by asymmetric, or privately revealed information. Yet it may also occur in situations of symmetrical information, such as when informed parties are engaging in so-called “hold-ups” and “hold-outs”.

The exogenous-endogenous dichotomy of transaction costs is exactly what Battigalli and Maggi (2002), and Horn, Maggi and Staiger (2006; 2005) refer to as “rigidity” and “discretion” costs of writing contracts.

3.1.1.2 Bounded rationality

An alternative, possibly complementary, explanation for the presence of contractual uncertainty is the assumption of bounded rationality. At the core of this view is the simple realization of “human contracting error”. Human error unfolds in three ways. First, contracting parties regularly fail to anticipate the existence of certain future circumstances – they lack the imagination to think about possible states of the world, be

---

117 Asymmetric information settings may give rise to ex ante costs produced by “adverse selection” (hidden knowledge), or to ex post “moral hazard” behavior (hidden action). In standard terminology, the propensity to deviate from joint-surplus maximizing behavior in the presence of asymmetric information is called “moral hazard” when the distortion involves actions or information revelation ex post, and “adverse selection” where ex ante private information leads only those transactors with less desirable characteristics to transact (Masten 1999, p. 28). An extensive review of this earlier incomplete contract literature can be found in Hart and Holmström (1987).

118 Hold-outs are a natural characteristic of dynamic bargaining problems (Rubinstein 1982 was the first to formalize the dynamic aspect of bargaining): They occur in situations where there are multiple outcomes that two (or more) parties prefer to having no agreement at all. In a sequence of offers and counteroffers one party may strategically delay the resolution of the contract in hope that the other side will make concessions. Parties thus procrastinate so as to reap a more preferred contractual option, i.e. a favorable distribution of gains from trade. Knowing that delay is costly (it results in more time spent without the benefits an agreement would bring; it also increases the risk that one side will break off negotiations entirely and look for other trading partners), the contracting party that values the agreement more (has higher opportunity costs), is at a disadvantage and prone to being held out by the player with the lower cost for non-cooperation (cf. Fearon 1998, pp. 278).

Hold-ups occur in situations where one or both of two trading partners make sunk relationship-specific ex ante investments (“reliance investments”) at the time of the contract conclusion. Reliance investments enhance the efficiency of this specific trade but have considerably less outside value – they are partly sunk. These investments cannot be contractually fixed, either because they are non-verifiable by a court, or because future contingencies are imperfectly foreseen. Hold-ups then occur when one party (usually that which has made no investments) imposes contract renegotiations on the investing party. Threatening to cancel the deal, the injuring party can partially expropriate the quasi-rents generated from efficiency gains of the relationship-specific investment. Anticipating this opportunistic behavior, potential hold-up victims react with underinvestment, which leads to an inefficient resource allocation and thus to an ex post inefficient outcome. The hold-up literature has a versatile field of application in economics (Edlin and Reichelstein 1996, p. 478). It spans industrial organization, labor, and comparative institutions (see for instance Hart and Moore 1988; Klein et al. 1978; Williamson 1985). Hold-ups also play a central role in more recent attempts to explain the boundaries of the firm (e.g. Grossman and Hart 1986).

119 The authors state that there are two types of contingencies, namely aspects of the environment (exogenous TC in our terminology) and aspects of behavior (endogenous TC). When writing a contract, contracting parties engage in a costly endeavor: Rigidity costs are incurred by adding an environmental contingency to the contract (explicating a future state of the world), whereas discretion spells out permissible or prohibited behavioral responses (policy instruments) in connection with an external contingency.
it environmental situations or human actions. Consequently, they erroneously assign a zero percent probability to a positive-probability contingency. Second, signatories assign the wrong probabilities to possible contingencies, or mis-specify their effects or magnitude. Transactors often omit crucial details or neglect the dynamic effects of their contractual regulations. Third, contracting parties regularly write down contradictory clauses and agree on terms that are subject to opposite interpretation.

As we will show shortly, bounded rationality may lead to inadvertent (in contrast to foreseeable) incompleteness of the contract. Although situations of factual ambiguity and accidental contract gaps are rather intuitive and per se trivial propositions, the discipline of economics is not well equipped to deal with them. In order to buy into this explanation, one must be willing to assume actors are of limited cognition and perception. Economists are very uncomfortable with this idea. More on this issue infra (subsection 3.1.4).

3.1.2 What types of “uncertainty” can be distinguished? The concept of incertitude

Parties exposed to uncertainty are unable to write complete contracts. The previous subsection dealt with sources of uncertainty. This subsection is geared towards an adequate categorization of the concept. Distinguishing properly the different facets of uncertainty will help us to structure contractual incompleteness more effectively in the next subsection.

In order to classify and catalog what contract theorists refer to as “uncertainty”, we venture a brief excursion into the discipline of technological risk management. Scholars committed to the study of technological risk have a more nuanced vision on the limits of certainty: They distinguish between knowledge of outcomes (effect, or results of contingencies), and knowledge of probabilities of occurrence of a contingency. Along those two axes, and under the general heading of “incertitude”, risk management experts map the concepts of “risk”, “uncertainty”, “ignorance” and “ambiguity” (see Chart 3.1).
### Chart 3.1 Incertitude in risk management – risk, uncertainty, ignorance, ambiguity

<table>
<thead>
<tr>
<th>Knowledge about likelihoods</th>
<th>Knowledge about outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes well-defined</td>
<td>Outcomes poorly defined</td>
</tr>
<tr>
<td><strong>Firm basis for probabilities</strong></td>
<td>RISK</td>
</tr>
<tr>
<td><strong>No basis for probabilities</strong></td>
<td>UNCERTAINTY</td>
</tr>
</tbody>
</table>

Source: Stirling (1999), Figure 4

Risk is defined as “a condition under which it is possible both to define a comprehensive set of all possible outcomes and to resolve a discrete set of probabilities (or density function) across this array of outcomes” (Stirling 1999, p. 16). Uncertainty on the other hand “applies to a condition under which there is confidence in the completeness of the defined set of outcomes, but where there is acknowledged to exist no valid theoretical or empirical basis for the assigning of probabilities to these outcomes” (ibid.). Ambiguity applies to situations where outcomes are ill-defined, yet there is some basis for probabilities of the event occurring. Ignorance, finally, “applies in circumstances where there not only exists no basis for the assigning of probabilities (as under uncertainty), but where the definition of a complete set of outcomes is also problematic. In short, it is an acknowledgement of the possibility of surprises” (ibid.).

Based on this more comprehensive representation and conceptualization of incertitude, we now propose a similar classification of what contract theorists usually refer to as “contractual uncertainty”, but what we will call incertitude. Classifying incertitude in contracts is a bit more complicated than in management of technological and environmental risk. In contracts, parties are dealing (a) with various contingencies that need to be anticipated, (b) with knowledge about contingency outcomes, (c) with knowledge about outcome probabilities, (d) with the (a)symmetry of contingency revelation ex post.
Consider Chart 3.2, which represents a structure along these four relevant dimensions: (i) specification of the contingency; (ii) definition of outcome and effect (“knowledge about outcomes” in Chart 3.1); (iii) definition of probability of occurrence (termed “knowledge about likelihoods” in above chart); and (iv) symmetrical revelation of contingencies (information observable and/or verifiable).

**Chart 3.2 Incertitude in contract theory – a novel taxonomy**

Source: author based on Sterling 1999, Figure 4

Notes: Real-life contracts are examined along four relevant dimensions: (i) Can a contingency be specified/forecast?; (ii) can the outcome be defined?; (iii) do contingencies happen with a previously known probability (-density)?; (iv) are contingencies observable/verifiable? Depending on whether these criteria are fulfilled (Y)/(*Poorly*) or not (N), different kinds of contractual incertitude can be distinguished. Note that only one branch of the tree leads to a complete contract; all the other types of incertitude (framed) inevitably lead to incompleteness in the contract.

---

Outcomes are ill-defined, because they are liable to more than one interpretation, explanation, or meaning, which cannot be determined from the given context.
As stated before, a contract can only be called complete, if signatories can comprehensively specify all relevant present and future contingencies. Consequently, the first question to address is precisely: (i) Can all contracting parties specify in advance all future contingencies, actions/policies, and events?\textsuperscript{121} If so, the next relevant question is criterion (iv), i.e. whether the anticipated contingency – once it happens – is revealed by nature to both (all) signatories (observable information), or whether it is private knowledge to one party.\textsuperscript{122} If the information is in fact observable, is it verifiable by an impartial third party?\textsuperscript{123} Whenever the contingency is not verifiable or non-observable, there must be “uncertainty” as to the better informed party’s actions. “Asymmetrical information” in the form of moral hazard or adverse selection then makes it impossible to write a complete contingent contract.

If contingencies are poorly specified, we shall call the resulting contractual incertitude “ambivalence”: Relevant passages in the agreement referring to certain contingencies are ambiguous and subject to discussion, dispute, and potentially to opportunism.\textsuperscript{124}

In case it is impossible to forecast or outline exactly the relevant contingencies, we shall talk of “uncertainty in the broad sense”. The subsequent question to tackle then is (ii) whether parties, although they cannot conclusively “nail down” the nature of the contingency, can nevertheless define the likely outcome (which entitlement is affected and with what magnitude of disruption?). Whenever neither contingency nor outcome can be specified, we shall call that type of incertitude “unforeseeability”.\textsuperscript{125} If the outcome is poorly defined, we follow Stirling (1999) and call this type of incertitude “ambiguity”. Just like ambivalence, ambiguity leads to misunderstandings, opposite interpretations, opportunism, and hence to disputes.

If the signatories are capable of clearly anticipating/outlining the effect or outcome of all unspecified contingencies, the contract displays “unforeseen contingencies”. The follow-up question now is (iii) whether parties can assign a probability (or probability density)

\textsuperscript{121} An event is a bundle of concomitant environmental or behavioral contingencies.
\textsuperscript{122} Criteria (ii) and (iii), which deal with the knowledge of outcomes and their probabilities, are superfluous issues: If all contracting parties can define every possible contractual contingency, they can \textit{ipso facto} specify outcome and outcome probabilities.
\textsuperscript{123} The distinction between observability and verifiability only makes sense in those contractual contexts where signatories can prosecute a claim in front of a third party, e.g. a court.
\textsuperscript{124} Checking for criteria (ii)-(iv) seems ineffective, because ambivalently defined contingencies are a shaky fundament to build on. Subsequent questions (ii)-(iv) would be equally ambivalent.
to the outcome. If so, we may speak of a “risky” contingency. We can further distinguish (iv) whether the contingency will be revealed symmetrically or not. Symmetrical revelation (observable information) leads to what we call a “type A risk”. Asymmetrical revelation shall be referred to as incertitude of the “type B risk”. Whenever parties are ignorant about the probability density function of the unspecified contingency, we shall call the result “uncertainty in the narrow sense”. Uncertainty in the narrow sense can then be divided according to the revelation of the contingency. Symmetrical revelation leads to “type A uncertainty”, private revelation to “type B uncertainty”.

3.1.3 Contractual incompleteness: A taxonomy

It is easy to see from Chart 3.2 that any situation other than the one featuring a fully specified set of symmetrically revealed contingencies displays contractual incertitude. A contract affected by incertitude is necessarily incomplete. Here is the logical nexus between incertitude and incompleteness: Incertitude is a condition which one or more signatories are exposed to ex ante, while incompleteness is the contractual outcome. If a contracting party is under any type of incertitude concerning relevant contingencies, s/he will not be in a position to write a complete contingent contract, try as s/he may. Chart 3.2 (overleaf) represents this result by framing all final nodes that lead to incompleteness.

125 Whenever outcomes cannot be defined ex ante, their probability cannot be defined either. The same goes for the revelation of a contingency. If the state of nature is not foreseeable, its revelation cannot be foreseen, either. Thus, criteria (iii) and (iv) would drop out of the equation.

126 Note that previously unspecified contingencies will automatically be non-verifiable to external third parties. Courts have a very hard time dispensing justice on issues that are not foreseen, and hence not mentioned in the contract. Parties cannot point to a specific contingency rule or obligation to prove contractual infractions.

127 Every contract is incomplete in many ways, but only certain facets of incertitude are of issue to signatories: Incompleteness is relevant (and hence needs to be addressed) only if the revelation of certain contingencies results in regret, or if it opens the floodgates to opportunism.
We are now ready to develop a rigorous taxonomy of contractual incompleteness.

1. **Strategic incompleteness:** This form of contractual incompleteness is an outflow of what we called incertitude of the “asymmetrical information”-type in Chart 3.2. Strategic incompleteness is given when - despite the fact that parties can comprehensively specify all possible contingencies – some important contingencies...
are *asymmetrically observable* (revealed privately to some parties, but not to all) or *non-verifiable* by a neutral third party.128

One party may have private knowledge of the occurrence (or magnitude) of the contingency at hand, which it can strategically withhold. When information is *unobservable* to one of the signatories, parties are not able to condition on it, and obviously will not even attempt to write an enforceable contract on the set of unobservable contingencies. The issue to which the information relates to is “noncontractible” (Schwartz 1992, pp. 279). Alternatively, where the occurrence of contingencies or certain actions is observable but *not verifiable* (cf. footnote 57 above for the difference between observability and verifiability), parties can informally condition on them, but are not able to write a legally enforceable contract.129

In sum, rational, anticipating parties will refrain from conditioning performance on asymmetrically revealed states of the world, or actions, since the better informed party can be expected to opportunistically abuse its information edge by misrepresenting the truth. Asymmetrical information will thus result in the impossibility of writing complete contingent contracts.

2. Accidental incompleteness. The second type of contractual incompleteness which we want to identify is *accidental incompleteness*. It is derived from *ambivalent/ambiguous* treaty language. Whenever the nature of the contingency is mis-specified, signatories have to deal with an uncertainty that we called *ambivalence* in the last subsection. Whenever the resulting outcome is poorly delineated, parties must live with what was termed *ambiguity*. *Accidental incompleteness*, though probably very common, is in need of further research (at least in the field of economics and L&E).130 Little is known about how and when ambiguous, contradictory, erratic, or incoherent language emerges, and how to take precautionary measures to prevent these types of incertitude from happening.

---

128 Another way of seeing it is that all contracting costs except for *endogenous ex post* transaction costs are absent (or assumed away).

129 Masten (1999, p. 28) submits: “The concern posed by non-verifiability is that, with the court no longer able to determine whether some aspect of promised performance has occurred, transactors stand to gain by strategically withholding information or by altering their behavior in ways that yield private benefits but reduce joint gains.” Salanié (1997, p. 175) confirms: “It is no use conditioning the contract on a variable if nobody can settle the dispute that may arise.”

130 To understand better the concepts of ambiguity and ambivalence, one should probably delve into social science disciplines (psychology, sociology) and linguistics.
3. **Ignorance.** This type of incompleteness stems from an incertitude type that we referred to as *unforeseeability*. As Chart 3.2 illustrates, whenever a contingency can neither be defined nor its outcome specified, contracting parties are simply at a loss: They are completely ignorant about the existence (let alone the occurrence) of a contingency. Consequently, they are not able to protect themselves properly against the incidence of unforeseeable contingencies. We will call the resulting incompleteness *ignorance*. By definition it is difficult to study ignorance in contracts, simply because no individual knows *ex ante* what s/he does not know. In a state of ignorance it is always possible that things happen which have been entirely excluded from previous consideration. Even in hindsight it is often difficult to track down the origins of surprise contingencies, and to detect underlying patterns.

4. **Inexorable incompleteness.** In Chart 3.2 we defined *uncertainty in the narrow sense* as contracts where contingencies cannot be specified, yet their outcome can be foreseen though not their probability. These conditions lead to what we call *inexorable* contractual incompleteness. We choose the nomenclature *inexorable*, because this kind of incompleteness is extremely difficult to come by (see *infra*). After all, how are parties supposed to deal with an event of known magnitude but unknown probability of occurrence? Potentially there exist hundreds and thousands of contingencies – should they all be considered in the contract?

Whenever the contingency revelation can be expected to be observable we speak of *inexorable incompleteness* of type *A*. A contract bestowed with asymmetrical revelation of information gives rise to *inexorable incompleteness* of type *B*.

5. **Type A or efficient incompleteness.** Whenever faced with a contract situation in which symmetrically-revealed contingencies of known effect, magnitude and probability distribution arise, signatories usually *choose* incompleteness out of rational cost-benefit considerations. We shall call the result *Type A*, or *efficient incompleteness*.

Based on the methodology of Battigalli and Maggi (2002), Horn and colleagues (2006; 2005) show in a formal incomplete-contract model that in a dynamic, non-stationary world it can be both rational and efficient for contracting parties to deliberately leave contractual gaps, and to refrain from writing a fully contingent contract. The transaction costs involved in researching, writing and bargaining over contractual obligations and permitted instruments under the full range of possible environmental conditions make contractual completeness simply impractical. The
costs involved outweigh the benefits of doing so.\footnote{It is not a new insight that the TC involved in considering all contractual contingencies are prohibitive, even if their probabilities are known \textit{ex ante}. This was argued by various authors (see e.g. Ayres and Gertner 1989 at p. 92; Macneil 1978 at pp. 871; Shavell 1980 at p. 468; Williamson 1985 at p. 70). Horn et al.’s contribution to contract theory lies in the formalization and operationalization of this important insight.} This is typically the case when contracts are relatively complex and display a large number of low-probability contingencies that can affect the value of contractual performance.\footnote{It is not a new insight that the TC involved in considering all contractual contingencies are prohibitive, even if their probabilities are known \textit{ex ante}. This was argued by various authors (see e.g. Ayres and Gertner 1989 at p. 92; Macneil 1978 at pp. 871; Shavell 1980 at p. 468; Williamson 1985 at p. 70). Horn et al.’s contribution to contract theory lies in the formalization and operationalization of this important insight.} \textit{Efficient incompleteness} is also chosen, whenever efficient responses to contingencies vary substantially and cannot easily be specified in advance (cf. Cohen 1999 at p. 81; Hadfield 1994 at footnote 15).

Hence, under the burden of significant transaction costs, contracting parties accept uncertainty (the term \textit{risk}, as discussed above, is more precise) over future conditions of the world and over possible responses. It is efficient and rational for signatories to conclude incomplete contracts, since it is cheaper to leave contractual gaps and to refrain from dealing with contingencies until they happen. As a result, contracting parties have to enter into \textit{ex ante} negotiations on how to deal with unanticipated contingencies. More on strategies of dealing with incompleteness \textit{infra}.

6. \textbf{Type B or necessary incompleteness.} In contractual situations where both outcomes and probabilities of contingencies can be specified (albeit at a cost of doing so), but some important contingencies are privately revealed to one party, \textit{type B or necessary incompleteness} is the result. Conceptually, this type of incompleteness is a combination of \textit{strategic} and \textit{efficient incompleteness}: \textit{Efficient incompleteness} means living with, accepting, the risk posed by future contingencies for reasons of prohibitive endogenous transaction costs. \textit{Strategic incompleteness}, now, is caused by asymmetrical information, either between signatories or between signatories and courts. Necessary incompleteness, then, combines the insights of these two types.

Under these circumstances contractual incompleteness is unavoidable: Parties neither have the chance of writing a complete contingent contract (as they could under \textit{efficient incompleteness}) nor of devising an elaborate incentive scheme (as they could under \textit{strategic incompleteness}). A contract cannot be made contingent (directly or indirectly) on something that is not foreseen (i.e. that has not been specifically provided for in the contract). Unforeseen privately revealed contingencies, in short,
3 Incomplete contracts

... provoke a kind of contractual incompleteness which is impossible to “cure” or “overcome”, even if signatories are assumed to be fully rational actors (Schwartz 1992). This will be consequential for our later examination of the WTO contracting context.

It now makes sense to define groups of contractual incompleteness:

Ignorance and accidental incompleteness together represent what we shall call inadvertent incompleteness: Both kinds of incertitude result in a contractual incompleteness which is haphazard, or accidental in nature. This set of types of incompleteness occurs frequently in real life (if one is willing to accept that actors are of limited, or bounded, rationality): Transactors think they have “nailed down” an issue, i.e. specified it comprehensively and in unambiguous terms. Yet the language chosen is in fact incompletely contingent, open to opposite interpretations, or inadvertently leaves gaps that can be abused by a party acting in bad faith. It may also be the case that contingencies occur that have not been anticipated to ever happen at all.

Next, and possibly complementary to inadvertent incompleteness, is what can be called foreseeable incompleteness. As Chart 3.3 shows, this term clusters strategic, efficient, necessary, and inexorable (type A and B) types of incompleteness. The idea here is that reasonably rational actors, i.e. those that have trust in their contracting capabilities, but may not always be able to contractually specify in advance all possible contingencies/outcomes, will sometimes accept and anticipate incertitude as a fact of life. They foresee that incompleteness is something they must deal with, and that, for better or worse, signatories must develop strategies for dealing with incompleteness in the contract.

The difference between inadvertent and foreseeable incompleteness is that in the case of the former, parties can only assume that incompleteness is looming, whereas in the latter case they know this for a fact; they can even assess the outcomes thereof. In other words, in situations of inadvertent incompleteness signatories are facing unforeseeable events, or “unknown unknowns”. In situations of foreseeable incompleteness, transactors are

---

132 If the probability of a contingency (or class of contingencies – an event) is low, then it may be less costly for contracting parties in the expected sense to resolve difficulties only on the chance that they arise, than to bear the costs of providing for the contingency with certainty by including the respective passage in the contract (cf. Shavell 1980 at note 7 for a numerical example).
faced with what can be called “known unknowns”. As we will see infra, this distinction makes a big difference when parties try to address and remedy contractual incompleteness.

This is not the only classification of incompleteness. Other authors have come up with different categorizations. However, we find these classifications to offer no analysis of incompleteness to speak of. Rather, they are incoherent strings of symptoms explaining the absence of an ideal contract. The classification of contractual incompleteness herein is “more complete”, because it is rooted in a more rigorous and structured analysis.

3.1.4 Do we have the technical toolkit for dealing with contractual incompleteness?

Now that we have distinguished six different types of contractual incompleteness resulting from six different kinds of incertitude, we are ready to assess how incompleteness has been (or can be) addressed by formal economic contract theory.

Incomplete-contract theorists from the “industrial organization” (IO) discipline in economics have examined in considerable depth those contractual relationships where strategic behavior is key – in fact where the presence of endogenous TC is the only contracting imperfection assumed to be present (see Tirole 1994 for a survey of the literature). In other words, IO predominantly deals with contexts of incertitude that give rise to strategic incompleteness: A previously well-defined contingency is either asymmetrically revealed, or observable but not verifiable by an independent court. Examples are situations of moral hazard, adverse selection, hold-out, or hold-up.

133 This may remind the reader of a notorious quote made by former U.S. Secretary of Defense, Donald Rumsfeld, given at a press conference on February 12, 2002. Certainly, his statement borders on poetry: “Reports that say that something hasn’t happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don’t know we don’t know” (taken from the official transcript of the press briefing; available at http://www.defenselink.mil/transcripts/transcript.aspx?transcriptid=2636).

134 Schwartz (1992, pp. 278) notes five “causes” for contractual incompleteness: Ambiguous language, undeliberate [sic!] omissions, efficiency reasons, asymmetric information, and adverse selection. Shavell (1980, p. 468) only mentions efficiency and verifiability reasons; Tirole (1994) distinguishes between unforeseeable and unforeseen contingencies that can give rise to incompleteness. Ayres and Gertner (1989 at footnote 29) distinguish two “ways for a contract to be incomplete: a contract may fail to specify specific future contingencies [and it may be] insensitive to relevant future contingencies”, even though parties’ duties are fully specified. Salanié (1997, chapter 7) mentions efficiency reasons, non-verifiability and bounded rationality as possible reasons for incompleteness. Dunoff and Trachtman (1999, p. 34) allege three causes of contractual incompleteness: (i) inadequate knowledge of the future, (ii) rational cost-benefit deliberations, and (iii) asymmetrical information.

135 We re-emphasize that models featuring strategic incompleteness do not assume what we termed above as uncertainty in the broad sense (cf. Chart 3.2). Unanticipated events of any kind are notably absent (usually because non-stationarity of the environment is assumed away). Under this type of incompleteness parties (and
Economists have devised ingenious strategies aimed at overcoming the strategic incompleteness of those types of contracts. In order to make up for the asymmetrical information gap, which one or both contracting parties (or the court) are confronted with, contract economists suggest that signatories draft elaborate payment schedules or sharing rules, that either force the better-informed party to reveal its information, or that align the informed party’s behavior with the interests of the less informed party/parties. These information-forcing and incentive alignment contracts, however, can only mitigate, but never make up for, welfare losses relating to the contractual incompleteness.\footnote{\textbf{Chains} designed to elicit voluntary performance of unverifiable actions depart from the Arrow-Debreu ideal in leaving gains from trade potentially unrealized relative to the cooperative (nonstrategic) outcome” (Masten 1999, p. 28).}

With regard to type A or efficient incompleteness, some literature can be found that addresses issues of “breach” and remedy in contracts which are beset by symmetrically revealed but previously unspecified contingencies of a known probability density. We mentioned above the important contribution of Horn et al. (2006; 2005) to incomplete contract theory. Horn and his colleagues have devised a methodology to \textit{endogenously} derive contractual “uncertainty” over future events (we called that type of incertitude risk) – and consequently the incompleteness of the contract. They do so without having to jettison the assumption of rationality. This is important, since the economic toolkit functions best if the rationality precept is maintained.\footnote{This is a significant result, since it refutes the objections of influential “complete-contract theorists” (Masten 1999, p. 28), such as Maskin and Tirole (Maskin and Tirole 1999; Tirole 1994). Complete-contract scholars, although acknowledging the presence of uncertainty in the real world, are wary of embracing the concept as long as scholarship lacks an explicit theory of uncertainty, and hence fails to operationalize it.} As was demonstrated above, in a situation of “risk-cum-symmetrical-information”, even rational parties find it too difficult to anticipate, evaluate and write down every possible detail that the future may bring.

Several authors of contract theory have brought forth formalized models which describe ways of dealing with situations that lead to what we refer to as efficient incompleteness. IO scholars examine one-off (non-repeated) contracts involving reliance investments (see Edlin and Reichelstein 1996, and references therein; Klein, et al. 1978; Klein 1996).\footnote{Edlin and Reichelstein (1996, p. 481) find that conceptually it does not matter if the relationship-specific investment is non-contractible due to non-verifiability by external third parties, or because its description (which depends on future contingencies) is prohibitively difficult.} Efficient incompleteness, too, has been discussed in repeated-interaction contexts, where self-enforcement, instead of sharing rules, can sustain cooperation. In

\textbf{the court} know exactly what the environmental and contractual context is; they just don’t have access to private information, cannot “see inside the other party’s head”.

\textbf{136} “Contracts designed to elicit voluntary performance of unverifiable actions depart from the Arrow-Debreu ideal in leaving gains from trade potentially unrealized relative to the cooperative (nonstrategic) outcome” (Masten 1999, p. 28).

\textbf{137} This is a significant result, since it refutes the objections of influential “complete-contract theorists” (Masten 1999, p. 28), such as Maskin and Tirole (Maskin and Tirole 1999; Tirole 1994). Complete-contract scholars, although acknowledging the presence of uncertainty in the real world, are wary of embracing the concept as long as scholarship lacks an explicit theory of uncertainty, and hence fails to operationalize it.

\textbf{138} Edlin and Reichelstein (1996, p. 481) find that conceptually it does not matter if the relationship-specific investment is non-contractible due to non-verifiability by external third parties, or because its description (which depends on future contingencies) is prohibitively difficult.
WTO scholarship, several formal contributions examined those contractual situations (Ethier 2001a; Furusawa 1999; Rosendorff 2005; Rosendorff and Milner 2001; Sykes 1991, Appendix A).

Although contracting parties are often confronted with exactly those situations of necessary or type B incompleteness (as Schwartz 1992 convincingly shows), we submit that this is an under-researched area in contract theory (Grossman and Hart 1986; Klein 1996; Salanié 1997). With respect to contract situations of previously unforeseen and asymmetrically revealed contingencies of a known probability density, Salanié submits (ibid., p. 188): “In my view, [practical approaches] should eventually study the consequences of incomplete contracting when information is asymmetric.” Economic contract theory has so far been reluctant to formally tackle this kind of incompleteness (notable exceptions include Shavell 1980; Herzing 2005; Copeland 1990; Hungerford 1991; Kovenock and Thursby 1992; Horn 2006).

For all the other kinds of contractual incompleteness – accidental, and inexorable type A/B incompleteness and ignorance – the economic toolkit is empty. Assessing these types of incompleteness poses non-trivial technical challenges to the discipline of economics:

- **Inexorable incompleteness** originates from situations of uncertainty in the narrow sense: Although the outcomes of contingencies can be specified, the probability of occurrence is unknown. Without being able to assign a probability to contingencies, economic actors have no means of assessing expected costs and benefits. Econometric logic of rational choice and utility maximization, which fundamentally relies on actors’ rational cost-benefit calculations, breaks down in those situations.

- When it comes to accidental gaps (caused by poor specification of contingencies, and probabilities of outcome), the canon of economic theory has again reached its limits: Some pundits have argued that whenever parties accidentally leave contractual gaps, they cease to be rational. Tirole (1994) warns against giving up the assumption of rationality in the absence of any workable theory of “bounded rationality”. He contends that any theory of bounded rationality and human error should be able to specify whether errors occur, why they occur, and what the consequences are for all signatories. Tirole rejects an “errors-just-happen” approach as arbitrary and unscientific. Without any workable theories as to the nature and content of accidental gaps, so the argument goes, the discipline of economics must fail to rationalize
contractual incompleteness *without* venturing the risk of making arbitrary (inoperable) assumptions, or of giving up the economic toolbox.\textsuperscript{139}

- The same goes for those forms of contractual incompleteness, where the contingency can neither be specified, nor its outcome anticipated. We called this type of incertitude *unforeseeability* and the related concept of incompleteness *ignorance*. Economists feel very uncomfortable with *ignorance*: Unforeseeability is a challenge to the precept of rationality which fundamentally drives all results in economics. Contract economists claim that bounded rationality (not to speak of irrationality) is an unworkable concept, given that the economics profession has made very little progress in understanding, let alone modeling it.\textsuperscript{140} The concept of unforeseeability renders economic methodologies and solution techniques futile: How can you assign probabilities to something you are ignorant of *ex ante*? How can you write a contract on issues that you are not aware of? Unforeseeable contingencies by definition do not affect contracting parties’ decision-making or actions (cf. Ayres and Gertner 1989 at footnote 34).

In sum, the toolkit which economics has to offer for tackling contractual incompleteness is quite scant. Chart 3.4 illustrates the realm of incompleteness that economic contract theory has *not* been able to grapple with (shaded area).

---

\textsuperscript{139} See also Maskin and Tirole (1999), who claim that without clear theories of how uncertainty comes about and what issues exactly contracting parties are (un)certain of, the use of contractual uncertainty is methodologically arbitrary. Unless scholarship comes up with an explicit theory of bounded rationality and uncertainty, complete-contract proponents submit that full rationality and unlimited foresight is the more accurate methodological benchmark.

\textsuperscript{140} “Unfortunately, while many authors have insisted on the need for such an approach, little progress has been made yet” (Salanié 1997, p. 188).
How can formal contract theory deal with the fact of not being able to tackle these (arguably most interesting) aspects of contractual incertitude and incompleteness? Three possible answers are conceivable:

- A first way to deal with the limited scope that economic approaches to contract theory can offer is to completely drop the assumption of actors’ rationality. This, however, comes at a cost to structure and with a loss of economic tools. Basically, it means
3 Incomplete contracts

leaving the realm of *economic* contract theory and moving to *non-economic* explanations.

- Second, maintain the orthodox concept of full rationality, and discuss only those contract types that contemporary contract theory can master. This is a feasible, technically clean, and legitimate way of doing research. It is much the way that the discipline of IO has progressed. IO literature focuses on relatively simple hierarchical principle-agent transactions featuring narrowly defined sets of problems such as employment-, R&D-, or franchising contracts (Salanié 1997).\(^{141}\) Thanks to the extreme rationality assumptions of unlimited foresight and cognition which prominent IO strands of research maintain, agents are assumed to treat contracts “as if they were complete”: Actors form expectations over all possible states of the world and are thus able to solve all sorts of hidden knowledge/hidden action problems.

The problem, however, with this approach to contract theory is the inherent trade-off between modeling convenience, economic dogma, and “getting it right”. A substantial loss of explicatory power and significance may well ensue when researchers follow a “we can’t deal with it, so let’s pretend it isn’t there” approach. For example, although mathematically elegant, IO approaches to incomplete contracts are hardly helpful in dealing with complex, long-term, repeated-interaction, and multi-party contracts.

- The third and most effective way of dealing with contractual incompleteness is to let go of economic orthodoxy and try to be more pragmatic. Instead of lamenting over missing theories of bounded rationality, contract scholars may just take it as given (treat as exogenously) that contracting parties cannot foresee all future contingencies and in addition regularly make errors when negotiating contracts. The essential question then is to figure out ways that *reasonably rational* transactors can deal with any unforeseen (or even unforeseeable) situation at hand. Prominent proponents of the L&E discipline have taken this pragmatic route. For decades, L&E scholars have been concerned with assessing strategies – economic and non-economic – for filling contractual gaps in continuing contractual relations. Whenever contracts are incomplete, contract law (or a court) must be in a position to refer to rules and strategies for resolving any issue which has not been explicitly addressed by the

\(^{141}\) As stated above, in IO literature incomplete contracts are either an outflow of current asymmetry of information (e.g. in employment contracts), or caused by the indescribability of the future states of the world (e.g. in R&D contracts).
parties. Contractual L&E literature assesses *intra*-contractual and judicial strategies of gap-filling in incomplete contracts.\textsuperscript{142}

Every scholar is ultimately free to tackle the issue of incompleteness in whatever way s/he thinks is most convenient, feasible, and/or promising. As we stated *supra*, it is not absolutely necessary to give up the tenet of rationality to believe that contractual incompleteness occurs in real life. Even completely rational transactors may decide that it is efficient to leave contractual gaps with respect to low-probability or low-impact contingencies.\textsuperscript{143}

However, it is not our objective to examine incompleteness as economists see it, but to describe how contracting parties in real life cope with contractual incompleteness (*infra*). More to the point, we want to assess what kinds of flexibility mechanisms WTO parties have designed (and should be designing), given the Agreement’s inherent incompleteness.

Therefore, we favor a pragmatic approach to incompleteness, which is to consider a broad range of real-life types of incertitude that hinder contracting parties. In order to be as encompassing as possible, and especially in order not to exclude *inexorable incompleteness* from our consideration, we derail from the assumption of perfect rationality and instead embrace the notion of *reasonably rational transactors*.\textsuperscript{144} A key

\textsuperscript{142} Issues addressed by L&E scholars are, for instance, whether courts should be interventionist or reticent (Calabresi and Melamed 1972; Cohen 1999; Kaplow and Shavell 1994; Schwartz 1992; Hermelin and Katz 1993); whether courts should encourage efficient, fair, or socially just allocations (Ayres and Gertner 1989; Craswell 1999; Macneil 1978; Scott 1987); how the optimal contract remedy or damage measure should be calculated (Barton 1972; Mahoney 1999; Rogerson 1984; Shavell 1980, 1984).

\textsuperscript{143} In addition, if “rational” incompleteness is combined with asymmetrical information, the tenet of rationality can also be maintained. The combination of Horn et al.’s (2006; 2005) approach to *efficient incompleteness* with that of asymmetrical revelation of contingencies may represent such a pragmatic and promising avenue for further research: Under the assumptions of *risky* contingencies and *asymmetrical* information, contracting parties are conscious of their incapacity to write a complete and flawless contract. Signatories acknowledge the fact that private information may provoke strategic misrepresentation of the truth, and consequently lead to opportunistic behavior. Anticipating this, rational trade negotiators must take contractual precautions in the form of fallback rules that apply in case a previously unspecified contingency occurs.

\textsuperscript{144} We prefer the newly-created term of “reasonable rationality” to that of “bounded rationality”. The term “bounded rationality” was coined by Herbert Simon (1955), who used it to question the standard rationality assumption of orthodox economics. Over the years, however, the term has assumed a life of its own and spiraled in two connotative directions that we do not approve of: For some, bounded rationality is equivalent to *irrationality* of actors (e.g. Kahnemann 2003), while for others it is related to concepts of *constrained choice* and Bayesian updating of otherwise perfectly rational actors (e.g. Rubinstein 1998). We do not wish to maintain either of these two extreme assumptions. Actors are neither irrational, nor superhuman utility maximizers. Reasonably rational actors are guided by rational choice: They have complete, transitive, and continuous sets of preferences and engage in thorough cost-benefit analyses. Yet they are not perfect at doing so, and human errors in judgment, evaluation, foresight and decision-making may slip in occasionally. In short, although actors do the best they can to act rationally, they are prone to erratic decisions.
tenet of the concept of reasonable rationality is that individuals are not rational supercomputers, but are not foolhardy either: Important decisions are taken after careful deliberation of the costs and benefits they may entail. However, information constraints as well as the complexity of the contractual environment make the occurrence of contractual gaps inevitable. Signatories of the reasonably rational type may lapse, but are cognizant of this fact, and learn from previous errors. Therefore, contracting parties try to minimize the mutual welfare losses triggered by contractual incompleteness through the best possible design of the contract.

3.1.5 The effects of incompleteness on contracting behavior

Incertitude and the resulting incompleteness of the contract make contracting difficult. In this subchapter we want to assess how and to what extent incompleteness may impact on the contracting behavior of signatories.

In general, we find that contractual incompleteness creates two types of problems: On the one hand, it provokes gaps and gives rise to legal loopholes where there should not be any. On the other hand, incompleteness creates rigidity, where the complete contingent contract (CCC) would instead mandate flexibility. The first type of error harms victims of \textit{ex post} non-performance, whilst the latter harms injurers. Both error types are prone to reduce the \textit{ex ante} cooperation level of transactors.

3.1.5.1 Victims’ willingness to cooperate

Incomplete contracts may not sufficiently consider abusive reactions to environmental eventualities or opportunistic party conduct. They often contain legal loopholes, and misdraw the line between \textit{intra}-contractual, permissive- and \textit{extra}-contractual, illegitimate behavior. The victim party is unable to tell whether an observable outcome (e.g. the quality of a contractor’s performance, or a country’s level of protectionism) was provoked by an outside shock (in which case the injurer’s action can be seen as legitimate), or by opportunistic (and hence prohibited) behavior (cf. Copeland 1990). In other words, the contingency is not sufficiently observable to the victim.

\footnote{The following quote is apt in this regard: “So mistakes will be made […] Our assumption that there are no systematic mistakes means only that actors will not miscalculate in the same way every time they confront a similar set of circumstances. The subjects in our models are not systematically myopic, systematically gullible, or systematically naïve about the responses of others” (Grossman and Helpman 2001, p. 15).}
Alternatively, some action by the injurer is clearly discernable as opportunistic, but its enactment has not been anticipated in the letter of the agreement (or the contract language is blurry). Hence, although the injurer acts in contravention to the spirit of the contract, the victim cannot point to a specific contractual provision that explicitly outlaws such behavior. The contingency is non-verifiable and the victim cannot bring the injurer to trial for contractual misconduct.

A third option is that the injurer’s action is found to be in contradiction with the contract, but the damage caused to the victim cannot be easily assessed by anyone except the victim – who has an incentive to exaggerate the magnitude of the impact.

In all three cases, the victim’s ability to enforce the contract is imperfect, because efficient enforceability (observability, verifiability, quantifiability) of the contract cannot be safeguarded. Recall that enforcement is a function of enforceability and enforcement capacity (cf. subsection 2.1.3). So, even if a victim signatory possesses the sufficient enforcement capacity to punish the injurer (or have him punished), she is likely to be hesitant to cooperate extensively, if the enforceability of the contract is porous. Chart 3.5 illustrates this point: It corresponds to Chart 2.2 above, but shows the contracting constraints from the victim’s point of view.¹⁴⁶

The “hit-and-run”-curve (H&R) represents the victim’s disutility caused by the injurer’s defection from the terms of the initial contract (in Chart 2.2 above this corresponded to the injurer’s utility from defection). The C-curve represents the victim’s expected benefits from continued cooperation. The C&C-line represents the liquidated-damages punishment the injurer suffers when defecting from a contract enforceable by an impartial third party. Chart 3.5 illustrates the impact of weak enforceability on the victim’s willingness to enter into contractual obligations. Whenever the contract contains a gap, or displays a legal loophole, it may well happen that defection is imperfectly detectable, verifiable, or quantifiable. This skews the H&R-curve to the left. Spurious enforceability reduces the victim’s proclivity to cooperate. The scenario with insufficient enforceability results in a much less enforceable contract: For the self-enforcement case compare the darkly shaded area produced by the intersection of the dotted H&R and C-curve with the lightly shaded area resulting from the intersection of H&R’ and C curve. If signatories can rely on external enforcement, the cooperation space is reduced from $C^{law}$ to $C^{law'}$ through the westwards movement of the H&R curve.

¹⁴⁶ Cost and benefit curves are interchanged in comparison to Chart 2.2 above.
3 Incomplete contracts

Chart 3.5 The impact of contractual incompleteness on the victims’ commitment

Source: author, based on Keck and Schropp (2007, Figure 2)
Notes: This is a slightly modified version of Chart 2.2. The graph illustrates the impact of weak enforceability. It represents, from a victim’s perspective, cooperation payoffs and disutility from defection (vertical axis), and levels of cooperation (horizontal axis). The H&R curve represents the victim’s disutility if the injurer defects from the contract. The C-curve represents the Continuation value of the game, the discounted value from future cooperation. The C&C-line represents the injurer’s expected disutility for defection if there is third-party enforcement in place.

In summary: The problem with contractual loopholes created by contractual incompleteness is that too few constraints confine the injurer. He can thus go on and engage in opportunistic breach of contract in situations where the CCC would have deterred such behavior. Anticipating such opportunism, the victim is less inclined to engage in extensive ex ante commitment.

3.1.5.2 Injurers’ willingness to cooperate

From the injurer’s perspective, a serious problem of contractual incompleteness is that of overly rigid constraints, or over-regulation (e.g. Goldstein and Martin 2000; Smith 2000; Setear 1997). In a dynamic and ever-changing environment “regret contingencies” are pre-programmed, caused by insufficient, or insufficiently clear, contract language.

Regret contingencies occur whenever performance of the actual terms of the agreement leaves gains from trade unrealized ex post, i.e. given the information available to
parties/courts at the time performance takes place. Injurers will wish to seize regret contingencies, since this increases their welfare.\footnote{A word on the concept of “regret”: A signatory experiences regret when an \textit{ex ante} envisioned transaction value is not realized in light of the newly revealed information. An unanticipated contingency arises which – had it been known \textit{ex ante} – would have changed the initial content of the contract. Termed differently, regret occurs in instances where pursuit of the CCC would have excused performance, but the provisions of the real (incertitude-ridden, incomplete) contract erroneously mandate it. Regret is a function of the magnitude of the unexpected contingency, or shock (the “regret contingency”), and of the level of \textit{ex ante} commitments. Note that the concept of regret is strictly different from that of opportunism: Whereas giving in to (or “acting on”) regret produces welfare gains (otherwise the CCC would not have mandated non-performance), opportunistic action by definition is welfare-depreciating (see footnote 54 above).}

Rigidity is to be understood as inadequate and insufficient consideration of future external and behavioral contingencies.\footnote{A contract that simply states “John sells me a bushel of wheat at the price of 5$ tomorrow” displays rigidity, and is therewith incomplete in the sense that it does not consider events like John falling into a coma, or a thunderstorm devastating John’s field, in which case it would be better for John not to perform, given he pays me adequate compensation.} It has a negative impact on injurers’ \textit{ex ante} willingness to cooperate. Consider Panel ‘a’ of Chart 3.6 (next page): It captures a self-enforcement situation from the perspective of the injurer. An unexpected, previously unspecified state of nature occurs in the form of a negative shock ($\theta$).

Imagine a simple repeated-interaction contract. The contract is sustained by the victim’s enforcement threat. The injurer can be expected to cooperate as long as his hit-and-run advantage of doing so (the $H&R(\theta)$-curve in Panel ‘a’\footnote{The size of the hit-and-run payoff is dependant on the magnitude of the negative shock.}) does not exceed the punishment (the discounted opportunity costs of defection, depicted by the $P$-line).\footnote{Rosendorff (2005, pp. 393) provides some intuition of the curvature of the $H&R(\theta)$-curve. The $P$-line, notably, is flat, since the punishment is independent of today’s revelation of the shock $\theta$.} The rational injurer will defect as soon as the regret contingency exceeds some threshold-value ($\theta^*$), where the hit-and-run benefit exceeds the costs of defecting and prompting punishment by the victim. Depending on the probability of occurrence and of its expected magnitude, the presence of regret (here in the form of environmental shocks) can significantly decrease the stability of a contract.

More to the point, the injurer’s willingness to enter into a rigid contract and (if so) his level of cooperation, is likely to be a function of the victim’s enforcement capacity and the expected shock level. To see this, see panel ‘b’ of Chart 3.6: If the victim is not able to enforce the rigid contract at all, i.e. is either lacking enforcement capacity or enforceability, or both (not pictured), the injurer’s commitment level does not depend on
it either. The injurer will be willing to commit to what he perceives to be his optimal level of commitment, \( C^{\text{opt}} \), no matter how high future shocks will be.

**Chart 3.6 (a;b) The impact of contractual rigidity on injurers’ commitment**

![Chart 3.6](chart.png)

Source: author (top panel based on Rosendorff 2005, Figure 1)

Notes: This chart is seen from the **injurer’s** perspective. Panel ‘a’ plots the magnitude of some unanticipated, exogenous shock \( (\theta) \) on the horizontal, and foregone utility from cooperation and utility from defection on the vertical axis. \( P \) stands for the disutility from being punished for defections. Panel ‘a’ shows the threshold shock level above which the injurer prefers to defect \((D)\) instead of cooperating \((C)\). Panel ‘b’ plots the expected shock level \([E(\theta)]\) against the injurer’s cooperation level. It illustrates how the injurer’s willingness to cooperate is a function of the victim’s enforcement capability and the magnitude of exogenous shocks (yielding \( C^{\text{opt}} \) and \( C^P \), respectively).

However, things change if the victim possesses sufficient capacity to enforce her claims under the (overly rigid) letter of the contract: Under full self-enforcement power on the part of the victim, but expecting shocks of a small magnitude, the injurer commits to the cooperation level \( C^P \).\(^{151}\) Whenever regret contingencies are expected to be sizeable, the injurer refrains from entering into the rigid agreement in the first place. Alternatively, if the expected magnitude of shocks is higher than a threshold level \( E(\theta^P) \), the injurer’s inclination to cooperate is likely to decrease, since he will not be able to seize regret

\(^{151}\) Note that \( C^{\text{opt}} \) is always higher than \( C^P \).
contingencies under a rigid contract that prohibits any escape *ex post*. If the injurer expects contractual regret to reduce (or even outweigh) future gains from cooperation, his willingness to make *ex ante* concessions declines.152

If the injurer can foresee neither the probability of occurrence nor the magnitude of future exogenous shocks (shocks hence come as a surprise), it can safely be argued that the injurer’s level of *ex ante* concessions is likely to decrease in the face of a rigid contract. Injurers will not engage in “risky”, but instead only in “safe” (read: less shock-ridden) transactions, which brings down the scale and scope of *ex ante* promises, i.e. the level of commitments.153

To conclude this section on the impact of incompleteness on signatories: If not addressed properly, contractual incompleteness is bound to bring down signatories’ *ex ante* willingness to cooperate, and therewith the level of gains to be had from cooperation. In a contract where both (or all) parties know with certainty whether they will be victims or injurers, each signatory will conduct a private calculation of expected costs and benefits to making contractual *ex ante* concessions. The resulting deal is then likely to be decided by the most reluctant cooperator, i.e. by the party least willing to give concessions (see above the accompanying text of footnote 72). The most reluctant liberalizer (victim or injurer) effectively decides on the common level of exchange.154

### 3.1.6 Summarizing the findings on contractual incompleteness

It was our aim to ground this paper with an adequate structure and characterization of contractual incompleteness. We rooted incompleteness in the concept of incertitude and were able to differentiate various categories of incertitude, which in turn gave rise to corresponding types of contractual incompleteness. We showed the extent to which economic contract theory is able to deal with various aspects of incompleteness, and

---

152 The concave curvature of the cooperation line is drawn for demonstration purposes. However, it is intuitive that the injurer’s willingness to cooperate shrinks at a decreasing rate.

153 Under *uncertainty in the narrow sense* (specification of outcome is possible, but not of probabilities), injurers who want to optimally shape their *ex ante* commitments can engage in various techniques, such as scenario analysis, sensitivity analysis, or fuzzy logic. Their level of cooperation is then a function of the quality of the available tools, and of their level of risk aversion.

154 If, at the conclusion of the contract, parties are uncertain about their future roles, or the frequency of becoming victim or injurer, they must assess *ex ante* the probability of assuming the role of victim or injurer. The result, however, is likely to stay the same: Either the contract is never entered into or the most reluctant liberalizer will set the level of commitment for all signatories.
hinted as to serious shortcomings in the economic toolbox. We assessed the impact of contractual incompleteness on signatories’ *ex ante* behavior.

Why is it at all important to examine the sources of incompleteness and to categorize the concept? It may be argued that ultimately it does not matter what determines contractual incompleteness, since the result – inefficiently contingent contracts – stays the same. We disagree. Consider the following reasons:

First, being able to distinguish various types of incompleteness helps researchers to make accurate use of and draw upon adequate literature. When examining an incomplete contract situation of some sorts, the researcher should be conscious of the type of underlying incompleteness that his/her subject of research is affected by. Only s/he who understands her/his problem can assess the scope and limits of different solution strategies, methods and approaches, and consequently take advantage of them. S/he will know which literature to apply – and at what “cost” of doing so (in terms of loss of explanatory scope).

Different strands of contract theory literature have dealt very selectively with issues of incertitude and incompleteness, for example by taking into consideration some transaction costs and assuming away others, or by assuming complete, bounded, or reasonable rationality of actors. Consequently, scholars have modeled very different types of incomplete contracts: IO theorists have focused on asymmetrical information and *strategic incompleteness*; some scholars applied economic tools examining what we call *efficient incompleteness*; L&E scholars have more pragmatically dealt with *necessary* and *inexorable* variants of incompleteness. The results and insights generated from various strands of contract theory do not easily carry over to any desired contractual situation, but only to similar contracting contexts.

Second, it seems quite important to understand the nature of incompleteness and to classify it, because neither signatories nor researchers should treat all forms of incomplete contracts as if they were nails to be hit with the same contracting hammer.

---

155 When reviewing the literature it was at times frustrating to find out that scholars hardly ever make explicit their underlying concept (or “type”) of incompleteness. This omission makes it notoriously difficult to compare different incomplete contract models/approaches and to assess and compare their findings.

156 As we will show much later in Chapter 6.3, approaches to the WTO that only consider *strategic* or *efficient incompleteness* and assume away other kinds of market imperfections, are only of limited explanatory value in the context of the world trading system.
3 Incomplete contracts

Different types of contractual incompleteness require different remedies, i.e. different ways of dealing with them. This will be the topic of the next subchapter.

3.2 How to deal with contractual incompleteness: Strategies of gap-filling

The presence of various types of contractual incertitude precludes the existence of Pareto-optimal complete contingent contracts in real life. Every contract is incomplete and necessarily contains gaps. As pointed out, these gaps can be provoked by asymmetrical information settings, uncertainty over future environmental contingencies, mishap (ambiguities in language and unforeseeable contingencies) or a mix of those reasons depending on the contractual entitlement under examination, the contracting situation at hand and – not least – on the methodology one subscribes to.

In section 3.1 we developed the concept of the reasonably rational actor, who was characterized as an individual that knows that she cannot contract flawlessly: Try as she may, some contingencies will be omitted (either accidentally or for efficiency reasons), some mis-specified, and others imperfectly observable/verifiable. Hence, reasonably rational contracting parties know that their initial substantive agreement is mute to a range of contingencies that may occur in the contract performance phase – their contracts most definitely will contain gaps.

Gaps (insofar as they are not pure “information gaps” produced by asymmetrical information) give way to the sensation of regret. Regret must be taken seriously: On the one hand, parties anticipating un-seized regret contingencies may decide to abstain from the contract, or to scale down their cooperative ambitions ex ante. On the other hand, capturing regret produces welfare-enhancing gains from non-performance. The

---

157 For the definition of the concept of “regret”, see footnote 147 above.
158 As explicated above (subsection 3.1.5.2) contractual rigidity causes injurers to reduce their ex ante cooperation concessions.
159 A simple example may illustrate the significance of regret contingencies and the Pareto-superiority of ex post non-performance in an incomplete contract setting (adapted from Shavell 1980, p. 467): Suppose i) a buyer enters into a contract with a seller to produce and deliver a machine, and pays at the outset; ii) the value of owning the machine is worth $200 to the buyer; and iii) the relevant contingencies concern the production cost, which will become apparent to the seller before he actually begins the production process. Assume that the cost of production will be $100 with probability 0.99 and $1000 with probability 0.01. Consider first a contract that requires the seller to perform regardless of production cost (a rule of inalienability), and that the price paid at the outset is, say, $150. Then the expected value of the contract to the seller is $150 – [0.99($100) + 0.01($1000)] = $150 – $109 = $41. The value to the buyer is $200-$150 = $50. Consider now the alternative contract that allows for regret contingencies and requires the seller to perform only if the production cost is $100, and that the
problem with regret, however, is to distinguish it from flat-out opportunism: Injurers may suffer a true regret contingency, yet they may also just backtrack from some contractual obligations for reasons of personal enrichment.

The prospect that contracts (as written) may leave gains from trade unrealized, leads to the issue of how parties are to respond to opportunities for mutually advantageous *ex post* adjustment. In other words, contracting parties *ex ante* have to concern themselves with the question of how to deal with contractual gaps that unfold *ex post*. They have to devise a series of gap-filling strategies to address various instances of incertitude relevant to the contract at hand.

Albeit unachievable, the Pareto-efficient complete contingent contract is the normative benchmark that contracting parties aspire to, since it constitutes the first-best outcome. Reasonably rational contracting parties, cognizant of their inability to conclude a CCC, are faced with the challenge to find the most efficient, yet achievable, substitute for the CCC in the face of contractual incompleteness. When entering the design phase of the contract, transactors strive for finding the second-best governance structure that all involved parties give their consent to and that takes into account the contextual circumstances. When faced with contractual incompleteness, signatories can choose four basic strategic trajectories:

(i) minimizing the number of gaps through *comprehensive contracting*;
(ii) seizing regret through drafting *flexibility mechanisms*;
(iii) minimizing the room for regret by means of *additional contract language and relational contracting*;
(iv) *delegating responsibility* to a third party.

As we will show below, comprehensive contracting (strategy (i)) is geared towards replicating the *original*, namely the CCC itself. Seizing regret (strategy (ii)) aims at replicating the *outcome* of the CCC by means of designing contractual rules of flexibility (contingency measures and rules of default). Minimizing room for regret (strategy (iii)) basically builds on the precautionary principle, and on endogenously strengthening trust in and cooperation of the relationship. Strategy (iv) finally consists of commissioning a contract price is lowered to, say, $145. Then, the expected value of the contract to the *seller* is $145 – 0.99($100) = $46, and its expected value to the *buyer* is 0.99($200) – $145 = $53. The second contract is Pareto-superior to the first: Both seller and buyer strictly prefer the second contract that allows for non-performance in the case of regret (cf. also Sykes 1991, Appendix A for a similar example).
neutral third party with gap-filling responsibility. In the following, we discuss the applicability and merit of each of these strategies of contractual governance.

3.2.1 Circumnavigating incompleteness: Comprehensive contracting

Comprehensive (quasi-complete) contracting is the strategy of choice between negotiating parties, when the incompleteness of the contract is “bridgeable” (Cohen 1999). This means that the level of incompleteness is not too high and the costs of writing the detailed contract not so extensive, such that they would outweigh the gains of doing so. Typically, comprehensive contracting is assumed to tackle situations of strategic incompleteness, featuring extremely simple contracting situations beset by asymmetrical information. Alternatively, risky contingencies (cf. Chart 3.2) giving rise to type A (efficient) incompleteness can also be bridged, although this strategy is not always a practical option due to the high costs involved in doing so (see our discussion in subsection 3.1.3 at point 5).

Under strategic incompleteness – i.e. in the presence of observability or verifiability imperfections – contracting parties can fully anticipate contractors’ ex post behavior. To prevent information-induced opportunism from happening, parties ex ante agree to circumvent the information gap. They either do so by “pegging” the contractual exchange to some related observable information (e.g. a worker’s performance outcome instead of his effort), or by devising incentive compatible (payment or transaction) schedules that force the better informed party to reveal its information edge.

In the presence of type A risk (leading to efficient incompleteness), parties can also try to turn the incompleteness into completeness. To this end, transactors should have to nail down every environmental and/or behavioral contingency by specifying exactly its outcome, by assigning the according probabilities and by prescribing detailed behavioral responses to be taken by all signatories (Horn, et al. 2005). This is a costly endeavor that requires a lot of upfront research and bargaining. Hence, efficient incompleteness can only be overcome if the set of relevant environmental contingencies is small, that is, if there is a manageable number of high-probability, high-impact contingencies.
The comprehensive contracting strategy assigns a *course of action to every possible contingency* – just like the CCC.\(^{160}\) The comprehensive contract lays out a full plan of action *ex ante*, exhaustively specifying *substantive* entitlements, and spelling out in complete detail the corresponding protection belt of *auxiliary* entitlements. Complete contracts leave no room for discretion *ex post*; they are by definition renegotiation-proof.\(^ {161}\) They mandate a compulsory specific performance of contracting parties at any point in time, which is another way of saying that – just like a CCC – they are protected by an enforcement rule of inalienability.

Comprehensive contracting strategies aim at replicating the CCC or, at least, relevant aspects thereof. Being able to do so presupposes particularly simple contractual situations – or indeed a significant abstraction from reality on the part of the researcher. Given the limited application scope of comprehensive contracting it is quite surprising how much attention contract theorists in the economic disciplines of microeconomics and IO have expended on it. It is probably fair to say that the nature of most real-life contracts is too complex to be “bridged” by some state-contingent incentive schedule or sharing rule.\(^ {162}\)

Whenever the route of mimicking the CCC through exhaustive substantive rule-making is forestalled by the constraints of the contracting context, parties are forced to devise other ways of addressing the problems caused by the contractual incompleteness. The other three alternative strategies to comprehensive contracting introduced above (and elaborated below) might be called *incomplete* contracting, since they do not aim at

---

\(^{160}\) The term “comprehensive contract” originates from Hart (1995, p. 22), who aptly notes that under a comprehensive contract “there will never be a need for the parties to revise or renegotiate the contract as the future unfolds”.

\(^{161}\) “Although contracts designed to elicit voluntary performance of unverifiable actions depart from the Arrow-Debreu ideal in leaving gains from trade potentially unrealized relative to the cooperative (nonstrategic) outcome, economists generally regard contracts optimally designed to deal with information asymmetries as complete in the sense that such agreements (i) still fully specify each party’s performance obligations for every possible contingency, and (ii) yield the best possible outcome given the information available to the courts at the time the agreement is carried out and thus ‘never need to be revised or complemented’” (Masten 1999, p. 28, citing Holmström and Tirole 1989, p. 68, in the end of the quote).

\(^{162}\) Indeed, as Masten (1999, pp. 28) points out, comprehensive contracting has been disappointing as a positive theory: “Aside from the broad prediction that efficient sharing rules will balance incentives for one party against inefficient risk bearing by that party or the incentives of trading partners, asymmetric information models yield few testable hypotheses. One reason for this is the ‘extreme sensitivity’ of optimal incentive schemes to slight changes in the relation between actual performance and verifiable information [...]. Complete contract theory also fails to account for the observed simplicity of sharing rules in most real world contracts [...A]ctual contracts incorporate few if any explicit contingencies and generally use simple, typically linear pricing schemes.”
3 Incomplete contracts

replicating the CCC. Instead, they strive – in one way or another – to cope with the inevitable contractual incompleteness in the best possible manner.

3.2.2 Seizing regret: Drafting flexibility mechanisms

Contractual flexibility mechanisms are designed with the aim of efficiently seizing the gains that ex post regret contingencies pose. As we mentioned in the introductory Chapter 1, flexibility mechanisms are intra-contractual, that is, legal provisions of ex post discretion which legitimize a departure from original performance as promised.

The strategic difference between comprehensive contracting and the design of flexibility mechanisms becomes apparent: Whereas comprehensive contracting dictates a complete plan of action and thereby exhaustively specifies the transaction terms, flexibility provisions lay down general contracting goals that do not specify the contractual exchange in all its detail, but rather outline the basic rules of the game. Whenever contracting parties craft flexibility mechanisms, they usually leave the substantive entitlements rather unspecified and instead focus on the design of auxiliary entitlements, or rules of entitlement protection. Contractual flexibility instruments aim at reproducing the CCC’s outcome, but not its content. Their role is not to achieve completeness, but to “heal” the contract without remedying incompleteness itself.

There are two sorts of flexibility mechanisms: Contingency measures intend to circumscribe ex ante contingency outcomes which allow for non-performance on the part of injurers. Contingency measures are additional auxiliary entitlements. Default rules, on the other hand, are general fallback rules of entitlement protection. By assigning rights of ex post discretion to victims and injurers, signatories attempt to trigger efficient ex post behavior by spelling out general rules of conduct. Default rules do not distinguish between welfare-improving and -depreciating situations in explicit language, but set in place a general incentive structure that fosters appropriate post-contractual behavior.

3.2.2.1 Contingency measures

Whereas comprehensive contracting was described as a strategy geared towards the specification of substantive entitlements (by comprehensively defining each contingency, anticipating its outcomes and probability of incidence, and by assigning the exact plan of response action taken by signatories), contingency measures are a less ambitious way of remedying incompleteness. They are driven by the desire to circumscribe the outcome of certain groups of (previously unspecified) contingencies or
events, and to prescribe in exact terms the permissible action to be taken in response by signatories. Contingency measures are fine-tuned “dependant auxiliary entitlements” that are integrated into the contract (see discussion in subsections 2.2.1 and 2.2.2): Each contingency instrument is pegged to a single substantive entitlement, has a unique level of conditionality (preconditions) and application scope, as well as an entitlement protection rule. The aim is to specify in as much detail as possible – but without having to specify contingencies themselves – instances where the use of contractual non-performance is welfare-enhancing. In short, contingency measures lay out the broad contours of regret.

The use of contingency measures is encumbered by a number of potential problems:

- First, contingency provisions are often difficult to operationalize. In many instances, it is not clear whether a contingency has occurred, and whether reacting to the event falls under the ambit of the contingent escape clause, or constitutes a violation of the terms of the agreement.

- Second, contingency measures only apply to certain types of contractual incompleteness, namely those with signatories able to define and observe the effect of a future state of nature: The property of circumscribing clusters of regret contingencies, and of prescribing a unique course of action, makes contingency measures unfit for dealing with inadvertent incompleteness (the cluster consisting of accidental incompleteness and ignorance; cf. Chart 3.3 above), since for these types contingency outcomes are not readily predictable.

In addition, whenever the contingency provoking a certain outcome cannot be symmetrically observed, the legitimate source of regret cannot unambiguously be distinguished from illegitimate and opportunistic behavior. Disputes are a natural result. This rules out instances of necessary or type B inexorable incompleteness. Hence, contingency measures only seem applicable to type A (efficient) incompleteness and to inexorable incompleteness of type A.

---

163 Art. XX of the GATT is a good example of a contingency measure: It lays out the circumstances under which general exceptions to tariff liberalization can be enacted, what must be considered when doing so, and how ex post discretion can be exercised by the injuring party.

164 Disputes between signatories occur precisely because the nature of the gaps cannot be put into words ex ante. The circumstances under which unforeseen contingencies may occur, and the desired responses, can only be sketched.
Third, writing down contingency measures is a costly enterprise for signatories – with decreasing marginal returns at that. A lot of upfront research and legal drafting has to be devoted to codifying contingency measures. Only with this can parties find exactly those clusters of regret contingencies that warrant the effort. After the auxiliary entitlement is crafted, rules of entitlement protection and enforcement have to be supplied. Contingency measures are suitable for high-probability-high-impact groups of contingencies. However, probability of occurrence is known \textit{ex ante} only in situations giving rise to \textit{efficient incompleteness}. Not so in contracts displaying incompleteness of the \textit{inexorable-type A} sort, where transactors cannot know which clusters of regret contingency to focus on. Writing contingency measures of flexibility must often be seen as “a shot in the dark”.

Fourth, contingency measures are not only costly to write, but also costly to enact for the injurer. It is usually the injurer who has to prove that some regret contingency occurred and that this eventuality actually falls under the purview of a certain contingency measure. Doing so may often entail some serious up-front, or signaling costs, which do not entail any efficiency-enhancing added value (they are sunk costs). Thus, the presence of high enactment costs may have a “chilling effect” on injurers: Although they experience regret and know that non-performance at face value would be efficiency-enhancing, they refrain from doing so, just because the fixed costs of enacting the respective contingency instrument are prohibitively high.

Fifth, contingency measures will not cover the entire scope of regret, simply because of the considerable costs involved in researching, drafting and writing them. Some outcomes provoked by regret contingencies are never captured, especially those of low probability and low impact (if the probability is initially known to the transactors). This sort of regret is either lost, or a \textit{second line} of more general flexibility mechanisms is in place to seize previously unspecified types of regret.

Finally, in addition to an insufficient application scope, contingency measures are also prone to human contracting error. Adding contractual language is always a hazard. The more explicit and elaborate a contractual clause gets, the more it is prone to contracting errors. Two types of contracting error may occur when parties draft contingency measures of entitlement protection: A \textit{Type-I error} (or false positive) is the tendency to include states of nature under the purview of a contingency measure,
where non-performance does not lead to a mutual welfare-increase, but instead opens the floodgates to opportunistic abuse.\(^{165}\) A *Type-II error* is a false negative: The language of the contingency measure is overtly strict, or rigid. It wrongly prohibits non-performance, where the contracting ideal, the CCC, would mandate non-performance to be welfare-enhancing. Regret cannot be captured (see section 3.1.5.2 above). Box 3.1 explains the general vices and virtues of express contract language.

**Box 3.1: The double-edged sword of express language in incomplete contracts**

Contracting parties often try to fight contractual incompleteness by striving to make their contract ever more complete: They attempt to anticipate and assess future contingencies and/or outcomes by devising complicated contract schemes that involve refined substantive and additional auxiliary entitlements, mixed entitlement protection regimes, and other behavioral prerogatives. Negotiating and writing up contingencies is costly and time-consuming. Verifying the realization of contractually included contingencies is also costly for uninvolved third parties. Moreover, a drive towards contractual completeness via express language is prone to severe pitfalls:

- **Ambivalent catchall phrases.** The more explicit wording transactors chisel out, the more they risk interspersing textual ambivalence and ambiguities. This poses a dilemma for the framers of a trade agreement: Efforts to fill gaps contractually by adding treaty language will prove counterproductive if they introduce further ambiguities. Cohen (1999, pp. 80) points out that parties striving to complete an incomplete contract often resort to catchall phrases that may sound stringent and are seemingly broad enough to achieve completeness, but often result in the exact opposite. Clauses like “best effort”, “gross inequity”, “serious injury”, “unforeseen developments”, “like products”, or “appropriate countermeasures” bear the inherent need for interpretation. They entail incompleteness rather than completeness, and create gaps rather than closing them.\(^{166}\)

- **Risk of “reverse hold-up”.** Quite often, parties think they have anticipated and forestalled all relevant future contingencies that may give rise to opportunistic behavior. Elaborate and detailed contract schemes protect the potential victim party. However, unanticipated regret contingencies may occur which make adherence to these elaborate contractual terms generally suboptimal. The very party protected by the contract terms as a potential victim now uses contractual rigidity to blackmail her

\(^{165}\) Think of the contingency measures of AD and CvD in the WTO: The ambiguities inherent in these contingent events give rise to protectionist abuse of these “unfair trade” instruments. National bureaucracies have substantial leeway to “interpret” the evidence. The porous language of ADA and SCM allows political players to “shape” the laws and regulations which govern the work of these bureaucracies (Finger, et al. 1982).

\(^{166}\) Masten (1999, p. 34) submits: “[C]ontracts that use terms such as ‘best efforts’, ‘gross inequity’, or ‘substantial performance’ to describe contractual obligations leave the parameters of acceptable performance ultimately to the courts.”
partner: The victim threatens to insist on the overly rigid terms of the original contract. Klein (1996) calls this opportunistic victim behavior a “reverse hold-up”.167

- **Risk of “petty litigation”**. Excessive contract language may be prone to Type-I errors. Explicit substantive language creates the presumption of completeness. In case of a dispute, haphazard gaps may not be recognized by the courts for what they are. This may then allow “trigger-happy” signatories to initiate disputes that are formally correct, but foil the spirit of the agreement (WTO 2007, section II.D.3.b). The result of this over-reliance on the letter of the agreement, some observers maintain, is a violation of the initial spirit of the contract and the loss of shared sense of cooperation (Charnovitz 2002c). It is argued that a fully legalized, over-regulated system could lead to excessive litigation, a hostile atmosphere, vengeful parties, and growing mistrust. Issues of “legal” and “illegal” threaten to replace considerations of “fair”, “sensible” and “feasible” (Setear 1997).

- **Neglecting default rules**. In general, parties that focus on drafting a “watertight” complete contract risk overlooking the need for including default rules of entitlement protection (cf. subsection 3.2.3 infra). This is an unfortunate situation since disputes arising from the revelation of unanticipated gaps will have to be settled by courts. Courts will need to interpret language and to supply gap-filling clauses and are likely to be overwhelmed by the task. Notably, if contingencies are revealed in a manner not verifiable by the court, judges will not be able to sensibly fill the gap, even if they have the best intentions (Schwartz 1992).

As a general conclusion to these pitfalls of express language, we submit that explicitness at times risks replacing **rigidity** for **flexibility**: If contractual contingencies are molded into express language they are usually inflexible against environmental changes and prone to prompt inefficient party behavior and costly disputes. Flexible adaptation to a dynamic world becomes much more difficult in these situations (Downs et al. 1996; Goldstein and Martin 2000; Koremenos, et al. 2001; Rosendorff 2000; Rosendorff and Milner 2001). Rational contracting parties can be expected to anticipate problems of inefficient *ex post* adjustment that excessive explicitness entails: Intuitively, if a party will have to assume that pseudo-completeness will forestall flexible *ex post* adjustment to regret contingencies, and instead lead to a reverse hold-up or petty disputes, it will scale back its contractual efforts. This signatory will either refrain from entering into such a contract or at least scale back its contractual obligations and engage in less-than-full *ex ante* commitments (cf. section 3.1.5.2 *supra*).

---

167 “[W]riting something down to be enforced by the court creates rigidity. Since contract terms are necessarily imperfect, once something is written down transactors can engage in a hold-up by rigidly enforcing these imperfect contract terms, even if the literal terms are contrary to the intent of the contracting parties. This is what occurred in the Fisher Bodies-General Motors case [the classic example of reverse hold-up], where the written contract terms that were meant to prevent General Motors from holding up Fisher were actually used by Fisher to create a much greater hold-up of General Motors” (Klein 1996, p. 467).
Various L&E scholars have recognized the pitfalls of overzealous contracting and of designing ever more fine-grained contractual entitlements. They suggest that explicit language should not attempt to eradicate contractual incompleteness, but to give structure to it by formulating general rules and entitlements. Hviid (1999, p. 55) suggests that parties should focus on two topics: That of defining in broad terms the proposed outcome of co-operation, namely the object and purpose of the transaction (substantive rules of entitlement in our nomenclature), and that of designing the punishment strategies (i.e. the default rules of entitlement protection). The more complex a contract, the more contractual emphasis should shift from devising “a detailed specification of the terms of the agreement to a more general statement of the process of adjusting the terms of the agreement over time – the establishment, in effect, of a ‘constitution’ governing the ongoing relationship” (Goldberg 1976, p. 428).

3.2.2.2 Default rules

A possibly complementary way of providing ex post flexibility for signatories of an incomplete contract is the design of default rules of flexibility. Default rules (DR) are imperatives “that define the parties’ obligation in the absence of any explicit agreement to the contrary” (Craswell 1999, p.1). They are also known as “backstop, enabling, fallback, gap-filling, off-the-rack, opt-in, opt-out, pre-formulated, preset, presumptive, standby, standard-form, and supplementary rules” (Ayres and Gertner 1989, p. 91).

DR are unspecified provisions of entitlement protection, i.e. non-specific rules of ex-post adjustment, apt to resolve any issues that have not been explicitly addressed in the contract before. Signatories assign rights of ex post discretion by specifying general rules of entitlement protection ex ante. This allows parties to seize regret contingencies by engaging in mutually advantageous ex post contract modification after the revelation of any previously unanticipated contingency.

168 Shavell (1980, p. 488) warns against the use of express language and suggests that more specified contract terms should be contemplated only if “the use of an incomplete contract together with a damage measure [i.e. an unconditional default rule] would lead to significant inefficiency – when it would induce parties to act in a way that departs substantially from how they would act under a Pareto efficient complete contingent contract.” This can for example be the case when the (personal, subjective) value of a good to the victim is very large compared to the costs of breach to the victim.

169 The term “default” is somewhat unfortunate, since it has nothing to do with “defaulting” from (i.e. violating) the terms of the agreement. The concept of default rather originates from IT-programming, where default rules are rules that a program follows in “default” of an explicit choice by the user to have some other principle apply (Ayres and Gertner 1989 at footnote 24).

170 An interesting twist here is that a contract staffed with a set of DR (one DR for each traded entitlement) is comprehensive in the sense that no matter what happens, it prescribes a course of action. However, it is incomplete in that some of these actions involve ex post discretion by the contracting parties.
The major advantage of unspecific or unconditional default rules is that they can be applied to various types of incompleteness. The only precondition for drafting a workable DR is thereby that the outcome be definable *ex ante*. Hence, default rules of flexibility can be utilized to overcome all types of incertitude that give rise to what we clustered in the term *foreseeable incompleteness* (*strategic-, type A or efficient-, type B or necessary-, and inexorable-type A and B* variants of contractual incompleteness; cf. Chart 3.3 above). DR may even work for the *ignorance* type of incompleteness, as long as parties agree that they are really dealing with a contingency that they previously did not fathom.\footnote{In case parties disagree on the occurrence or the nature of the contingency, default rules may not work. Note that once the contingency has occurred, it presents a *fait accompli* that one (or more) party/parties may benefit from. This party/these parties may have little desire to change this outcome through the application of an *ex post* DR (if they expect to lose out from doing so). As an example: Imagine a land-lease contract between lessee *A* and owner *B*. The contract stipulates that *A* can harvest the spoils of the land, provided he pays his lease to *B*. Suppose, a DR is in place stipulating: “In any event not considered in the contract, *A* and *B* divide gains and costs to equal shares.” Next day, *A* finds a pot of gold on the leased land. This event was not foreseen by the parties. *Ex post*, however, *A* has an incentive to declare the pot of gold part of the spoils of the land to which he is entitled to. *B*, on the other hand, claims that the discovery falls under the purview of the default rule, and wants the DR to take effect.} *Accidental incompleteness* (caused by incertitude types of *ambiguity* and *ambivalence*), however, can hardly be tackled properly with a DR, if only because signatories will be in disagreement as to the nature of the contingency. If *ex post facto* parties cannot agree on the occurrence, nature, and magnitude of the contingency at hand, they will have trouble assigning the proper DR.

As contractual rules of flexibility, default rules are very versatile in dealing with contractual incompleteness. Theories on default rules and on the choice of efficient DR have attracted considerable academic attention (see Ayres and Gertner 1989; Craswell 1999; and Dunoff and Trachtman 1999 at chapter III.D for literature overviews). DR theories basically examine the contractual and contextual circumstances under which intra-contractual non-performance should be permissible, and at what price this option should come for the party engaging in the “breach”.\footnote{While most of the academic discussion on DR circles around what DR impartial courts should apply in case parties have failed to specify contractual contingencies, it is clear that sophisticated contracting parties will devise strategies that lay out what to do in case a gap occurs during the performance of their contract. This goes even more for the international realm, where courts are notoriously weak (Pauwelyn 2006).} We will dedicate section 3.3 to this salient question.
3.2.3 Minimizing regret: The principle of precaution

The use of flexibility mechanisms was described as a strategy to seize the regret potential of an incomplete contract by designing efficient rules of *ex post* discretion. A different strategy is pursued by signatories whenever they engage in precaution.\(^{173}\)

Cognizant of the inevitable incompleteness of the contract, and to prevent possible acts of opportunism from happening, signatories engaging in precautionary strategies aim at minimizing the room for regret, should contractual gaps occur nevertheless. Opportunism can be curbed by *endogenously* strengthening trust and cooperation of the relationship. To achieve these twin goals, three strategies are at hand: diligence, preamble language, and relational contracting.

Although the precautionary principle may help address all types of contractual incompleteness, it seems particularly apt in situations of *inadvertent incompleteness* (cf. Chart 3.3), i.e. in contracts fraught with incertitude of the types of *ignorance*, *ambivalence*, and *ambiguity*.\(^{174}\)

### 3.2.3.1 Diligence and preamble language

*Diligence* as a strategy of remedying contractual incompleteness basically consists of transactors’ efforts to write the best contract they can. Issues connected to the use of this strategy have been discussed at length above (*comprehensive contracting* in subsection 3.2.1 and Box 3.1). Writing “good”, thorough, contracts is substantially more difficult than writing extensive contracts. Upfront research, information gathering, benchmarking, bargaining and drafting are key – and ultimately costly endeavors. Yet the incurred benefits from performing diligence are unknown and undiscovered to transactors due to the difficulty of constructing a counterfactual: How can parties assess the payoffs of having spared themselves disputes arising from, say *ambiguity* and *ambivalence*?

Another precautionary strategy is a carefully drafted *preamble language*, which lays out the general spirit of the contract, and common intentions of the parties. Just as with the drafting of default rules, signatories use general, non-contingent, language. Although preambles are usually at a level of generality that make them un-informative when it comes to using them concretely to interpret the various legal provisions committed

---

\(^{173}\) On the use of the precautionary principle in science and technology, see Stirling (1999).
Incomplete contracts (Mavroidis 2007, chapter 1, p. 10), they can nevertheless be helpful in two ways: First, they can serve neutral third parties (courts) as a point of reference when interpreting the contractual gap at hand, and second, just like commandments, they may prevent signatories from engaging in opportunistic behavior in the first place – a merit that is again difficult to assess due to the impossibility of constructing a counterfactual.

3.2.3.2 Relational contracting

Another application of the precautionary principle in incomplete contracts is the strategy to establish a tightly-knit relationship between signatories. If a contractual gap surfaces or becomes an issue in a relational contract, parties are said to rarely resort to contract terms or courts to settle their disputes about obligations. Rather, mutual norms of the relationship fill this void, and help transactors resolve disputes in an amicable and non-opportunistic fashion (Hviid 1999, p. 53).

There is a growing literature in the legal, L&E, and even economic realm dealing with relational, or fiduciary, contracts. A relational contract is characterized by longevity, a continuing relationship and substantial incompleteness. As a consequence, relational contracts are said to be governed by shared norms, values, and a sense of “what is right” rather than through explicit rules and elaborate provisions. It is contended that it is often impossible to try settling disputes in relational contracts by resorting to the letter of the contract, since no uninvolved third party can verify implicit contract details, or manage to track the history of behavior. In fact, appealing to an impartial court and pointing to contractual provisions is often counterproductive, because it sours the atmosphere of cooperation.

Due to repeated interaction within a particular well-defined group of actors, and due to a good deal of contractual incompleteness, an implicit set of relationship norms may

---

174 This is so, because precautionary measures are relatively unspecific, or undirected, in nature. They help improve the general climate of contracting. Under contractual situations of foreseeable incompleteness, however, parties can actually anticipate outcomes, and thus can apply more targeted schemes in the form of pinpointed contingency measures, for example.

175 Although there is no generally accepted definition, scholars typically take relational contracts to be complex, long-term, incomplete, and repeated-interaction agreements (cf. Schwartz 1992 at footnote 1; Hviid 1999 at p. 46; Goetz and Scott 1981 at pp. 1091). Goetz and Scott (ibid.) adopt a stricter definition contrasting a relational (incomplete) to a complete contingent contract: “A contract is relational to the extent that the parties are incapable of reducing important terms of the arrangement to well-defined obligations. Such definitive obligations may be impractical because of inability to identify uncertain future conditions or because of inability to characterize complex adaptations adequately even when contingencies themselves can be identified in advance […]Long-term contracts are more likely than short-term agreements to fit this conceptualization, but temporal extension per se is not the defining characteristic.”
3 Incomplete contracts

replace the letter of the contract. In Box 3.2 we will address the interesting question of whether “going relational” actually represents a self-standing gap-filling strategy which differs distinctly from other forms of governance. We argue in the negative: Relational contracting is not the result of some ingenious contracting scheme, but rather the outflow of contextual constraints (longevity, repeated interaction, complexity). The strategy which signatories ultimately resort to – implicitly or explicitly – is that of designing apposite rules of contractual flexibility.

Box 3.2: Relational contracts – a concept apart?

Contributions to relational contracting (literature reviews can be found in Baird 1990; Goetz and Scott 1981; Scott 1987; 1990; Hviid 1999; Craswell 1999) tend to treat relational contracting as a concept separate to all other concepts of contract design. We disagree and contend that relational contracts abide by the same basic contractual logic and follow the same fundamental precept as all contracts. The only thing special about relational contracts may be its heavy reliance on ex post flexibility mechanisms.

The L&E literature is in conflict regarding the nature of relational norms. Relational contract theory falls in two camps: an internal relational school and an external one. Scholars of the external relational school believe that contracting parties are guided by non-economic/non-efficient norms that transcend the relationship.176 According to authors like Macneil (1978; 1981; 1982), contracting parties choose external frames of reference as relationship norms: They are directed by general and socially accepted (codified and non-codified) norms of what society thinks is fair, adequately participatory, or distributionally just (Schwartz 1992, p. 275).

A second group of scholars believes that it is internal relational norms that guide signatories in a fiduciary contract. A non-economic strand of the internal relational school holds that contracting parties derive their behavioral guidelines not from external norms, but from the norms and history intrinsic to the common contractual relationship (Hadfield 1990). Another, more economically-minded branch applies standard contractual logic to relational contracts (e.g. Goetz and Scott 1981; Scott 1990; Craswell 1999). Proponents of this view contend that relationship-specific norms may differ in form, but not in substance, from written contractual norms: Relational norms are behavioral guidelines intrinsic to the relationship, and apt to fulfill the objective of the contract as efficiently as possible. The same Paretian welfare-maximizing logic that is common to all contractual rules also applies to relational agreements.

Does this make relational contracts a concept apart, and as such incomparable to “classical” contracts? Rather not. As for the internal relational approach, Schwartz (1992, p. 276) has persuasively shown that the concept of relational contracting

necessarily collapses with standard Paretian contractual logic, and hence does not add substantially to the analysis of contracts in general: Internal relational theories mandate the protection of parties’ normative expectations. This means that in the presence of contractual gaps contracting agents base their decisional criteria on what they would have expected of each other, had the gap been known *ex ante* (“hypothetical consent”). Contracting parties strive to maximize their utility and as such cannot be expected to ever assume an obligation that is not in their best interest. Norms that derive from these expectations will be efficient, i.e. they will maximize the parties’ collective welfare. Thus, rules representing the norms of the relationship must adhere to mutual efficiency, which is exactly the view that L&E cherishes (see also Craswell 1999, p. 13).

Relational contracts are certainly more complex than simple barter-like transactions or one-off principle-agent contracts. They are more difficult to model, and most certainly contain more gaps (of various types of incompleteness) than simple classical contracts. This implies that there is an inherent need for flexibility. We believe that, when talking of inherent *norms*, relational contract scholars actually mean flexibility rules in the form of contingency measures and of default rules of entitlement protection. The more complex the agreement, the more important the concept of flexibility. DR may even be the crucial norms that structure any long-term contractual relationship. As Hviid (1999, p. 55) points out, norms are crucial as focal points since relational contracts are characterized by a multiplicity of possible equilibria. Issues of *ex post* flexibility may then gain prominence, because they direct signatories’ *ex post* actions, without having to explicate specific rules of behavior.

Although we think that the external relational approach is logically at fault, and oppose this perspective on methodological grounds, we note that the same concept of flexibility applies here as well: What external relationship scholars in fact have in mind when referring to “external frames of reference” and “relationship norms” is that parties rely on *external rules of default* in order to fill their contractual gaps.

---

177 This is probably why the discipline of economics has so far only approached relational contracts very tentatively. Economic scholars have tackled only two issues in connection with relational contracts, namely i) what makes relational contracts superior to a series of short-term contracts, and ii) how does self-enforcement in repeated interactions work (see Downs and Rocke 1995; Hviid 1999).

178 In the international sphere this study deals with, authors of the external relational approach (e.g. Chayes and Chayes 1990, 1993b) claim that public international law takes precedence in ordering the interaction of sovereign countries. Consequently, external relational contract scholars run into an “endogeneity problem” (Rosendorff 2005, p. 389): If States were guided by external norms, they needn’t have signed an international agreement in the first place, since they intended to comply anyway. Economic thinking, in turn, claims that countries comply because they have signed an agreement and take actions (or refrain from taking actions) contrary to how they would have acted in the absence of the agreement (Downs, et al. 1996).

179 The external relational approach is irreconcilable with the precept of methodological individualism and member-made law.
3.2.4 Delegating responsibility: Using courts as gap-fillers

A final strategy for remedying contractual incompleteness that has been suggested in the literature is to deliberately delegate gap-filling authority to previously uninvolved, impartial third parties (which we call courts for shorthand). Outsourcing this responsibility to courts, to our mind, is an appropriate governance strategy in cases of inadvertent incompleteness (the collective term that we assigned to accidental incompleteness and to ignorance). But will transactors freely choose delegation with respect to those categories that we bundled under the name foreseeable incompleteness (see Chart 3.3)? Will signatories ex ante take the conscious decision to refrain from writing preambles, devising pay-off schedules, drafting contingency measures, and designing default rules in contracting situations where they can actually foresee incertitude to be on the horizon?

At face value, delegating gap-filling responsibility to courts seems a smart idea – it absolves parties from having to think about future regret contingencies before they actually happen. Time and costs are seemingly saved for the contracting parties. However, this argumentation is ultimately a straw man. The idea of contracting parties deliberately yielding constructive gap-filling to courts in situations where it is not absolutely necessary collides with the reasonable rationality assumption of actors: When courts are charged with the task of completing a contract, they will do so first and foremost by interpreting the rules and norms intrinsic to the parties’ relationship.\(^{180}\) Thereafter, in a second step, they will consult norms external to the contract at hand.\(^{181}\)

Judges presume that rational parties expect them to engage in constructive gap-filling by concocting a clause (or giving an interpretation of an existing clause) in a way that signatories would have consented to ex ante, had they foreseen the existence of the contingency at hand. This “would-have-wanted” guideline for courts is called the “default rule of hypothetical consent” by Ayres and Gertner (1989, pp. 93). Since by contractual logic parties would have had an interest in maximizing their utility (by virtue

\(^{180}\) Contract-intrinsic gap-filling can refer to parties’ intent, current or prior conduct, or to the terms of the written agreement.

\(^{181}\) When referring to external norms, courts presume that parties intended to contract with references to some (possibly non-economic) standard external to the written contract. Courts assume that it is the parties’ desire to resort to external (socially accepted) default rules of flexibility. The U.S. Uniform Commercial Code (U.C.C.) has been said to be a collection of external default rules that parties wish to resort to, in absence of any indication to the contrary (Ayres and Gertner 1989, pp. 95). In public international law, the Vienna Convention on the Law of Treaties (VCLT) and the ILC Draft on State Responsibility may be seen as collections of external DR of transnational contracting conduct (Dunoff and Trachtman 1999, p. 35).

- 102 -
of the normative benchmark of the CCC), courts thus see it as their task to come up with the *ex post* welfare-maximizing solution at hand. More specifically, courts see it as their duty to fill the contractual gap in a way the Pareto-efficient CCC would have prescribed it (Craswell 1999; Goetz and Scott 1981; Scott 1990). Now, can it be in the reasonably rational transactors’ interest to let an uninvolved third party act as a constructive gap-filler *by default*? We think not:

- First, as a matter of methodological principle, we fail to see how it can be in the interest of contracting parties to cede responsibility to a court which may revert to some *external* standards of gap-filling. As we stated in the discussion of external relational contract theory *supra* (Box 3.2), it strikes us as odd that reasonably rational actors would opt for contract-extrinsic rules of default instead of shaping their own ones. Recourse to an external frame of reference conflicts with the assumptions of methodological individualism: Self-interested contracting parties prefer making their own rules and regulations which are in accordance with each participating member’s short-, medium-, and long-term interests. The primary logical frame of reference for parties thus consists of intrinsic (endogenous) norms, rules and regulations – not some exogenous legal or normative codex.182

- Second, even if the court applies the dogma of “hypothetical consent” when filling contractual gaps, this is not in the principal interest of rational contractors. Courts are effectively charged with concocting a default rule that parties could have crafted themselves (and probably even better): When courts want to elicit *contract-intrinsic* joint maximizing terms (that would have been agreed to by perfectly informed rational parties in a complete contract situation), they necessarily have to engage in *contextual* interpretation (Cohen 1999, p. 83). Interpreting an incomplete contract, now, is *conceptually identical* with clarifying which party owns which entitlement, and establishing how that entitlement is to be protected (Craswell 1999, p. 15). But how can a court clarify the contractual entitlement regime sensibly without detailed guidance from the contracting parties? Parties must tell the court how it is to interpret their intentions. They do so by writing down *ex ante* the traded entitlements (primary rules of contracting), and the rules of protecting them (secondary rules of contracting).

---

182 Parties *do* regularly give effect to an external set of default rules (such as the U.C.C. or a standard tenancy agreement). Yet they can be expected to mention precisely which set of norms they want to be referred to, and when exactly it is to apply. In addition, not every contracting situation lends itself to the application of easy-to-use standard codices of default that contractors can resort to.
3 Incomplete contracts

This is a task not to be ceded to a third party. Ultimately, there is no benefit for parties in ceding primary and secondary rules of rule-making to uninvolved third parties.

- Third, courts operate not necessarily with perfection,183 are expensive to establish and to maintain. In addition, it is costly to provide court members with the information necessary to come up with a balanced judgment.

- Finally, we note that in situations that we referred to as necessary incompleteness, even flawlessly operating judges cannot fill contractual gaps – try as they may. If a court cannot verify the facts of the case (such as the allegation that an employee “shirked” his contractual duties), it cannot complete the contract in any meaningful manner (see Schwartz 1992).184

In sum, courts are an indispensable instance in contracts, whenever inadvertent contractual gaps need to be filled and ambiguous contract language needs to be interpreted. However, when it comes to situations of what we termed foreseeable incompleteness, contracting parties must give courts detailed directions in the form of entitlement protection rules, or rules of default. Reasonably rational actors are not expected to deliberately cede responsibility for gap-filling to a third-party litigator.

3.2.5 Summary: The art of dealing with contractual incompleteness

In this subchapter we introduced and assessed strategies of gap-filling in situations beset by various types of incompleteness: Comprehensive contracting tries to overcome (or bridge) incompleteness by replicating the contracting ideal of the CCC; flexibility provisions (contingency measures or default rules) seek to mimic the outcome of the CCC by efficiently capturing regret contingencies; precaution is geared towards minimizing the incidence of regret and disputes; and delegating gap-filling to courts is the strategy of addressing issues only after they occurred by ordering an impartial third party to deal with them.

Chart 3.7 summarizes these different gap-filling strategies. Our analysis has shown that there are in fact only two strategic trajectories that signatories may choose when

---

183 The existence of judicial imperfection opens the door to opportunistic conduct designed to contrive cancellation, evade performance, or otherwise force a backtracking from existing commitments (Masten 1999, p. 33).

184 Shavell (1980 at footnote 6) notes: “Reliance on [courts...] can act to fill in gaps in contracts only with respect to those readily observable contingencies for which the agreement that parties would have come to can be fairly confidently imputed.”
addressing instances of contractual incompleteness: The first route is that of \textit{striving for completeness} and \textit{fighting incompleteness}; the second route is that of \textit{accepting incompleteness} and \textit{embracing the opportunities it offers}. Signatories may make \textit{tactical} choices along each of these two strategic trajectories: Will they give the incomplete contract additional structure by \textit{supplying detailed and explicit contract language}, or by \textit{laying down general and unspecific rules of conduct}? The strategic and tactical choices now span a $2 \times 2$ matrix:

**Chart 3.7 Overview of gap-filling strategies in incomplete contract situations**

<table>
<thead>
<tr>
<th>Strategic trajectory of gap-filling</th>
<th>Tactical trajectory of gap-filling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Towards completeness (fighting incompleteness)</td>
<td>General contract language</td>
</tr>
<tr>
<td>Preamble</td>
<td>ComCon</td>
</tr>
<tr>
<td>Courts</td>
<td>Diligence</td>
</tr>
<tr>
<td>DR</td>
<td>CM</td>
</tr>
</tbody>
</table>

Source: author

Note: The chart spans two dimensions: Strategy of gap-filling (vertical continuum), and gap-filling tactics (horizontal continuum). Within this matrix space, six gap-filling strategies are plotted: \textit{Preamble} stands for “preamble language”; \textit{ComCon} for “comprehensive contracting”; \textit{DR} for “default rules”; \textit{Courts} for “delegating gap-filling to courts”; \textit{CM} for “contingency measures”. \textit{CCC} is the short form for the contracting ideal of the “Pareto-efficient complete contingent contract”.

The \textit{CCC} as the contracting ideal is placed in the top right corner, next to the strategy of \textit{comprehensive contracting}. Emulating the CCC, \textit{comprehensive contracting} is geared towards prescribing a complete plan of action and leaves no room for \textit{ex post} discretion. \textit{Diligence}, i.e. trying to be as precise and elaborate as possible when drafting a contract, is placed in the same quadrant. \textit{Preamble language}, in the top left panel, aims at lending the contract more substance and structure, yet does so by adding general, unspecific language. Parties are reminded to use their \textit{ex post} discretion in a cooperative and non-opportunistic fashion. Opposite, in the bottom right quadrant, is the strategy of drafting \textit{contingency measures}. Signatories add specific auxiliary entitlements that elaborate the conditions under which and to what extent contracting parties may engage in \textit{ex post} discretion.
discretion. Assigning gap-filling to *courts* is placed in the bottom-left quadrant. It is a
generalist incomplete-contracting approach to deal with incompleteness. However, as we
discussed above, assigning gap-filling competence to courts is a strategy that logically
applies to only a subset of the various types of contractual incompleteness, namely to
those constituting the group we labeled *inadvertent incompleteness*. Also, courts need
guidelines of operation, especially default rules. This is why ordering courts to fill gaps
is placed above that of designing DR. Writing *default rules*, finally, implies embracing
the possibilities that incompleteness entails by prescribing general entitlement protection
rules.\(^\text{185}\)

This examination of strategies of gap-filling is highly schematic and abstract. In real life,
signatories do not necessarily apply one strategy of gap-filling only, but staff their
contracts with various gap-filling techniques, depending on the nature and number of
exchanged entitlements. In doing so, contracting parties design an institutional frame or
governance structure for their contract. Ideally, transactors assess every single
entitlement they traded, try to anticipate their risk exposition to various types of
uncertainty and incompleteness, and shape the most efficient set of gap-filling provisions
possible.

Depending on the nature of the contractual exchange (number of original entitlements,
number of issue areas covered, depth of commitment level), as well as the complexity of
the original contracting situation (number of signatories, duration of the contract,
magnitude of contractual interaction, etc.), the resulting governance structure will be
more or less complicated. A range can be drawn from comprehensive contracts that
meticulously specify all substantive rules of transaction, to complex multi-entitlement
contracts, and to a simple “optimally indefinite contract”.\(^\text{186}\)

---

\(^\text{185}\) Note that in Chart 3.7 we do not include the strategy of relational contracting: Above, we established that the
strategic choice behind relational contracting is in essence the generation of (internal or external) default rules.

\(^\text{186}\) Incomplete contracting taken to an extreme leads to the concept of an “optimally indefinite contract” (Charny
1991; Cohen 1999). An optimally indefinite contract is the exact opposite of a comprehensive contract and need
not be any less efficient a substitute for a CCC (“In a world in which contract formation is costly and
adjudication costless, a perfectly indefinite agreement, rather than comprehensive Arrow-Debreu bargains
becomes the ideal contract”, Masten 1999, p. 33). Contracts of this type consist of only one substantive clause (a
rule of entitlement specifying the objective of the contract and the transaction or exchange involved), one DR of
entitlement protection, and one rule of enforcement. Optimally indefinite contracts exist in situations of flaw-
and costless arbitration procedures, or if the instituted DR leaves no room for opportunism.
3.2.5.1 Finding out the accurate type of incompleteness

Two questions remain for signatories of an incomplete contract to tackle: How can they know \textit{ex ante} which traded entitlement is affected by which type(s) of incompleteness? And what constitutes the best gap-filling strategy/strategies with which to respond?

Ultimately, it is an empirical issue, or at least a matter of experience, to determine which types of incertitude are most prevalent/critical in which sorts of contracts. It is also an empirical matter, to determine which strategy of gap-filling is best suited to remedy a certain type of incompleteness. Empiricism may be an avenue for future research, but we do not pursue this issue any further. We conjecture that a rule of thumb holds: The more numerous, deeper and broader the entitlements traded, the longer the contract stretches out into the future, the more volatile the relevant environment, and the more express language to the contract, the more \textit{unforeseen} contingencies happen. In contracts displaying \textit{foreseeable incompleteness}, signatories know what to expect – they may just not know the probability of the occurrence of events.

Regarding the signatories’ decision about which gap-filling strategy is best suited to remedy which kind of incompleteness, we have shown in this section: In principle, \textit{inadvertent incompleteness} can be addressed (though hardly eradicated) by means of dispensing diligence, craft ample preamble language, writing default rules, and resorting to external courts as gap-filling strategies. \textit{Foreseeable incompleteness} can be addressed by pay-off schedules and information-forcing strategies (comprehensive contracting), contingency measures, and default rules. Chart 3.8 gives an overview of gap-filling strategies.
3 Incomplete contracts

Chart 3.8 Gap-filling strategies broken down by types of incompleteness

Source: author

Notes: This graph is based on Chart 3.3 above. It shows which gap-filling measures (comprehensive contracting, diligence, preamble language, courts, default rules, contingency measures) are suited to address which types of incompleteness.

3.2.5.2 The importance of default rules of flexibility

For the remainder of Chapter 3 we will be analyzing default rules of flexibility. We do so for various reasons:

First, the research interest of this paper is geared towards the study of trade policy flexibility mechanisms. Contingency measures, the alternative flexibility instruments to DR, are actually subsets of default rules: Contingency measures, like default rules, are
entitlement protection provisions (secondary rules of contracting), but contingency measures display an added layer of conditionality, i.e. special enactment thresholds and application scope limitations. Much of the discussion on default rules can also be applied to contingency measures. However, as shown above (subsection 3.2.2.1; cf. also Chart 3.8), the contingency measures as a gap-filling strategy are far less versatile, and prone to serious stumbling blocks.

Second, when it comes to instances of foreseeable incompleteness, the use of default rules is vastly superior to any other measure of gap-filling. Crafting DR is the most versatile gap-filling strategy, since it successfully applies to four of five types of incompleteness that we bundled into the term foreseeable incompleteness.\footnote{DR may be of additional use in situations of ignorance (cf. footnote171).} Instances of efficient, necessary, and inexorable (type A and B) incompleteness are probably the most interesting forms of contractual incompleteness, since outcomes and effects are measurable.

Furthermore, as our discussion of gap-filling techniques revealed, the “strategies” of relational contracting and delegating responsibility to courts are ultimately little more than strategies connected with the design of default rules.

In summary: When incompleteness is unbridgeable, i.e. in instances of foreseeable incompleteness (sans strategic incompleteness), there is no way around studying DR as a fallback rule of flexibility.\footnote{Default rules are less relevant when we study cases of strategic incompleteness. However, we will not consider instances of strategic incompleteness any further. As said above, situations giving rise to this category of incompleteness can be expected to be quite rare (simple, one-shot, single-entitlement contracts in an environment of low volatility) and are thus of little use to the study of the WTO contract. In addition, scholars of IO have already extensively dealt with these issues and introduced ingenious schemes of comprehensive contracting.} As will be shown in Part 2 of the study, those situations feature prominently in the WTO contract. In focusing on the study of DR, we must leave aside a closer examination of inadvertent incompleteness. Interesting as the field of inadvertent contractual gaps and strategies for remedying them may be, we will not pursue this issue any further: At this point of time, no feasible theory is emerging in that area.\footnote{Uncertainty categories that we referred to as ambiguity, ambivalence, and unforeseeability (cf. Chart 3.2) are admittedly difficult to study: There is no methodology in place that could predict when and how signatories conduct contracting errors and lapses. We hope that contract theory will make progress in this area. However, we submit that inadvertent and foreseeable incompleteness are complementary fields of study. The insights that we gather in future sections of this study hence come without a loss of generality, even though we will not consider instances of inadvertent incompleteness.}
3.3 Crafting rules of flexibility: Inalienability, specific performance or liability?

Default rules of flexibility lay down whether and how parties can react to changing circumstances that have not been considered explicitly at the time of the contract formation. In this subsection, we will examine more closely the intricacies of default rule design for incomplete contracts. For reasons that will become clear in the course of our engagement with the WTO, we are especially interested in situations where:

(i) a contract exists between reasonably rational parties;\(^\text{190}\)
(ii) the \textit{ex ante} commitment (level of concessions) is non-trivial and discrete;\(^\text{191}\)
(iii) contracting parties live in a non-stationary (read: dynamic) world, and are faced with a situation of \textit{unforeseen contingencies};\(^\text{192}\)
(iv) enforcement as promised is possible;\(^\text{193}\)
(v) for simplicity, only a single entitlement is exchanged.\(^\text{194}\)

Although this may look like a string of five rather confining assumptions, the majority of real-life contracts would be contained in this assumption set.\(^\text{195}\)

\(^{190}\) For our current purpose we presuppose the existence of a contract. Much of the economic property rights literature and the L&E literature on default rules (the liability-vs.-property rule debate, see \textit{infra}) deals with non- or pre-contractual issues. For example, it is discussed in the literature how property should be protected in general, or whether entitlements in cases like traffic accidents or environmental pollution should be organized by way of liability or property rules (see Krauss 1999 for a literature survey). Although an interesting topic, results of non-contractual issues hardly carry over to contractual situations.

\(^{191}\) The exchanged “good” (better: promise) is not binary, but a matter of degrees; see footnotes 59, and 83 and accompanying text.

\(^{192}\) This is to say that for simplicity – but without loss of generality – we assume away all other types of contractual incertitude and focus on \textit{unforeseen contingencies} (see Chart 3.2 above). The presence of unforeseen contingencies gives rise to \textit{efficient, necessary, inexorable Type A, and inexorable Type B incompleteness} (cf. Chart 3.3). Think of contingencies as some kind of negative exogenous environmental or behavioral shock which is either revealed symmetrically or privately to one contracting party. The (previously unanticipated) shock might hit one or all parties, and will create regret in at least one signatory (whether succumbing to the contingency is welfare-enhancing or welfare-depreciating is a different question).

\(^{193}\) This assumption essentially implies that parties do not commit to contractual concessions that cannot be enforced, either by themselves or by a neutral third-party enforcer. We are not concerned with issues of self-enforcement capacity, and of the stability of the contract.

\(^{194}\) As stated before, every real-life multiple-entitlement contract can logically be unbundled into a series of independent substantive entitlements. So, examining only one substantive entitlement can be understood as analyzing a very simple contract, or one aspect of a wider and more complex agreement.

\(^{195}\) For example, imagine the contract between a tourist and a tour guide, where each of the above assumptions (i)-(v) hold: (i) The two signatories conclude a contract. (ii) Each party goes into advance commitment: The guide blocks his day, the tourist makes a down payment. (iii) Unforeseen contingencies abound; and (iv) enforcement is possible (in case the other party fails to perform, each party can complain to the tourist office). (v) The single-entitlement is the tourist’s right to see the entire city; in return, he owes money to the guide.
When designing efficient default rules, the objective of signatories is to mimic the outcome of the unattainable CCC, but not its substance (replicating its substance would be the prerogative of comprehensive contracting). This naturally means having the same contractual goals as the CCC. Contractual flexibility is designed to exhaust all gains from trade by inducing optimal ex post escape (i.e. non-performance in cases where the CCC would also mandate excuse) and by deterring inefficient, opportunistic ex post behavior. Safeguarding efficient “breach” will provide for full assurance to contracting parties who are thus willing to maximize their upfront commitments, namely the scale and scope of their ex ante concessions (including ex ante reliance investment in the relationship).

Essentially, the objective of a flexibility rule is (i) to seize the potential presented by possible regret contingencies in a way the CCC would mandate it; (ii) to forestall opportunism; and (iii) to safeguard the maximal ex ante commitment level. So far, so good. The difficulty consists in finding a rule of entitlement protection ex ante which is able to set the optimal ex post incentives for both victim and injurer. In other words, a DR must separate situations where non-performance is mutually welfare-enhancing from those contingencies where non-performance is opportunistic and thus results in welfare-depreciation.

In subsection 2.2 we introduced the basic stages of contract design. It was shown that signatories must lay down their primary rules (definition of entitlement), secondary rules (rules of intra-contractual entitlement protection), and tertiary rules (entitlement-enforcement provisions). It is now easy to see that the process for designing contractual flexibility mechanisms is indispensable for defining rules of efficient entitlements.

---

196 The fact that environmental or behavioral contingencies are unanticipated and the contract therefore is incomplete does not mean that parties’ intentions are uncertain: The Paretian logic, namely joint welfare maximization, common to all contracts is and stays the contractual objective. The CCC stays the welfarist benchmark.

197 To recapitulate: Reliance investments are pre-contractual relationship-specific investments, which are (i) partly sunk, and (ii) enhance the efficiency of the contractual exchange.

198 Mahoney (1999, p. 118) rightly contends that the importance of default rules for non-performance is not to be underestimated, since the choice of a DR alters “the incentives facing the breaching party, which will directly affect the probability of performance and indirectly affect the number and type of contracts people make, the level of detail with which they identify their mutual obligations, the allocation of risks between the parties, the amount they invest in anticipation of performance once a contract is made, the precautions they take against the possibility of breach, and the precautions they take against the possibility of a regret contingency.”

199 This objective is exactly at the heart of Ethier’s (2001) “reciprocal conflict problem” (cf. footnote 23): The injurer wants to be able to seize opportunities generated by the vision of non-performance, whereas the victim wants to rest assured that she does not suffer from that escape action. An adequate ex ante governance structure must be in place to set the right incentives to each contracting party.
3 Incomplete contracts

protection (Dunoff and Trachtman 1999, p. 32). Chart 3.9 illustrates that crafting contractual rules of flexibility is a key ingredient of the contract design process.

Chart 3.9 Designing contractual default rules

<table>
<thead>
<tr>
<th>Primary Rules (I)</th>
<th>Secondary Rules (II)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commitment</td>
<td>Flexibility mechanism:</td>
</tr>
<tr>
<td>(level of ex ante concessions)</td>
<td>Default rule of entitlement protection</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Tertiary Rules (III)</td>
<td>(Self-)enforcement instruments</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Objectives of contract design:
I: Maximize ex ante commitments
II: Seize regret contingencies, compensate victims ➔ Flexibility
III: Forestall opportunism ➔ Enforcement

Source: author
Notes: This chart is a modification of Chart 1.2, which introduced the concepts of flexibility, enforcement and ex ante commitment in the trade context. The chart above shows that for incomplete contracts in general that the determination of an efficient rule of DR is indispensable for finding the best enforceable entitlement protection, namely one that gives the right non-performance-incentives to injurers while compensating the victims such that both (or all) signatories are willing to pre-commit to the efficient level of contractual concessions.

Chart 3.9 recapitulates the process of contract design in general as consisting of:

- fixing a substantive entitlement and the level of mutual cooperation concessions (primary rule of entitlement);
- finding a viable entitlement protection rule (flexibility rules, especially an efficient default rule), and agreeing on enforcement rules that safeguard intra-; and
- deter extra-contractual behavior by all signatories (tertiary rules).

As stated in subsection 2.2.2, the literature usually distinguishes three pure types of entitlement protection, and consequently of default rules: If one party is allowed to opt out of the agreement unilaterally (a liability or pay-or-perform rule), the discretion and most likely all the ensuing gains from non-performance reside with the injurer.

---

200 Conceptually, the question of ex post flexibility is the flipside of the level of legal entitlement protection. Flexibility determines the “action space” of contracting parties. In particular, flexibility mechanisms lay down whether and how parties can react to changing circumstances that have not been considered explicitly at the time of the contract formation. So, while flexibility provisions nail down the legitimate behavior of the active party, the entitlement question is concerned with the scope of protection granted to the passive party.
Whenever the potential injurer is under a specific performance obligation (property rule), he must engage in renegotiation in order to buy off the right of non-performance from the victim. Both ex post discretion and gains from non-performance are then a matter of negotiation. In case ex post adjustment behavior is strictly prohibited, all signatories are under an unconditional specific performance obligation (or rule of inalienability).201

When searching for the most efficient substitute for the CCC in a situation of contractual incompleteness, the ultimate question that signatories have to settle effectively is: “Who owns the gap, and at what price?” The level of entitlement protection determines the ex post discretionary action space of contracting parties, and also fixes how ex post gains from non-performance are to be divided between the parties: A liability rule of protection advantages the injurer and (at best) insures the victim as a residual claimant against suffering harm. Under a property rule (PR) of entitlement protection, the victim’s best alternative to no agreement is to insist on the injurer’s performance as promised. Her threat to have enforced the original terms of the agreement makes her “owner” of the gap. Under the PR victims are likely to appropriate a larger share of the ex post gains from contractual flexibility (cf. our discussion in footnote 92 and accompanying text). Under an inalienability rule, no one owns the gap: The ex ante distribution of gains is identical to the ex post distribution, since contractual adjustment is prohibited.202

In the next three subchapters we will focus on four sets of questions concerning the design of default rules:

• First, in which instances of incomplete contracting is an inalienability rule of entitlement protection superior to a DR of ex post discretion? (subsection 3.3.1)
• Second, what determines the rational decision between a liability and property rule of default? (subsection 3.3.2.2)

201 As a side remark: The choice between the three entitlement protection rules (liability, property, inalienability) corresponds to three distinct strategies of dealing with technological risk in the discipline of risk management: Opting for an inalienability rule means choosing a strategy of robustness, defined as “the capacity to sustain performance under external perturbations by maintaining an established internal structure” (Stirling 1999, p. 27). A PR is akin to a preference for adaptability, defined as the “capacity to sustain performance under external perturbations by changing internal structures” (ibid.). A rule of liability, finally, is a choice for flexibility (in the narrow sense), namely “to retain as many options for as long as possible in advance of commitment, and (when commitments are made) to retain an ability to withdraw without great penalty if prohibitive conditions arise” (ibid.).

202 Dunoff and Trachtman (1999, p. 35) confirm: “Thus, default rules […] are not simply neutral background rules designed to facilitate agreements; rather, they have important distributional implications.”
• Third, what are efficient levels of *intra*-contractual remedies accompanying a rule of liability? (subsection 3.3.2.1)
• Finally, what additional modalities should be tied to a choice of a DR? (subsection 3.3.3)

3.3.1 *Inalienability or efficient non-performance?*

Whenever signatories decide on a rule of inalienability, they choose to prohibit any *ex post* flexibility: Despite the occurrence of external or behavioral shocks, the originally established internal contractual structure is to be maintained once and for all. If the contract happens to have a gap, performance as mandated is nevertheless the only permissible action. The contract is renegotiation-proof; it must not be modified.

We saw above that a rule of inalienability is the adequate entitlement protection rule for the CCC, as well as the comprehensive contract. However, inalienability can also be imperative in contracting situations featuring *unforeseen contingencies*, but where the exercise of *ex post* discretion would be globally welfare-depreciating. A general rule of inalienability is apposite whenever *ex post* deviation from obligations is

(i) always inefficient;
(ii) contract-annihilating; or
(iii) immoral, i.e. in contravention with basic external norms.

*(i) Choose inalienability whenever ex post escape is always inefficient.* Shavell (1980) shows that there are instances where rational welfare considerations warrant unconditional specific performance, since *ex post* non-performance of contractual obligations is always welfare-depreciating. It is welfare-optimal for signatories to tie their hands whenever two conditions hold cumulatively: First, the “breach” decision has to be less important than *ex ante* commitment. One or all of the signatories’ *ex ante* commitments and/or pre-contractual relationship-specific reliance investments are fragile, and disproportionally more important (in terms of the sum of expected values) than *ex post* regret contingencies provoked by temporary shocks or long-term trends.

---

203 Whenever the incompleteness is bridgeable, the contract is “near-complete”, for example in instances of *strategic incompleteness*. In such circumstances, a rule of inalienability protects the complete set of entitlement protection rules. We showed above that complete contingent and comprehensive contracts assign a course of action to every possible contingency. There exists a comprehensive plan of action; therefore *ex post* discretion would by definition be welfare-depreciating.
Second, any *ex post* escape action by an injurer is apt to frustrate (or crowd out) the victim’s *ex ante* reliance or up-front commitment.

Jolls (1997), building on the work of Fudenberg and Tirole (1990), discusses instances where contractually sanctioned *ex post* non-performance can be welfare-depreciating, even if both parties find it in the mutual interest (*ex post*) to escape or renegotiate the initial terms of the agreement. The author argues that “breach” should be unconditionally prohibited from the outset, so as to protect initial efficiency-enhancing relationship-specific investments, and to frustrate anticipatory strategic behavior in the form of underinvestment. Indeed, post-contractual non-performance would lead to *ex ante* underinvestment in reliance, and to a significantly lower *ex post* value of the contract.

**(ii) Choose inalienability whenever *ex post* escape is contract-annihilating.** Whenever a unique level of commitment is indispensable for the functioning of the contract, a rule of inalienability is pertinent. This is typically the case in instances where the contract is a binary agreement, or if the exercise of *ex post* discretion can be used to completely and irreversibly abrogate the terms of the accord. Consider the example of international arms control agreements, such as the Strategic Arms Limitations Treaty (SALT), or the Treaty on Anti-Ballistic Missiles (ABM): Anything else than a default rule of inalienability would frustrate or annihilate the treaties’ original terms, indeed the essence of these agreements (Rosendorff and Milner 2001, p. 830). No country would accede to SALT or ABM if they posited that any Member may “temporarily opt out of the Agreement”: “Temporary” contractual escape would be tantamount to canceling the respective treaty altogether.

**(iii) Choose inalienability whenever *ex post* escape is immoral.** Whenever clear, unambiguous peremptory norms of societal conduct are threatened by *ex post* discretion, an absolute protection of the initial entitlement, and the according strict prohibition to take away, or trade, the entitlement is apposite (Calabresi and Melamed 1972, pp. 1111). Human rights are an example of a set of ubiquitously accepted inalienable rights that cannot be taken away or traded: A person cannot sell himself into slavery, or have his freedom of speech taken by someone in return for monetary compensation.205

---

204 The contract knows only two states of the world: Compliance with, or deviation from, the terms of the agreement.

205 Another way of arguing why peremptory rules of social conduct should be protected by inalienability is given by Calabresi and Melamed (1972, pp. 1111). These rules do not lend themselves to objective and non-arbitrary measurement: “This nonmonetizability is characteristic of one category of external costs which, as a practical matter, seems frequently to lead us to rules of inalienability. Such external costs are often called moralisms.”
Transposed to the international realm, default rules of inalienability are pertinent whenever peremptory norms of international law (ius cogens) are in danger (Pauwelyn 2006). In addition to ubiquitous inalienable rights, every society or group has a set of generally accepted norms and values it perceives as untouchable.

To sum up: A rule of inalienability is in the signatories’ best interest and indeed an equilibrium outcome of contractual negotiations, whenever the level or substance of the ex ante commitment decision is sufficiently more important than the escape decision. Whenever ex post backtracking from previously made commitments frustrates contractual intent, inalienability is apposite.

3.3.2 Liability or property rule?

In most contractual settings, ex post discretion to temporarily and partially withdraw from previously made concessions is superior to “locking in” for good the initial terms of the agreement. In response to unforeseen contingencies or unanticipated party behavior, post-contractual non-performance can be mutually welfare-enhancing (in terms of expected discounted values), if the present DR implements the right set of incentives and constraints. It is a heatedly debated topic in contract theory, whether, and under which circumstances, a liability or a property rule of entitlement protection yields the better outcomes.206

Before we can go on to examine when and under what circumstances an LR or rather a PR makes for the more suitable default rule in an incomplete contract of the kind assumed in this subsection, we must find out what type of liability rule is to be pitted against a PR. As we said above, any LR of flexibility must be accompanied by an indemnity provision, or intra-contractual remedy payable to the victim (also called

206 In contract theory the controversy “liability vs. property rule” features prominently (see Kaplow and Shavell 1996b at note 1; Dunoff and Trachtman 1999; Krauss 1999 for literature reviews). The debate has circled around three different contexts: First, in the debate of pre-transactional allocation of property rights, where scholars are concerned, how – in the face of serious transaction costs – property rights shall be assigned, and how to design efficient rules on the “taking of property” (injunction). In economics this literature features under the rubrics of “residual rights assignment”, “comparative institutions” and “boundary of the firm” (see e.g. Coase 1937; Grossman and Hart 1986; Hart and Moore 1988; Williamson 1979). The second context of the liability-vs.-property debate is in the literature on rules of prescriptive jurisdiction: L&E scholars discuss the design of court-ordered rules of default, where parties to a contract have failed to specify their DR ex ante (e.g. Ayres and Gertner 1989, 1992; Craswell 1999; Johnston 1990). Third, and more to the point, scholars have broached the issue of liability and property provisions as flexibility rules in contracts. Two aspects are of main concern: The effect that contractual rules of breach and remedy have on relationship-specific investments in private contracts (the hold-up issue; see footnote 118 above), and the ease of use and administration of
compensation, or damage payments).\textsuperscript{207} We must thus examine which remedy provision signatories will rationally design so as to complement a rule of liability.

3.3.2.1 \textit{Which remedy best complements a liability rule?}

\textit{Intra}-contractual remedies can be located on a continuum that ranges from zero to infinite (cf. Chart 2.3 above). Most common are the \textit{zero}-, the \textit{restitution}-, the \textit{reliance}-, and the \textit{expectation} damage measures (Mahoney 1999, pp.121):\textsuperscript{208}

The \textit{expectation} damage measure aims to put the victim of contractual escape in as good a position as if the injurer had performed. It is equivalent to the replacement value that exactly makes the victim indifferent between the injurer’s performance and his default – the remedy insures her against any dynamics that unfold \textit{ex post}.\textsuperscript{209}

The \textit{reliance} measure of indemnity is intended to restore the \textit{status quo ante the \textquoteleft\textquoteleft breach\textquoteright\textquoteright} for the victim. The difference between expectation damages and reliance is that reliance damages compensate the victim for direct harm suffered, but leave aside indirect effects and foregone opportunities (such as transactional efficiency gains that would have accrued in the case of normal contractual performance).\textsuperscript{210}

The compensation measure of \textit{restitution} damages restores the \textit{status quo ante the contract}. An injurer must reestablish the Nash-level that persisted before the contract in the non-cooperative past (this may or may not correspond to the zero-cooperation

\textsuperscript{207} The reader is reminded that under a PR of entitlement protection, injurers and victims negotiate over the size of the damage payable to the victim.

\textsuperscript{208} Signatories also often opt for the \textit{liquidated damages} measure, which is a contractually fixed (\textit{ex ante} negotiated) indemnity. A liquidated damages clause usually refers to a certain sum, or action, that falls due whenever one signatory deviates from its promised behavior (Masten 1999, p. 27). While liquidated damages are an interesting concept, it is difficult to generalize their impact and outcome, simply because their design and magnitude are unique to each contractual setting.

\textsuperscript{209} Expectation damages are essentially what is called “tit-for-tat” or “commensurate damages” in game theory. Mavroidis (2000 at p. 800) argues that expectation damages are strictly higher than \textit{damnum emergens} (direct harm suffered). Expectation damages must be interpreted as \textit{lucrum cessans}: \textit{All} further efficiency costs (opportunity costs or losses in value-added) caused by the partial breach of the agreement over and above direct effects must be indemnified (see also Schropp 2005 at footnote 10 and accompanying text).

\textsuperscript{210} To be sure, it is these contractual efficiency gains (transaction cost efficiency, risk transfer, transaction efficiency) that motivated transactors to conclude the contract in the first place (see footnote 52 and accompanying text). One conceivable way of “paying” reliance damages is to seize a contested measure and return to fully cooperative behavior.
3 Incomplete contracts

It is evident that restitution damages are smaller than reliance, whereas reliance damages are strictly smaller than remedies amounting to a victim’s expectancy.

What is the appropriate amount of indemnity that reasonably rational negotiators will agree on at the outset of a contract? How can the victim best be compensated for her incurred losses? Mahoney (1999, p. 120) perceptively points out that the choice of intra-contractual remedy largely follows the same efficiency criteria as does the choice for an efficient rule of default. A particular damage measure can be termed efficient, if it creates an incentive for signatories to take the identical decisions and concessions they would have taken under the Pareto-efficient CCC. It is generally accepted that the preferred measure of damages is the expectation measure, since it induces injurers to refrain from performance only in situations where it is globally efficient to do so. At the same time it does not make the victim any worse off after contractual default occurred.

To illustrate why expectation damages usually Pareto-dominate all other rules of remedy that can potentially accompany a liability rule of flexibility, let us consider a simple yet generalist model of a contract: Suppose a single-transaction, one-shot contract between two symmetrical, risk-neutral, signatories. The agreement is thereby a perfectly indefinite contract that contains a description of substantive entitlements to be exchanged, and a single default rule of liability. The contract notably contains no provisions for contingencies (cf. footnote 186). The possibility of renegotiation is not considered. In accordance with our above assumptions (footnotes 190-194 and accompanying text), we consider reasonably rational players, discrete commitment levels by both parties, optimal enforcement, and the presence of foreseeable incompleteness.

Now define:

\[ c \]

as level of cooperation, agreed to ex ante by each player. \( c \) can be perceived abstractly as “willingness to cooperate”. Assume a vector of some type of cooperation areas (production, trade liberalization, service, etc.). It is a non-

---

211 Often restitution damages mandate the injurer to pay back any pre-contractual reliance investments made by the victim.

212 This model is used for exposition only. It is an abridged, adapted, and simplified version of the models used in Schropp (2007) and Mahlstein and Schropp (2007), which are based on the intuition of Shavell (1980, pp. 473).

213 Stipulating a one-off contract is for expositional convenience only. Alternatively, we could assume a repeated-interaction contract, where players are uncertain if, when and how the next interaction will take place.

214 For the purpose at hand, it is inconsequential whether the probability density function of contingency-shocks is known to players or not. Equally, it is of no consequence as to whether the contingencies are revealed symmetrically or privately to the players. It also does not matter whether or not victim and injurer know their respective roles as victim and injurer ex ante.
negative set and will be endogenously determined, whereby the most reluctant liberalizer sets the common level of cooperation;

\( \theta \) as an unforeseen (previously un-specified) contingency that hits exactly one player. \( \theta \) is a scalar that can be interpreted as an exogenous shock of some magnitude. The player affected by the shock will experience regret, and consequently takes the role of the injurer;

\( B \) as “breach” set, that is, all those revelations of \( \theta \) that induce the injurer to escape his contractual obligations and to decide for intra-contractual non-performance; \( \{ \theta \mid \text{the contract will not be performed} \} \);

\( W_C(c, \theta) \) as value enjoyed by the injurer if he gets hit by the exogenous shock, but performs his contractual obligations as promised (Cooperation payoff);

\( W_D(c, \theta) \) as value enjoyed by the injurer if he gets hit by the exogenous shock and escapes from certain contract obligations (Defection payoff);

\( V_C(c) \) as value enjoyed by the victim if the injurer performs his obligations as promised (Cooperation payoff);

\( V_D(c) \) as value enjoyed by the victim if the injurer escapes from certain contract obligations (“Sucker’s payoff”, or “Scrap value” of the contract);

\( V_N(c) \) as value enjoyed by the victim in the pre-contractual non-cooperative past (“Nash payoff”).\(^{215} \)

The game plays out as follows: Both signatories conclude a contract by fixing symmetrical ex ante investments/commitments and a rule of post-contractual escape, backed by a certain remedy. An unanticipated contingency occurs which hits one party that thus assumes the role of the injurer. Thereupon, the injurer decides whether to engage in non-performance (“breach”), or perform as promised. Anticipating the injurer’s “breach” behavior and the damage payable to the victim, both parties shape their ex ante commitments accordingly.

Consider the first-best contract that signatories could possibly conclude: The Pareto-efficient complete contingent contract (CCC) is defined as the one maximizing the sum

\(^{215} \) We note that \( W_D(c, \theta) > W_C(c, \theta) \), and \( V_C(c) > V_D(c) > V_N(c) \).
of the expected values of the contract to the injurer and the victim. This implies that there will be an excuse to perform in a contingency, if, and only if, that situation will raise the sum of the values enjoyed by both injurer and victim. More precisely, the Pareto-efficient “breach” set, $B_{opt}^C(c)$, is given, where

$$B_{opt}^C(c) = \{ \theta \mid V^C + W^D \geq V^C + W^C \} .$$

(1)

Rearranging the terms, equation (1) yields

$$B_{opt}^C(c) = \{ \theta \mid W^D - W^C \geq V^C - V^C \} .$$

(1*)

In order to obtain the optimal “breach” set, the injurer’s regret contingency (the anticipated exogenous shock) must be large enough that his regret ($W^D - W^C$) outweighs the harm done to the victim ($V^C - V^C$). The CCC strictly mandates performance, and grants non-performances, in situations where it is mutually welfare-enhancing to do so. Since the CCC is the first-best contract, it also induces both parties to concede to optimal ex ante commitment levels ($c^*$).

Let us now leave the contracting ideal of the CCC behind, and proceed to consider the optimally indefinite incomplete contract backed by a liability rule of flexibility. A game of complete uncertainty over the future unfolds. We define in addition to the above variables:

- $d$ as damage measure (for simplicity, a monetary compensation payable by the injurer to the victim);
- $W^C(c, \theta)$ as value enjoyed by the injurer, if he gets hit by the exogenous shock, but performs the contract as promised;
- $W^D(c, \theta) - d$ as value enjoyed by the injurer, given “breach” and subsequent damage payment;
- $V^C(c)$ as value enjoyed by the victim, if the injurer performs his obligations as promised;
- $V^D(c) + d$ as value enjoyed by the victim, given “breach” by the injurer and subsequent damage payment.

To recapitulate (see also Chapter 2.4 supra): The CCC is the ideal contracting outcome assumed to result if players are endlessly rational, and there exist no transaction costs. Signatories thus anticipate any conceivable environmental (or behavioral) circumstance or contingency ($\theta$), and tailor the contract optimally to any conceivable realization of $\theta$. 

---

216 To recapitulate (see also Chapter 2.4 supra): The CCC is the ideal contracting outcome assumed to result if players are endlessly rational, and there exist no transaction costs. Signatories thus anticipate any conceivable environmental (or behavioral) circumstance or contingency ($\theta$), and tailor the contract optimally to any conceivable realization of $\theta$. 

- 120 -
Under an LR system of default, the injurer’s participation constraint is
\[ W^D - d \geq W^C . \]  
(2)

Equivalently, we can note
\[ B(c) = \{ \theta \mid W^D - W^C \geq d \} . \]  
(2*)

The injurer’s participation constraint basically states that the damage payment, or indemnity payable for non-performance, must be strictly smaller than the size of the injurer’s efficiency gains from non-compliance; otherwise the injurer is not willing to utilize his ex post discretion and refrain from “breaching” his obligations.

The victim’s participation constraint, next, is given by
\[ d \geq V^C - V^S , \]  
(3)

which essentially says that a reasonably rational victim must get compensated for her losses suffered from the injurer’s non-performance. Otherwise the contract is a loss-making enterprise for her, whenever the injurer escapes from his contractual obligations. Consequently, the victim will not agree to sign onto the agreement if (3) does not hold.

Inserting the victim’s participation constraints (3) in equation (2*) above yields
\[ B(c) = \{ \theta \mid W^D - W^C \geq V^C - V^S \} = B^{opt}(c) , \]  
(4)

which is equivalent to equation (1), the optimal “breach” set. In other words, in order to satisfy both signatories’ participation constraints, any feasible incomplete contract between two reasonably rational parties must feature the same (optimal) “breach” set as the CCC; otherwise the contract will be rejected ex ante by either one of the signatories.

Let us now compare different damage remedies \( d \) that this incomplete contract can feature:

Under the expectation measure \( d^{exp} = V^C - V^S \);

under the reliance measure \( d^{rel} = V^N - V^S \);
under the restitution measure \( d_{\text{res}} = VN - V^S \);

under the zero-damage measure \( d^0 = 0 \).

It is now easy to see that only the expectation damage measure yields a “breach” set which mimics the optimal “breach” set that the CCC provides for. Inserting the remedy variants \( d_{\text{exp}}, d_{\text{rel}}, d_{\text{res}} \) and \( d^0 \) into equation (2*), we can see the:

\[
\begin{align*}
B(c)^{\text{exp}} &= \{ \theta \mid WD - WC \geq VC - V^S \} = B^{\text{opt}}(c), \\
B(c)^{\text{rel}} &= \{ \theta \mid WD - WC \geq VN - V^S \} \supset B^{\text{opt}}(c), \\
B(c)^{\text{res}} &= \{ \theta \mid WD - WC \geq VN - V^S \} \supset B^{\text{opt}}(c), \\
B(c)^0 &= \{ \theta \mid WD - WC \geq 0 \} \supset B^{\text{opt}}(c).
\end{align*}
\]

Any remedy short of expectation damages violates the victim’s participation constraint, given by equation (3). In addition, equations (5*) to (5***) show that anything but expectation damages induces the injurer to escape more often than is optimal (the resulting “breach” sets are larger than the optimal “breach” set \( B^{\text{opt}}(c) \)). Faced with reliance-, restitution-, or zero-damages, the injurer engages in over-“breach” – he defaults inefficiently often, compared to the optimal “breach” set \( B^{\text{opt}}(c) \). This is equivalent to saying that the injurer engages in non-performance at the occurrence of shocks of a lower magnitude than the CCC would allow for default to happen. He takes the occurrence of relatively mild shocks to defect from the agreed-upon rules of the game. Hence, the injurer commits an opportunistic “breach” in situations where the CCC would mandate strict performance.

Absent an efficient “breach” set, the victim (as most reluctant transactor) will be unwilling to commit to the optimal level of ex ante concessions, \( c^* \). Cognizant of the presence of some sort of shocks, she will anticipate over-“breach” by the injurer. Thereby, any efficiency loss from escaping the original contractual obligations will be shouldered by the victim, and she will not have a stake in any efficiency gain seized from non-performance either. Excessive escape behavior on the part of the injurer will leave her in a worse state than had he performed. Factoring into her calculation this

---

218 Reliance damages reestablish the status quo ante the breach, the restitution remedy reestablishes the status quo ante the contract. In our example of a one-shot contract without reliance investments, both damages are equivalent.

219 The corresponding “breach” sets \( B(c)^{\text{rel}}, B(c)^{\text{res}} \) and \( B(c)^0 \) form supersets of \( B(c)^{\text{opt}} \). In other words, smaller revelations of \( \theta \) induce the injurer to escape his contractual obligations than the CCC would prescribe.

220 Shavell (1980, p. 483) proves formally that any damage measure that is to induce the optimal level of ex ante commitment must safeguard an efficient “breach” set (“If there were a damage measure which always induced Pareto-efficient behavior, then in particular it would have to induce Pareto-efficient breach”).
opportunistic escape behavior will bring down her ex ante willingness to cooperate (her commitment level $c$); in fact, she will be hesitant to enter the contract at all.

Summing up the insights from these formal illustrations: Only expectation damages possess the necessary incentive compatible characteristics that effectuate efficient non-performance by the injurer on the one hand, and optimal ex ante commitment by all signatories on the other. Behind a veil of ignorance, commensurate remedies in the form of the victim’s expectancy emerge endogenously.

To conclude, in this subsection we have illustrated (rather than proven formally) why “there does not exist a damage measure which is always Pareto superior to the expectation measure” (Shavell 1980, p. 483, emphasis in original) – at least in one-shot contracts where the need for temporary and partial escape is as important as, or more important than, the preceding commitment decision. An LR of entitlement protection is best complemented by an expectation damage measure of indemnity. Most other commonly used compensation instruments short of expectation remedies encourage opportunistic over-“breach” on the part of the injurer. This not only brings down the mutual welfare level (hence constitutes Pareto-inferior behavior), but also discourages the victim(s) of such escapist behavior and induces them to engage in suboptimal ex ante cooperation. This decreases the value of the contract for both players.

3.3.2.2 Property rule or liability rule? A question of transaction costs

Now that we have shown that a liability rule of default is best accompanied by the expectation damage remedy, we can proceed with our initial endeavor, namely to compare the respective suitability of LR and PR as efficient rules of default.

The choice of the efficient DR depends on post-contractual transaction costs: Those TC that unfold during the contract-performance phase – after the conclusion of the contract – determine whether an LR or a PR is more apt to address contractual incompleteness. In subsection 3.1.1 we distinguished between two groups of ex post TC: Coasian (exogenous, or objective) TC, and endogenous (behavioral, or subjective) TC.

---

221 In a world absent of ex post imperfections, the choice of default rule is inconsequential: Given that all post-contractual actions from bargaining to calculating damages are costless, the resulting resource allocation will always be Pareto-optimal. This is another application of the Coase Theorem (see Kaplow and Shavell 1996b). Contrary, this means that owing to the presence of ex post TC the choice of flexibility DR does make a difference to the outcome, and consequently to signatories’ choice of ex ante cooperation.
There are two groups of *exogenous ex post TC*: measurement costs of damages, and renegotiation costs. Each renders one of the two DR options more problematic. An LR is encumbered by calculation/measurement costs of damages, while renegotiation costs eschew a PR regime of flexibility:

A *liability rule* of entitlement protection backed by an expectation damage remedy requires the presence of an external arbitrator to calculate the harm done to the victim, and consequently to assess the appropriate size of remedies. This entails considerable transaction costs for all signatories: First, the injurer must engage in information gathering and processing costs in order to estimate the damage that his planned act of non-performance will cause to the victim. He must do so in order to assess whether opting out and paying expectation damages will be worth his time. Second, further TC are generated by the creation and maintenance of an impartial and capable legal authority which is to act as arbitrator. This is not only a costly endeavor, but also a collaborative expense prone to collective action problems. Third, appealing to an arbitrator may bring about litigation costs for both injurer and victim (preparation costs, costs of using legal advice, opportunity costs of time invested in litigation, etc.). Fourth, as a precondition for an LR to work, parties must agree that interpersonal comparison of utility is possible: The victim’s subjective value should be “monetizable”, and the victim must be willing to subordinate the subjective valuation of her incurred damage to the judgment of an impartial outsider (see Dunoff and Trachtman 1999, p. 31; Krauss 1999). A monetization of damages depends strongly on the underlying contractual exchange: Where commitments and contractual efficiencies are easily measurable (money invested, profit made, tariff-lines decreased, etc.), posting a price tag on the damage suffered is substantially easier than in cases where commitments are multi-faceted, emotional, or profoundly subjective (e.g. the value of a painting, the value of a limb). Fifth, serious exogenous TC ensue if the arbitrator commits systematic errors: Whenever arbitrators chronically under- or over-estimate the damage, injurers will either engage in opportunistic breach, or refrain from defaulting in cases where it would be welfare-enhancing to do so, respectively.223

---

222 We are not suggesting that compensation must always be paid in monetary units. Indemnity can take many forms: seizure of the measure in question, additional concessions by the injurer, an offer of consolation, etc. Our point here is simply that damage and compensation must somehow be inter-subjectively comparable in kind.

223 Interestingly, Kaplow and Shavell (1996a) find that an independent arbitrator need not be omniscient or operate flawlessly. As long as its judgment is not *systemically biased*, the arbitrator’s verdict will be acceptable to parties (see also Rosendorff 2005).
Under a *property rule* of default, parties can do without an arbitrator. Instead, a PR of entitlement protection relies on the instrument of renegotiation of the contract. This brings with it a different set of exogenous *ex post* TC: First, although an objective interpersonal comparison of utility is less of a problem in a negotiation between an injurer and victim, the parties nevertheless have to strike an *implicit* deal. In particular, parties have to reach terms as to the size of expectation damages, since this is the victim’s fallback option of negotiation, her best alternative to a negotiated agreement. Hence, the injurer has to incur significant information gathering and processing costs, and costs of verifying the victim’s expected harm resulting from his intended act of non-performance. Second, the victim also has to invest resources in identifying the potential efficiency gains from non-performance (of which she wants her share). This means that she has to assess the nature of the regret contingency, as well as its expected effect on the injurer’s well-being. This requires serious fact-finding efforts on the part of the victim. Third, in contrast to an LR, where the injurer is free to opt-out and immediately seizes his regret contingencies, bargaining induces time costs in terms of opportunities foregone: Whenever an injurer is hit by a significant shock, every minute lapsed without reacting to it (in the form of temporary withdrawal from previous commitments) will incur costs. If renegotiation drags on, efficiency gains from non-performance are lost. At the margin, renegotiations can persist until non-performance has become meaningless, and the prospective gains therefrom are lost (e.g. because the crisis has subsided). Fourth, the number of signatories and the complexity of the contractual entitlements are another factor to reckon with when considering the option of renegotiation: The larger the group of victims, the more heterogeneous their preferences and the more diverse and versatile the underlying contractual deal, the more difficult, complex, long-winded, and conflicting will be the ensuing negotiations over contractual escape.

This comparison of *exogenous ex post TC* gives a first indication of when and why it is better to apply an LR or a PR of flexibility. Yet there is a second group of transaction costs which must be closely considered. *Endogenous ex post TC* are costs induced by strategic gamesmanship on the part of one or more signatories. We distinguish two kinds of situations: Those in which *symmetrical* revelation of previously unforeseen

---

224 Rogerson (1984 at footnote 6) remarks: “Even though the agents can estimate the other’s cost or value, it may still be impossible or very costly to specify the contingencies in objectively verifiable terms suitable for legal enforcement.”
3 Incomplete contracts

contingencies induces parties to act strategically, and those that emerge from asymmetrical, or private, knowledge. Only the injurer’s regret contingency may be private knowledge (the contingency or its effect is privately observable by the injurer). Alternatively, or additionally, the harm caused to the victim can be private knowledge to her.

Under symmetrical revelation of previously unforeseen contingencies, alternative, or additionally, the harm caused to the victim can be private knowledge to her.

Under symmetrical revelation of previously unforeseen contingencies, two types of strategic behavior may occur:

1. Crowding out. Under symmetrical information and a PR mandating renegotiation, the victim can effectively blackmail the injurer. Without any contractually fixed sharing rule or a distinct renegotiation procedure, the victim as a reasonably rational actor will try to appropriate all efficiency gains of non-compliance through renegotiations. She will make efforts to leave the injurer exactly indifferent between performance and default. The victim thereby crowds out the injurer’s regret. A PR without a structured renegotiation process can then become structurally equivalent to a variant of the game in which no ex post discretion is permitted. This outcome of essentially prohibiting ex post discretion is not intended by signatories (otherwise they would have opted for an inalienability rule from the outset, and not for a PR of flexibility).

2. Hold-out. Even if there is some contractual sharing rule, or some sort of fixed renegotiation process (e.g. in the form of a timeline) in place, the victim under a PR nevertheless has an incentive to hold out her counterpart (cf. footnote 118 above). Knowing that time is precious for the injurer, the victim may engage in opportunistic procrastination (“foot-dragging”) with the aim of influencing the ex post distribution of the resulting non-performance gains in her favor (cf. Fearon 1998). This may result in either hold-out welfare losses and a “blackmail-premium” for the victim, or even

225 As pointed out before (see text around footnote 92), a rational victim has no interest whatsoever in a renegotiation outcome that does not at least put her in as good a position as if the injurer had performed (which exactly corresponds to the expectation damage remedy). She can threaten to cancel renegotiations and instead insist on the injurer’s specific performance obligation.

226 Symmetrical revelation of previously unanticipated contingencies gives rise to incompleteness of the efficient and inexorable-type A incompleteness variants (confer Chart 3.3 above).

227 The victim can be expected to do so, because she sees only upside potential in engaging in blackmailing, yet virtually no downside. Under a PR she has the bargaining leverage of threatening to exit the renegotiations and to insist on the injurer’s contract performance as promised at any point in time.
lead to an outcome in which the victim appropriates all the *ex post* gains from non-compliance.\textsuperscript{228}

3. **Overinvestment.** Under an LR backed by expectation damages, and in a situation of symmetrically revealed information, the victim may be tempted to engage in overzealous (and hence sub-optimal) *ex ante* levels of concession, or in excessive reliance investments.\textsuperscript{229} Under the expectation damage remedy the victim has nothing to lose from the injurer’s default. Yet she also has nothing to gain from contractual non-performance, since she does not enjoy any of the *ex post* gains from non-performance. Hence, the victim is not concerned by issues of non-performance altogether, and instead sees *ex ante* concessions as an investment with a contractually guaranteed rate of return (expectation damages act as “compensation floor”). This leads her to exceed the Pareto-efficient *ex ante* commitment level: The victim behaves as if her *ex ante* investment (in the form of commitment level or reliance) was not risky at all. This causes her to neglect the fact that even under the CCC *ex ante* commitment will not always produce efficiencies in instances of non-performance (Shavell 1980, p. 478).\textsuperscript{230}

In summary, we see that both non-performance regimes show weaknesses when it comes to withering strategic gamesmanship on the part of the victim in situations of *symmetrical* revelation of contingencies. The general verdict in the literature is divided: Whereas Rogerson (1984) and Edlin and Reichelstein (1996) see a clear superiority of the PR in situations of unilateral and bilateral reliance investments, Sykes (1991, Appendix A) and Ethier (2001a) show that an LR complemented by expectation damages will equally lead to efficient outcomes.\textsuperscript{231}

\textsuperscript{228} Aghion et al. (1990; 1994) discuss model situations, where one signatory appropriates all renegotiation surplus despite some (exogenous or contractually negotiated) renegotiation process (see also Chung 1991; Nöldeke and Schmidt 1995). Edlin and Reichelstein (1996) develop a renegotiation bargaining protocol that allows for “sharing rules”.

\textsuperscript{229} Note that the concern of underinvestment caused by the notorious *hold-up* behavior (see footnote 118 above) is *not* of issue under an expectation damage remedy. As we saw in the last subsection, overzealous “breach” and underinvestment occur when victims are not sufficiently compensated for their losses. This is the case for the reliance-, restitution-, and zero damage measures, but not for the remedy of expectancy.

\textsuperscript{230} “Because expectation damages ‘insure’ the [victim] against all breaches of the other party, [her] private return to reliance exceeds the joint return. As a consequence, levels of reliance are set excessively high” (Rogerson 1984, p. 40). Thus, overinvestment is a contract-theoretical version of the “tragedy of the commons” (cf. Hardin 1968).

\textsuperscript{231} The discrepancy in outcome between Rogerson and Edlin/Reichelstein, and Ethier and Sykes can be attributed to a difference in modeling: Ethier and Sykes assume that parties negotiate behind a veil of ignorance, that is, signatories do not know whether they will turn out to be injurers or victims later on. Rogerson and
Things change, once the contexts of asymmetrical revelation of contingencies (a situation that we previously termed necessary- and inexorable-type B incompleteness) are examined. Under asymmetrical information the following strategic party behavior problems may be observed:

1. **Overinvestment.** Under a regime of LR-cum-expectation damages, overzealous investment by the victim party may also occur in games of asymmetrical revelation of contingencies (Shavell 1980): The victim, ignorant of (and indifferent to) the nature and size of the injurer’s regret contingencies, takes its ex ante promises as an investment with a certain rate of return. This leads to inefficient over-commitment.

2. **Hold-out.** In situations where the victim possesses private knowledge of the damage incurred against her, hold-out behavior can occur under a PR of entitlement protection. The victim has a strong incentive to overstate her incurred damage, and to drag on renegotiation talks. Ideally, she will expropriate the injurer of all the (commonly known) gains from non-compliance.

3. **Reverse hold-out.** Whenever the revelation of a previously unspecified contingency or its effect is private knowledge to the injurer, he has an incentive to misconstrue reality, if the flexibility regime is a PR. The gains from non-performance are thus the injurer’s proprietary knowledge. Under the specific performance requirement this induces the injurer to make believe that the exogenous shock (his regret contingency) was small and insignificant, and that the efficiency gains from non-performance are minor. The victim has all the reasons in the world to mistrust his assertions. Instead of accepting the injurer’s compensation offer, she may make a counter-offer which apportions the (unknown) gains from non-performance substantially more to her favor. A sequential bargaining model unfolds, in which the injurer, caught in between wasting time and losing his bargaining leverage to the victim, tries to maximize his payoffs by signaling his true intentions.

4. **War of attrition or double-sided hold-out.** In situations where both the regret contingency to the injurer and the damage incurred by the victim are private

---

Edlin/Reichelstein, on the other hand, assume predefined, fixed roles. Note that none of these models consider the presence of any exogenous ex post transaction costs.

232 The victim possesses proprietary information whenever the continuation value of cooperation or her “sucker’s payoff” (value of the contract after escape has occurred) is private knowledge to her.
knowledge to the respective parties, renegotiations prescribed by a PR of default are even more difficult. Both parties try to misconstrue their real conditions in the renegotiation talks. Bargaining in situations of two-sided incomplete information has been termed a “war of attrition”, because each party tries to hold out against its opponent. In this case, solution concepts are very difficult to come by, indeed.\footnote{Grossman and Perry (1986) have described such a game, and found a novel solution concept called “perfect sequential equilibrium”. The mathematics involved is notoriously difficult. For our present purposes it suffices to note that signatories incur substantial opportunity costs in due course of a lengthy bargaining procedure.} For the purpose at hand it suffices to say that wars of attrition incur severe efficiency losses (time costs) and opportunity costs of bargaining.

Whenever asymmetrical information dominates the contractual relationship \textit{ex post}, a property rule of default loses its appeal: Entering into renegotiation triggers strategic hold-out behavior on the part of all signatories. A liability rule of default is superior in these situations. Under an LR the injurer opts out whenever he feels the gains from doing so exceed the expected compensation remedies payable to the victim. He may do so without being held up by a rent-seeking victim. \textit{Ex post facto}, an impartial arbitration instance calculates the damage awards to the best of its knowledge and ability. Recent contributions in the realm of WTO scholarship confirm the general superiority of the LR in situations of private information (Rosendorff and Milner 2001; Rosendorff 2005 for the case of victim’s private knowledge of damage; and Herzing 2005, chapter 3; Mahlstein and Schropp 2007 for the case of private revelation of the contingency to the injurer).\footnote{See Ordover and Rubinstein (1986), Osborne (1985), or Henricks and Wilson (1985) for overviews of the literature on wars of attrition in bargaining with asymmetrical information.}

Some conclusions emerge: It is ultimately a question of context as to whether an LR-\textit{cum-expectation-damage} or a PR of renegotiation is more appropriate to protect a certain entitlement in a situation of incertitude. Transaction costs that occur \textit{after} the conclusion of the contract and during the contract performance phase determine which of the two options is better suited to the circumstances at hand. We stressed the fact that not only Coasian, \textit{exogenous}, TC are an issue, but notably also \textit{endogenous} TC, caused by

\footnote{In contractual situations featuring asymmetrical revelation of the contingencies \textit{and} bilateral \textit{ex ante} commitment, an expectation-damage-backed LR neither leads to opportunistic breach (moral hazard) on part of the injurer, nor to excessive \textit{ex ante} commitment on the part of the victim – if the damage calculation by the arbitrator is not systematically biased. Opportunistic breach is ruled out, because the victim receives compensation amounting to her expectancy. Excessive overinvestment on the part of the victim is ruled out, because the injurer will want to fix the mutual \textit{ex ante} commitments on a lower level than the victim (Schropp 2007, Appendix). Being the more reluctant liberalizer of the two, the injurer manages to set the reciprocal commitments on the \textit{ex ante} efficient level. We shall return to this important finding in Chapter 7.}
strategic party behavior. We showed that in situations of *asymmetrical* information about contingencies and/or their effect on signatories’ utility, property rules lose their competitive edge. In such circumstances, renegotiations are costly and prone to strategic brinkmanship by both the injurer and victim.

### 3.3.3 Additional modalities of default rule design

When drafting a proper gap-filling strategy, the contracting parties’ decision to take one of the three DR options (LR, PR, or rule of inalienability) is only one, albeit important, question to consider. Signatories have to be mindful of additional questions on DR modalities:

- **First, there is the issue of enforcement:** Parties have to define the *extra*-contractual remedies that best complement the default rule at hand. It is evident that a rule of inalienability must be backed by coercive (or infinite) penalties (see Chart 2.3 above and accompanying text). Signatories must be discouraged to ever deviate from the terms of the default rule. A property default rule that mandates specific (but renegotiable) performance must also be protected with very high punishments. Else, some injurers would prefer to defect from the rules of the game, and subsequently sustain the subsequent penalty. How to protect a liability rule of default that is complemented by an expectation damage remedy? It is important that the *extra*-contractual remedy be more costly for the injurer than expectation damages. If so, a reasonably rational injurer will always prefer to adhere to the *intra*-contractual escape strategy instead of *violation-cum-punishments*.

- **A second issue of DR design is that of divided entitlement and level of conditionality:** Should the enactment of a DR be tied to any preconditions? Is it advisable to limit the DR’s scope of application? That is, should the entitlement protection be co-owned by both sides, injurer and victim in a situation of incompleteness? Probably not. The whole point of a default rule of flexibility is to act as a fallback for an unconsidered occurrence. How can signatories attach a conditionality to something they are insufficiently aware of *ex ante*? Attaching preconditions to a DR necessarily means limiting its scope. But large scope and far range are exactly the point of a DR.

---

236 If a PR were not backed by very high *extra*-contractual remedies, it would turn into a *de facto* LR of escape and thereby thwart the default rule’s initial purpose.
In addition, a DR tied to conditionality is conceptually similar to a contingency measure. And as stated above (subsection 3.2.2.1), contingency measures, though valuable, have a number of serious drawbacks. Thus, a rule of default based on divided entitlement is predestined for disputes. An accurate rule of thumb is to keep both the underlying entitlement and the rule of default as simple and clear-cut as possible.

- A third issue is that of divided entitlement protection: Does a hybrid rule between a *liability-cum-expectation damage* and pure property rule make sense? Should contractors *ex ante* craft remedies that are systematically lower or higher than expectation damages? The answer depends on the idiosyncrasies of the contracting context, as well as on the prevalent type(s) of incompleteness. In situations of *strategic incompleteness*, certain information-forcing rules of default probably make sense which systematically over-compensate the less-informed victim (cf. Ayres and Gertner 1989). However, under the set of assumptions used in this subchapter (see footnotes 190-194 and accompanying text), especially in the presence of unforeseen, unspecified and unanticipated contingencies, everything but an LR-cum-expectation-damage or a pure rule of renegotiation was shown to be an incomplete substitute for the first-best CCC.

- Fourth, do multiple rules of default make sense? Absolutely not. A DR is a fallback rule that allows for *ex post* discretion whenever an unforeseen or unspecified contingency occurs. A DR becomes effective in the absence of any explicit agreement to the contrary. Having two rival DR in place to protect the same entitlement is an arrangement that is bound to create havoc and disagreement.

The discussion in the previous two paragraphs leads us to a final consideration, namely that of the relationship and interplay between contingency measures and default rules. As explained in section 3.2.2, both measures are flexibility tools apt to remedy contractual incompleteness. Contingency measures are a special case of default rules and can be used as complementary tools of gap-filling. It may well be that a contractual entitlement is protected by several contingency measures and a single DR at the same time. However, two preconditions of an efficient system of multiple flexibility mechanisms

---

237 A high enactment threshold always limits the application scope and brings with it the issues of sunk costs. Furthermore, contingency measures do not work properly in situations of asymmetrical revelation of contingencies. Writing contingency measures also bears the risk of provoking contracting errors due to imprecise and ambiguous language.
must be heeded by transactors. First, signatories must effectively single out and unambiguously circumscribe all those bundles of outcomes they want to see protected by provisions other than the DR. Second, a clear hierarchy of flexibility rules must be kept in mind: DR should trump any contingency measure, which means that unless contingency measures unambiguously apply, a DR comes into effect whenever a previously unspecified contingency occurs.

3.4 A framework for theorizing about incomplete contracts: Revisiting the concept of the efficient “breach” contract

Chapter 2 gave a general introduction to the nature of contracts. Chapter 3 has so far dealt with the reasons for, and types of, contractual incompleteness; ways of addressing andremedying incompleteness were also reviewed. This section wraps up our discussion on incomplete contracts in general. Two distinct objectives shall be pursued:

First, we want to piece together the findings gathered so far in order to develop a framework of theorizing about incomplete contracts (subsection 3.4.1). This framework, on the one hand, shall guide us through the contract-theoretical examination of the WTO in the next chapter. On the other hand, it may lead to a better understanding of the nature of incomplete contracts in general, and to successful theorizing about ways of improving the governance structures of specific incomplete agreements.

The second objective pursued in subsection 3.4.2 is to revisit the concept of the efficient “breach” contract (EBC). The concept is important yet misunderstood in the contract literature. We want to identify what an EBC really is, and what it is not. The EBC shall be introduced as the second-best, the achievable first-best contractual governance structure that reasonably rational parties wish to conclude in the presence of incompleteness. The EBC serves as the normative yardstick for all incomplete contracts.

3.4.1 A framework for theorizing about incomplete contracts

It seems obvious to state that before reforming a contract one has to understand its nature. This requirement is occasionally overlooked, not least by WTO scholarship (which will be the focus of discussion as of the next chapter). Researchers regularly theorize about the quality of contract outcomes, yet fail to demonstrate an understanding of the underlying contractual constraints and systemic trade-offs. In the following
subsection we want to introduce a general framework that may help us to theorize about any incomplete contract. 238

We propose a three-phase procedure: The first phase is geared towards analyzing a given incomplete contract and understanding its nature. The second records relevant constraints and trade-offs, and assesses the existing contractual governance structure. The final phase introduces guidelines of reform by conducting a hypothetical bargain analysis.

3.4.1.1 Phase 1: Analyzing the nature and context of the contract

In order to grasp the true nature of any incomplete contract under examination, we divide the first phase of analysis into the following five steps of examination:

1. Signatories and contractual intent. The logical first step is an assessment of the identity of the signatories and their contractual intent. “Who is in charge of negotiating and signing the contract?” “What are the contracting parties’ utility functions to be maximized?” 239

2. Basic entitlements and level of commitments. The contractual intent is informed by the most important entitlements that signatories exchange (the substantive entitlements). Yet as was noted in the last chapter, a contract is usually comprised of more entitlements than just those specifying the contractual intent.

Complex contracts (bundles of entitlement) ought to be broken down into their constituent parts, the single entitlements. Relevant issues to tackle are: “What are the basic entitlements exchanged, and what is the level of cooperation the parties individually agreed to?” “Who owns these entitlements; under what circumstances does entitlement switch?” Some entitlements are appropriated to the signatories in a very detailed fashion (“if contingency x occurs, the Player A shall do y and Player B shall engage in z; if contingency xx happens, the Player A shall do yy and Player B shall engage in zz”); some entitlements are subject to exceptions; other entitlements are rather crude (“A owes B 250 $ payable in three weeks time at the latest”).

---

238 This proposed approach is sufficiently general to tackle contracts of various form, shape, and level of complexity. Yet it is detailed enough to generate added value and to motivate exact reform proposals.

239 In Chapter 2 we noted that contracts necessarily solve a problem. Therefore, finding the rationale for contracting, i.e. the very problem that the particular contract at hand is meant to overcome, is an important aspect of the analysis.
After having identified the most important entitlements, the researcher may decide to focus on the examination of one or several of them. S/he should thereby make explicit which entitlement(s) s/he focuses on and for what reasons.

Most entitlements are not of a binary, but rather of a discrete nature. It is therefore vital to get a notion of the relevant levels of commitment, and the readiness of parties to pre-commit to mutual cooperation. Commitment may be granted in the form of mutual concessions, up-front assurance, *ex ante* reliance investment, etc. Important aspects of commitment are the scale (depth) and scope (breadth) of cooperation. Signatories can decide to cooperate a little in one single field or to engage a lot in various fields of cooperation. The level of *ex ante* concessions, or the scale and scope of cooperation, is the essence of the contract. It defines the efficiency gains to be had from the contract.  

3. **Upfront constraints, prevailing types of incompleteness, and scope of regret.**

Above we introduced various types of contractual incompleteness resulting from different underlying categories of incertitude in the contractual environment. Which types of incertitude are present, acute, and relevant to the contractual relationship? No contract is complete, and every contract is exposed to all sorts of incertitude. However, cognizant of the contractual intent, the entitlement under scrutiny, and the relevant context, contract scholars must be able to separate important types of incompleteness from those that are not. Only those kinds of incertitude are of issue that produce significant regret contingencies or which open the door to substantial opportunistic action.

The quest for incompleteness should be carried out separately for every substantive entitlement under examination. Different entitlements fall prey to different types of incompleteness. Also, external or behavioral shocks differ from context to context. The “scope for regret” must be described.

4. **Existing governance structure.** After having examined the nature and the circumstances surrounding the contractual deal, it is pertinent to assess the existing governance structure which signatories have crafted. How do signatories intend to deal with contractual incompleteness: Do they condone *ex post* discretion or do they

---

240 “Did Party *A* commit all his efforts in return for Party *B*’s promise, or did he remain cautious and vigilant?”; “Did the buyer engage in the welfare-optimal pre-contractual reliance investments, or did she under-invest anticipating a later hold-up problem by the seller?”.
categorically prohibit such actions? The secondary rules of contracting are thereby vital.

This includes a discussion of “dependant auxiliary entitlements”: Many contracts feature additional auxiliary rights and obligations geared towards curbing opportunistic abuse of the system. Hence, the interested scholar needs to look for auxiliary entitlements, at which substantive entitlement they are tied to, and with what level of detail they are circumscribed. Special attendance must be directed to the choice of flexibility mechanisms. The design of contingency measures, as we stated above, forms a decisive part of contractual flexibility, as does the design of default rules.

5. Enforcement. A final step is to analyze how rules of entitlement and entitlement protection are enforced against extra-contractual misdemeanor. Do rules of enforcement make sense given the preceding rules of entitlement protection? How can contract infringement – if discovered – be punished, and what kinds of (self-) punishment mechanisms are in place?

3.4.1.2 Phase 2: Mapping post-contractual constraints; assessing trade-offs

The second phase of our proposed incomplete-contract analysis is to assess whether the existing contractual order fulfills the expectations of the signatories in terms of effectiveness and efficiency. Building upon the previous examination of actual contractual intent, context and institutional design of the contract, the researcher should strive to understand the relevant constraints and trade-offs unfolding during the performance of the contract. Loopholes and inconsistencies in the existing contractual regime have to be identified: Where are there efficiency losses, where is room for opportunism, where can we expect instances of ex post strategic gamesmanship? The costs of contractual flaws have to be assessed in terms of welfare lost and efficiencies foregone through inefficient levels of ex ante commitments. Chart 3.10 may prove helpful for mapping out and understanding contractual constraints and trade-offs:

241 In other words, the researcher has to understand the “game” that plays out in the contract.
3 Incomplete contracts

Chart 3.10 A stylized version of a contracting game under incompleteness

---

Source: author, based on Schropp (2007, Figure 1)

Notes: The chart represents a contracting game including signatories’ contract design options and possible ex post transaction costs which may unfold during the contract performance phase. The process has six stages. PR stands for “property rule”, LR stands for “liability rule”, TC for “transaction costs”.

Chart 3.10 divides a contracting game into six stages. In stage 1, two (or more) parties decide to cooperate in the form of a contract. The accord is necessarily an incomplete contract. In the initial terms of the agreement the signatories determine the entitlements they wish to exchange (but not yet the level of commitment), as well as the instruments of entitlement protection. Contracting parties negotiate over appropriate mechanisms of dealing with, and overcoming, contractual incompleteness, as well as enforcement provisions. As we found earlier in this chapter (see Chart 3.7), signatories have two strategic trajectories: Towards completeness, and towards incompleteness.

Depending on the previously agreed-upon governance structure, parties in stage 2 negotiate the level of mutual commitments. For example, country signatories come to terms with the level of reciprocal trade liberalization commitments. After the level of cooperation is fixed, the performance phase of the contract begins with stage 3. Countries start interacting according to the terms of the contract. At some point in time ($t_3$), whatever ex ante incertitude there is in the external environment is resolved. A

---

$^{242}$ Stages 1 and 2 take place concomitantly (but not simultaneously) in real life.
shock hits one or more parties and may create room for regret. Shocks can be exogenous (technical, economic, political, social, even personal) or behavioral (party A does something that triggers regret in party B). Shocks can unfold locally, regionally, nationally or globally.

In stage 4, depending on the terms of the initial agreement, an affected injurer then decides whether to react to this domestic shock by delivering as the contract stipulates. Alternatively, the injurer may default from previously agreed levels of performance. Intra-contractual remedies are enacted as prescribed by the agreement in stage 5: Should \textit{opt-out} (LR) be the stipulated flexibility tool, damages are paid to the victim; should entitlement be protected by a specific performance rule (PR), the injurer must barter with the victim to buy off her right to demand performance as prescribed. If the victim claims to have detected a violation of the agreement, if the injurer refuses to pay the remedies as previously agreed to, or if an injurer infringes upon an entitlement protected by an inalienability rule, dispute settlement and enforcement commence (stage 6).

In a perfect world this game would have a trivial outcome. In real life, parties are faced with many an imponderability at the time of the contract formation. When drafting their contractual governance structure, signatories’ anticipation of the future is obstructed by a veil of ignorance. In addition, various kinds of \textit{ex post} transaction costs may unfold in the contract performance phase. Uncertainty about future roles, the presence of various types of contractual incompleteness and post-contractual TC constitute trade-offs and constraints that signatories must take into serious consideration when designing the contract. Anticipating them shapes the contract’s governance structure at the time of contract formation. This is not an easy task and one prone to myopia, error, and misjudgment.

The examining researcher has to find out whether the actual choice of signatories’ governance structures is a good one or if it is rather apt to create behavioral loopholes

\footnotesize{\textsuperscript{243} The injurer is guided by his allegiance to the contract, by the fact that the shock was insignificant, by a rule of inalienability, or by the fact that expected compensation payments outweigh his gains from non-performance.\textsuperscript{244} He may be guided by opportunistic impulse coupled with a legal loophole, or by the contractual possibilities presented by the contractual flexibility rules.\textsuperscript{245} The concept of the “veil of ignorance” is important in long-term, repeated-interaction contracts. It implies that players do not know the future distribution of gains and losses from the initial agreement with certainty (Rawls 1971). This it to say that negotiations over the terms of the agreement between prospective Members take place in ignorance as to the identity of future signatories, ignorance of their role as injurers or victims, and generally about how future states of the world are going to impact on signatories’ well-being.}
and/or avoidable inefficiencies. If the contract indeed displays design flaws, s/he should assess their influence on the mutual ex ante commitments.

3.4.1.3 Phase 3: Hypothetical bargain analysis

Based on the findings gathered in the previous two phases of analysis, the researcher may finally proceed to make reform suggestions. Staffed with a firm understanding of the nature and context of the contract, as well as of the underlying constraints and trade-offs, s/he is now in a position to suggest better ways as to how – given the constraints and trade-offs inherent in the nature of the contract – specific entitlements, entitlement protection and enforcement instruments should be organized so as to yield more efficient outcomes.

The researcher should construct a hypothetical benchmark contract. Utilizing the same inputs as the real contract (same entitlements, same types of incompleteness, same timeline, same number of signatories, same transaction costs), the benchmark accord will suggest a novel governance structure apt to produce a more efficient output. In fact, the researcher constructs a model contract of the feasible, or achievable, first-best. S/he must thereby lay out in detail which clauses, provisions, or obligations need to be changed, how they should change, and what the expected effect of such a bundle of measures will be.

3.4.2 The efficient “breach” contract as the achievable first-best contract

What does the achievable first-best contract look like, and how is contractual flexibility best organized? Finding the contractual governance structure that best organizes various contractual flexibility mechanisms means looking for the efficient “breach” contract (EBC). As a matter of axiomatics: If the CCC is the unattainable Pareto-efficient complete contingent contract, then the EBC is the one – attainable – governance structure for incomplete contracts that best mimics the outcome of the CCC.

The concept of the EBC is quite often misunderstood, partly because of its ambiguous terminology, partly because the L&E literature applies the concept in a theoretical, abstract and often a-contextual manner. In the following paragraphs and as a conclusion

246 Are potential injurers allowed to escape from contractual performance only when it is efficient to do so? Are injurers induced to engage in efficient, over- or under-“breach” of their obligations? Do victims get satisfactorily compensated in case of escape? Are the enforcement mechanisms in place and working?
to our general introduction to incomplete contracting, we would like to set this record straight. In light of findings thus far we suggest putting the notion of the EBC into conceptual perspective.

- **Efficient “breach” is no breach.** As mentioned before, the terminology breach is somewhat unfortunate. This is so for at least three reasons. First, if parties ex ante agree on the permissibility of ex post non-performance under certain well-defined circumstances, this arrangement forms an integral part of the contract. Since this provision constitutes an intra-contractual arrangement, it is conceptually distinct from sanctionable defection from the contract. It is not a breach of contractual obligations. Second, the term efficient breach does not resolve the more vital question of how to organize non-performance efficiently – as a liability rule or a property rule, or a mix thereof. The third reason why the terminology breach is misleading is slightly more subtle: An efficient “breach” clause is geared towards replicating the CCC’s outcome. However, the CCC only knows inefficient breach. It defines an efficient set of performance and excuse obligations, which are optimal – ex ante, as well as ex post. Non-performance of the CCC provisions must by definition be inefficient. So, it is the CCC that logically defines breaching (extra-contractual) behavior, and not party behavior in an incomplete contract. Using the term breach is unfortunate when what is really meant is non-performance as would be prescribed by the CCC.

We acknowledge that the term breach as a generic category for intra-contractual escape is problematic and may be misleading. However, it is useful shorthand.

- **Efficient “breach” is not just about opting out of the agreement.** It seems that for many economists the concept of “efficient breach” is tantamount to that of a safety valve that allows for ready (uncompensated) opt-out: A party experiencing regret contingencies just withdraws from its previously made obligations; as a reaction, the victim might engage in vigilante justice and retaliate. This understanding of the concept is wrong. First, efficient “breach” can be achieved just as well under a

---

247 As was pointed out supra (footnote 93), efficient ex post adjustment to changing circumstances and unforeseen contingencies is possible under both a liability rule and a property rule.

248 Terms like efficient ex post default, efficient adjustment to contingencies, or efficient release from previously mis-specified obligations would definitely be more correct in the message they convey, but sound more circumstantial and less catchy. In fairness, the word breach does have linguistic advantages over terms like excuse or release: The connotation that breach bears is that of unforeseen/unexpected contingencies, and of ex post flexibility. Excuse or release, on the other hand, are terms more apt to the context of the CCC, where due to perfect anticipation all efficient action is ex ante predetermined, foreseen, and explicit. Parties ex ante define certain non-performance patterns of behavior (excuse), and prohibit any ex post discretion.
property-rule regime of renegotiation. Second, simple opt-out may constitute a breach, albeit not an efficient one: To constitute efficient “breach”, a flexibility mechanism must necessarily be accompanied by a compensation provision. Usually (depending on the present context and constraints) an expectation damage rule of remedy achieves efficient “breach” under an LR (see subchapter 3.3.2.1); alternatively, negotiated remedies under PR can achieve the same result. An uncompensated opt-out possibility is akin to a property rule for the injurer – logically a very different animal than a liability rule of entitlement protection (see Chart 2.3 and accompanying text).

- **An efficient “breach” contract is not an optimally indefinite contract.** L&E literature of efficient “breach” often suggests that *ex post* discretion will naturally result in efficient outcomes (cf. e.g. Bello 1996; Schwartz and Sykes 2002b for the WTO contract). Saying it this way means elevating efficient “breach” (usually understood as a general liability rule of entitlement protection) to the overarching principle. In other words, the concept of an EBC is degraded to a contract, where all gaps that may surface in the contract performance phase are to be filled by the same liability rule granted to the injurer (Friedman 1989). This notion is of an entirely different caliber: A contract of this kind is exactly what we referred to above (footnote 186 and main text) as an “optimally indefinite agreement”, namely one in which all entitlements are protected by the same unconditional liability rule of default.

This is a strong assumption, and one that we by no means wish to maintain. As we showed before, the optimal choice of gap-filling strategies depends on the nature of each traded entitlement, the contingencies affecting this entitlement, and on the *ex post* transaction costs involved. For a pure, across-the-board liability-rule system of gap-filling to be effective, *ex post* transaction costs related to all possible entitlements and contingencies would have to be such that a simple LR is the negotiation equilibrium. This is quite a rigid corset, and one that is not likely to work in many contexts. Realistically, an optimally indefinite contract may only be operable in extremely simple contractual situations with a very limited number of contingencies (e.g. the sale of harvested and stored crop, the lease of a good).

A maximally indefinite contract is *not ipso facto* an optimally indefinite contract: In more complicated contractual situations that involve various volatile entitlements, a
general liability rule of default may well be unworkable. Some L&E scholars (implicitly) contend that the concept of an EBC equals an optimally incomplete contract. This is a dangerous oversimplification of contracting reality. An optimally indefinite contract – just like a comprehensive contract at the other end of the completeness spectrum – is but one governance structure for organizing contracts. It is fair to say that for most real-life contracts the efficient governance solution hardly lies on the extremes, but resides somewhere in the middle: Most contracts are multi-entitlement accords and thus best protected by a mix of different default rules of flexibility.

- An efficient “breach” contract is not necessarily an efficient contract. Contracting parties seek to negotiate the best possible substitute to the CCC – at its best one that is equally efficient. The epithet efficient thereby refers not to the quality of the “breach”, but to the nature and quality of the contract in general. A contract may prompt efficient ex post non-performance decisions on the part of injurers, yet may well display inefficiencies compared to the CCC benchmark. It may prompt suboptimal mutual ex ante commitments or ex post welfare losses due to post-contractual transaction costs.

- An efficient “breach” contract is the ideal of an incomplete-contract governance structure. The L&E literature is cognizant of contracting parties’ inability to achieve the ideal of the CCC in a non-stationary world. In situations where ex post discretion can be welfare-enhancing, an efficient “breach” contract is the optimal substitute for the CCC. As argued in the previous paragraphs, this is not to say that every entitlement of a specific contract should be subject to ex post discretion. Despite contractual incompleteness, many entitlements are best protected by an inalienability rule.

An EBC is one that strictly fulfills two criteria: First, ex post non-performance in every entitlement is identical to what the CCC would have mandated. Secondly (and this is a point that has received little attention in the literature of default rules), the extent to which signatories commit to ex ante concessions (size and number of

249 Imagine for instance an employment contract. Suppose the employer is granted an unmitigated liability rule of non-performance. He could breach any of the employees’ rights by simply compensating them, including their basic human rights, their freedom of speech, their entitlement to have leisure and holidays, their entitlement to work breaks, their entitlement to form trade unions, etc. It is unlikely that an employee will ex ante give her consent to a contractual arrangement like this.
promises) must be welfare maximal – as prescribed by the CCC. In short, an efficient contract requires efficient “breach” *ex post*, and efficient reliance (commitment) *ex ante* (Craswell 1999, pp. 18). Efficient non-performance (opt-out or renegotiation) is a necessary but not sufficient condition for the contract to be efficient.

To conclude, the concept of efficient “breach” is an outflow of the maximization logic in a situation of contractual incompleteness. An efficient *intra*-contractual rule of non-performance aims at instilling mutually efficient *ex post* behavior irrespectively of what has (or rather: has not) been prescribed by the original contract. An efficient “breach” *clause* forms part of the contractual agreement. In a situation of contractual incompleteness, it constitutes a negotiated meta-rule for *ex post*-discretion that supersedes previously agreed substantive commitments.

However, the efficient “breach” principle, geared towards remedying the incompleteness with respect to a single entitlement, risks being misinterpreted as a *pars pro toto*. An efficient “breach” *clause* is not an invitation to opt out of all legal rights and obligations. In fact, an efficient “breach” *contract* is one in which each entitlement is protected by that (set of) rule(s) which generates the most efficient outcome and safeguards the optimal pre-contractual commitments. It is the perfect governance structure that can be crafted in the presence of insurmountable contractual incompleteness.

3.5 A first step towards a general theory of disputes?

In lieu of a conclusion, this chapter closes with some deliberations on disputes in incomplete contracts. Mavroidis (2006) argues that a theory of disputes is derived from the “primary law” (entitlements) that signatories have given themselves, from the protection of entitlements that is agreed upon, and from the degree of contractual completeness. This chapter has dealt extensively with exactly those issues. We have introduced a typology of incompleteness and distinguished between the two groups of *inadvertent* and *foreseeable* incompleteness. We have also presented several ways of bridging contractual incompleteness. The analysis led to an elaboration of the concept of the efficient “breach” contract, which is geared towards minimizing opportunism and

---

250 As Rogerson (1984, p. 40) points out, Shavell (1980) was the first author to claim that the choice of default rule will affect the *ex ante* incentives to “rely”, as well as the incentives to default *ex post*. Shavell was also the first author to establish that a comparison of the efficiency of various damage measures requires both factors considered.
seizing regret contingencies. To what extent can this chapter’s insights contribute to a theory of disputes?

Contract disputes can never be a negotiation equilibrium between rational signatories. They entail costs and per se create little added value. Their occurrence is an inevitable by-product of contractual incompleteness.²⁵¹

Interesting insights can be generated if we examine what triggers contractual disputes: A contractual disagreement may occur because there is general ambiguity in the contract. Due to textual ambiguity, incoherencies, contracting lapses and omissions, or due to the emergence of contingencies of unanticipated nature, magnitude, and effect, parties feel an objective need for clarification of contractual provisions. Parties are genuinely unsure whether the measure in question, or the occurred contingency, is within the limits of the contract, or constitutes an infringement. We shall call those instances “good-faith clashes”. Disputes can of course also be caused by “bad-faith clashes”: One contracting party, driven by self-interest, engages in opportunism, either intentionally or acting with gross negligence. This signatory deliberately “games” the system in search of private rents.

What complicates drawing a clear dividing line between good- and bad faith clashes is not the presence of incertitude per se, but deficiently drafted contracts, or contracting errors. Generally, four types of contracting errors can be distinguished:

(i) Ambiguous and ambivalent language;
(ii) Insufficient language;
(iii) Rigidity;²⁵²
(iv) Suboptimal remedies.

Ambiguous/ambivalent language is an outflow of accidental incompleteness resulting from poorly described contingencies and their outcomes (see Chart 3.3 above). As a consequence, legal loopholes may open up, or the need for clarification may become pressing. Contracts may also be flawed due to insufficient contract language, which means that the agreement fails to address gaps, either by explicit language or by flexibility rules. As a result, victims and/or injurers can engage in opportunistic behavior.

²⁵¹ A CCC is not expected to feature lengthy disputes. Being Pareto-efficient, any action in contravention of the letter of the contract is necessarily opportunistic and deserves uncompromising and instant punishment. No extensive litigation, interpretation, or gap-filling are required to solve the case.
Rigidity occurs whenever the contract demands performance in situations where non-performance would be welfare-enhancing. Finally, suboptimal intra- and extra-contractual remedies either cause injurers to “breach” excessively often, or deny them the opportunity of doing so when it would in fact be welfare-enhancing.

Contractual disputes emerge, because it is difficult to evaluate ex post whether a signatory’s behavior results from ignorance, opportunism, or a drive to create added value. Indeed, it is extremely difficult to tell good-faith from bad-faith behavior.

What can be learnt about contract disputes from this chapter? Our contribution may be considered a step towards a theory of disputes. Although we did not address issues of inadvertent incompleteness, and hence are unable to say much about disputes emerging as a consequence thereof, this chapter’s findings can help assess the quality of contracting when it comes to foreseeable incompleteness. Taking the EBC as a benchmark contract against which signatory behavior can be assessed, we are able to evaluate whether or not a contract is well-prepared to deal with instances of foreseeable incompleteness.

In a world absent of inadvertent incompleteness the EBC is free of disputes. Good-faith behavior will always be intra-contractual, while bad-faith behavior must necessarily be extra-contractual, illegitimate behavior. In an “EBC world” with inadvertent incompleteness, all disputes can be expected to emerge from instances of ambiguity, ambivalence, and unforeseeable contingencies. Signatories must find ways of dealing with these disputes, perhaps by establishing a competent and impartial court.

To conclude: Our findings are not a panacea. We have not found a way to deal with instances of inadvertent incompleteness. Yet we have come quite a way in showing how signatories to a contract can efficiently deal with situations of foreseeable incompleteness. We contend that contracting parties could free themselves from a number of disputes if they addressed their existing governance structure more seriously vis-à-vis issues of foreseeable incompleteness. Many contracting lapses caused by insufficient language, rigidity, and suboptimal remedies could be sidestepped if

---

252 Insufficient language was labeled Type-II errors in section 3.2.2.1, and rigidity as Type-I errors (cf. footnote 165 and accompanying text).
contracting parties tackled with rigor those types of incertitude that cause *foreseeable incompleteness*. We showed in this chapter how signatories can effectively do so by crafting efficient rules of flexibility, and especially of default.
Part 2

Theorizing about the WTO as an incomplete contract

Towards the end of Part 1 we laid out a framework for analyzing incomplete contracts. Following this guideline, we will examine the World Trade Organization (as we could any multilateral trade agreement for that matter) from a contractual perspective. This will show that the WTO is best understood as a relational contract, characterized by complexity, longevity, repeated interaction, and by incompleteness. The WTO has outgrown the narrow confines of the GATT 1947, and is now a bundle of agreements covering a wide array of issue areas, including services, intellectual property, government procurement and technical barriers to trade.

In order to comprehend and assess trade policy flexibility in the WTO, Chapter 4 will examine the nature of the WTO as an incomplete contract: We will moot the identity and preferences of WTO signatories, and will discuss the incentive(s) to conclude a multilateral trade agreement. This will be followed by a presentation of the basic entitlements, and the nature of mutual commitments exchanged in the contract. We then proceed to investigate the intricacies of contractual interaction, and the contingencies (or shocks) that might occur during the performance phase of the WTO contract. This will allow us to sketch the types of incertitude the trade agreement is beset with. Finally, we shall assess the scope for regret caused by the inevitable incompleteness of the contract.

Whereas Chapter 4 is concerned with basic entitlements, the primary rules of contracting, Chapter 5 looks at the contemporary regime of entitlement protection (the secondary rules of contracting) and of enforcement (tertiary rules of contracting). We will examine how well WTO signatories are dealing with the Agreement’s incompleteness by mapping the current governance structure of the contract – its auxiliary entitlements, the existent flexibility mechanisms, and the current rules of enforcement. We will discuss whether the existing system of non-performance makes

253 Some authors go as far as to liken the WTO to a constitution (e.g. Jackson 2004; Petersmann 2002, 1986). Other scholars treat it as a regime or an institution (see references in WTO 2007, section II.B.3). However, as was argued in subsection 2.3.2 (footnote 107 in particular), both constitutions and institutions are in fact contracts of sorts. They adhere to the fundamental contractual logic. It hardly matters what technical definition we give to the WTO, its essence is that of a treaty – the international equivalent of a contract.
sense, and if not, the potential consequences on parties’ *ex ante* commitments, and thus the value of the contract. Special attendance will be given to the protection of the “market access” entitlement, which is the most important contractual commitment exchanged between WTO Members.

Chapter 6 is a parenthesis of sorts. Based on the insights of the two previous chapters, we wish to enter into the academic “compliance-vs.-rebalancing” debate. At face value, this controversy is focused on the object and purpose of WTO *enforcement*. We will demonstrate that this is too narrow a scope. As a matter of fact, the debate spans the *whole nature* of the contract. This necessarily includes a discussion concerning the definition of basic entitlements exchanged in the WTO contract (primary rules), rules of entitlement protection (trade policy flexibility mechanisms; secondary rules), and rules of enforcement and dispute settlement (tertiary rules). The compliance/rebalancing rift, reinterpreted with regard to the *nature of the WTO contract*, will reveal that the two schools of thought actually describe two different contracts: The rebalancing approach to the WTO reduces the contract to reciprocal market access commitments, which are logically protected by a pure liability rule. It is agnostic about *extra*-contractual rules of enforcement, and draws no systemic difference between intra-contractual trade policy flexibility and dispute settlement/enforcement. In the eyes of compliance proponents, the WTO is a multi-entitlement “constitution”, where every entitlement is protected by a property rule of default and enforced by the strong language of the DSU. Yet the compliance view fails to discriminate between different entitlements, and largely ignores the systemic need for trade policy flexibility in the WTO contract.
4 Adding context: The WTO as an incomplete contract

The WTO emanated from the International Trade Organization (ITO, which never came into existence), as well as from the GATT and its subsequent extensions, which were laid down in codes.\textsuperscript{254} The WTO was the outcome of the Uruguay Round (UR) negotiations and came into being on January 1\textsuperscript{st} 1995. It is a single institutional framework organized around various pluri- and multilateral agreements, encompassing the GATT (as modified in the UR), all agreements and arrangements concluded under its auspices in the period between 1947 and 1994, and the complete results of the UR, most notably, the GATS, TRIPS, TPRM, TRIMs and DSU (TPRM stands for “Trade Policy Review Mechanism”). The “Marrakech Agreement to Establish the World Trade Organization”, or “WTO Agreement” in short, is sometimes referred to as the “Charter” or “Constitution” of the treaty. It is a multilateral contract among sovereign States that establishes the WTO as an international organization and defines various committees, bodies, and councils, as well as the duties of and relationships between these groups.\textsuperscript{255}

The WTO is a Member-driven organization with a compact Secretariat. It is mainly a negotiating forum for trade liberalization, a set of legal ground-rules of conduct in international trade, and a venue where signatories can debate and settle their trade disputes. The treaty’s prescribed mission and objectives include freer trade, non-discrimination, competitive markets, rule-of-law, predictability and stability of international trade, and economic development.

In this chapter we wish to thoroughly examine the WTO in detail from a contractual angle. To that end, we will apply the theoretical framework developed in section 3.4.1 to the context of the WTO. As a critical first step, the nature and the context of the contract will be assessed. Subsection 4.1 will single out the relevant contracting parties, their

\textsuperscript{254} We take for granted that the reader is acquainted with the WTO, its origin, genesis, design, and functioning. Elementary primers are Dam (1970), Jackson (1969; 1997a), Hudec (1990), or Hoekman and Kostecki (1995).

\textsuperscript{255} As a convention, we will utilize the terminology “WTO” or “WTO contract/treaty” to encompass the WTO Charter (Marrakech Agreement), including the various agreements mentioned in Annexes 1-4 of the WTO treaty. This terminology also at times used as shorthand for the WTO as an international organization.
preferences and their (likely) contractual intent pursuant to negotiating a multilateral trade agreement.

It is commonly known that the WTO is organized in multi- and plurilateral agreements that regulate the trade in goods, trade in services, trade-related aspects of intellectual property rights, government procurement, dispute settlement, antidumping, and so forth. However, in order to understand the WTO from a contract-theoretic point of view, it is important to shift from an agreement-centric view to one that distinguishes between different contractual entitlements, the primary rules of contracting. Subsection 4.2 will examine the basic entitlements that signatories exchange in the WTO. We will show that the entitlement to reciprocal trade (the “market access entitlement”), in which countries commit to granting mutual market access, exceeds all other entitlements in importance.

Subsection 4.3 will finally assess the nature of contractual interaction and possible contingencies that may occur in a non-stationary world. We will discuss which categories of contractual incertitude are present, acute, and relevant to the contractual relationship, and what types of incompleteness result. Comprehending the contractual incompleteness is a first and important step in understanding and criticizing the Agreement’s system of non-performance – trade policy flexibility mechanisms and enforcement provisions. This task shall be dealt with in the following Chapter 5.

4.1 Players, preferences, and contractual intent

It was the Appellate Body (AB) that stated in one of the early cases that the “WTO is a treaty, the international equivalent of a contract” (cf. footnote 2). The nature of the WTO is that of a relational contract characterized by longevity, open-endedness, repeated interaction, complexity, and incompleteness.²⁵⁶ In this section we will identify the key actors representing WTO contracting parties, their preferences and objectives. We will rely on the results of the political economy literature, a coherent theory of decision-making which has become standard in economic literature today.

²⁵⁶ We take these assertions to be uncontested: Open-endedness implies a long-term setting without clear termination date, and openness to acceding Members. Repeated interaction is a characteristic of a trade agreement in which countries do not interact on a one-off basis, but routinely and permanently engage in the trade of goods and services. Complexity of the contract is a natural result: With 151 Members currently participating (August 2007), multiple issue areas (such as services, goods, investment), many thousands of tariff lines and service commitments in four modes, and a multitude of protective rules and regulations in place, the WTO can rightly be called a complex setting. The criteria of incompleteness, finally, will be subject to a substantial discussion in this chapter.
4 The WTO as an incomplete contract

4.1.1 Players and preferences: Political economy theories of endogenous trade policymaking

The WTO is a contract between sovereign States or Territories. Who really takes the decision to join a multilateral trade agreement on behalf of a country, and who decides the conditions and modalities of the contract? Is it the electoral majority (the median voter) of a democratic country, the parliamentary legislative, a “benevolent dictator” or “social planner” that wishes to maximize general welfare, or is it rather the administration or government of a State that is in charge of conducting the initial trade negotiations? We follow the tradition of the political economy school in economics, and maintain that the WTO is a mutually advantageous political contract among self-interested political officials. While we feel safely backed by contemporary trade literature in that claim, we nevertheless prefer to formulate our opinion about signatories and their preferences as what it essentially is: an assumption.

Assumption: A negotiating party (would-be WTO Member) is represented by a single self-interested policymaker, who is driven by a re-election/re-appointment objective. That objective manifests itself in the policymaker’s strive for political support maximization.

This assumption comprises three elements, namely concerning:

i) the locus of decision-making;

ii) the manner of decision making;

iii) the specification of main actors’ preferences and objective functions.

(i) **Locus of decision-making.** First, we reduce complexity by asserting that there exists only one key decision-maker per country. This individual (or homogenous group of

---

257 Political economy as a discipline has its origins in the public choice literature, with its eminent forefathers Knut Wickell, Patrick Buchanan, Anthony Downs, Gordon Tullock, Mancur Olson, and Kenneth Arrow. It is the economic concept of methodological individualism and rational choice applied to the realm of political decision-making. The political process is understood as a means of achieving reciprocal advantages (rents). Political economy approaches are critical of classical and neo-classical economics: They dismiss the tenet of policy being shaped by high-minded, welfare-maximizing decision-makers (“benevolent dictators” or “social planners”) as naïve thinking. Applying the concept of methodological individualism more stringently, political economists contend that policy is drafted by self-interested government officials who care predominantly about maximizing their personal well-being, be it in the form of political support, re-election endeavors, receiving important information or maximizing their budgetary discretion (Grossman and Helpman 2001, chapter 1).

In the realm of trade policy, political economy approaches are considered the state-of-the-art methodology (see Grossman and Helpman 2001; Rodrik 1995; Staiger 1995 for an overview of the literature). It has become widely accepted in contemporary economic trade literature that members of a trade agreement are represented by rational, self-interested trade policymakers. In fact, it is hard to come up with any contemporary economic model of the WTO that would claim (and model) otherwise.
individuals) is in charge of shaping the domestic trade policy and representing a contracting country’s government in multilateral trade negotiations. This is a significant but reasonable abstraction from reality. Trade deals are usually shaped by institutions, such as DG Trade in the EU or the USTR in the U.S., under the guidance and supervision of their respective administration(s), and after having consulted the respective legislative(s). But what matters here is not how many people are involved in shaping a national trade policy, but whose decision ultimately counts. This is likely to be a small group of like-minded people (“the administration”, “the government”), or even, as we assume, a single person, the “policymaker”. We will use the words policymaker, government, and administration interchangeably.\(^\text{258}\)

(ii) Theory of decision-making. The second element of the above assumption is the approach of methodological individualism as a guideline for decision-making.\(^\text{259}\) All actors, be they policymakers, consumers, or producers, are homines oeconomici who behave self-interestedly, rationally (albeit not always perfectly so) and with reasonable far-sightedness when it comes to important decisions.\(^\text{260}\)

(iii) Actors’ preferences and objective functions. The third part of the assumption specifies the preferences of trade policymakers. Trade policymakers officially act in their capacity of representing a State or Territory. Yet we assume that they are ultimately interested in their own personal well-being. According to political economists there is no inherently logical reason why trade policymakers should be any different from other individuals in their behavior. Concerning their professional life, it is assumed that what

---

\(^{258}\) The alternative to assuming the presence of a single key decision-maker lies in explaining or modeling the domestic decision-making process. This implies inserting an additional layer of strategic decision-making to the trade-policy game. See Grossman and Helpman (2001, chapter 9) for an attempt to give the domestic trade process additional structure. For a general discussion of the right locus of decision-making see Allison (1971).

\(^{259}\) Methodological individualism is the fundamental theory of decision-making permeating all economic thinking (Hausman and McPherson 1996; Kirchgässner 2000, e.g.).

\(^{260}\) There are, of course, alternatives to the assumptions of individualism and rational choice. Non-rationalist, or “constructivist”, approaches to decision-making are norm-driven and collectivist in nature: In a nutshell, constructivists reject the method of rational choice and agent-centered methodological individualism that economic perspectives maintain. They claim that rationalist theories of choice fail on two accounts. Firstly, they neglect the formative influence of ideas, norms and values on behavior. To them, decision-making is much better explained by resorting to fundamental norms, shared ideas, inter-subjective beliefs, traditions, and habits. Perception, interpersonal communication, learning, and socialization to a large degree shape these norms and ideas. Secondly, rational choice theories allegedly fail to acknowledge the influence that the system has on the actors. The power of inter-subjective beliefs, shared understanding, culture and socialization, according to constructivists, is completely overlooked by agent-centric theories of rational choice. Hence, for proponents of constructivism, the dictum “actors’ preferences shape the outcome” is false. Instead, system and agents are strongly interdependent: the structure shapes actors’ perceptions, perception shapes agents’ preferences and
decision-makers want most is to be re-elected or re-appointed. The re-election/re-appointment assumption is broad enough to encompass various secondary preferences, such as policymakers’ drive to “make a difference”, to enter into history books, amass power, collect bribes, become rich, or to maximize a budget: Without being re-elected, however, a policymaker cannot achieve any of the latter goals.\footnote{Although the policymakers’ objective of re-election or re-appointment seems to be tailored to democratic countries, we would contend that this statement is valid for any regime. Take a dictatorship: “A dictatorship is a democracy with a much smaller electorate” (Richard Baldwin in personal communication with the author). Under a dictatorial regime the identity of the body of voters may differ from democratic regimes, and the kind of influence they exercise on the self-interested policymaker may be different. But ultimately, a dictator will have to be “re-elected”, or continuously supported, by the dominant powers of the state. For this reason, we assume re-election and re-appointment to be equivalent concepts.} Therefore, the assumption that policymakers’ preferences culminate in re-election/re-appointment is a solid and well-accepted proxy for their well-being.\footnote{There are alternatives to the re-election objective: Policymakers might exclusively act out of patriotism, a sense of duty, according to standard operating procedures or divine intervention, or out of sheer conviction that what they do is right. However, in the realm of trade-policymaking the re-election assumption is now standard in the literature. Ever since the seminal contributions by Stigler (1971) and Hillman (1982) it is readily accepted that trade policy is shaped by “incumbent politicians who make policy choices while being aware that their decisions may affect their chances for re-election” (Grossman and Helpman 1994, p. 834).}

Re-election chances are maximized if policymakers can rally as much political support from voters and pivotal domestic special interest groups (SIGs) as possible.\footnote{SIGs can be industry associations, labor unions, NGOs, single-issue lobbies (such as the National Rifle Association, or the Royal Society for the Prevention of Cruelty to Animals), or important organized societal groups such as retirees, guerillas, religious factions, or gays & lesbians.} SIGs are important in domestic politics, since they have successfully overcome collective action problems, are usually well-organized, and, importantly, are ready to invest time, effort, and money into influencing the ongoing political processes. Geared by their objective to maximize the welfare of their respective membership,\footnote{The policymakers’ aim at maximizing political support is reflected in the concept of “politically realistic objective functions” (PROF, Baldwin 1987): Governments are assumed to maximize some weighted average of general welfare (a measure that presumably safeguards votes) and SIG-welfare (which upholds political support by special interests). The relative weight that general welfare and SIG welfare assume in each policymaker’s calculation eventually depends on the personality of the respective consequently their behavior. Collective behavior can then have a “feedback” impact on the system (see for example Checkel 1998; Finnemore and Sikkink 2001; Wendt 1999a, 1999b).} SIGs can influence and shape domestic policies, especially if policymakers become reliant on their political support. Political support can take the form of financial contributions, ballot-box power, information provision, or coercive power (Grossman and Helpman 2001, chapter 1.1).
policymaker, the specific context, and, not least, on the micro-foundations (modeling specifications) of each model. Some models exogenously assume the source of policymakers’ political support (“social democrats are supported by the unions”), others assume that SIGs competitively vie for domestic influence in the form of “government auctions” (on that account, see Grossman and Helpman 1994). These assumptions are for modeling convenience only, however. It is certain that the channels of influence are never clear-cut and stable, and that policymakers’ utility functions are likely to be in a dynamic flux, depending on the policy situation, the context, the current political strength of different SIGs, and on the personality of the politician in charge.

Turning to the trade realm, the relevant players are trade policymakers (who are also trade negotiators), consumers/voters, and trade-SIGs. Trade-SIGs may be import-using (downstream) industries, exporters, import-竞争ing sectors (e.g. farmers), foreign exporters, or labor unions, and occasionally also environmental or civil-liberty SIGs. Special interest groups are interested in changing the domestic relative prices so as to maximize the (economic) welfare of their constituents. In order to do so, they are contesting for political access to the trade policymakers.

Trade decision-makers are primarily interested in maximizing their own political welfare. They are driven by a trade-specific PROF, a weighted average of general consumer and particular SIG welfare. Political entrepreneurs are assumed to exchange trade policy decisions (concerning issues of trade liberalization, subsidization, protectionism) in return for special interest group support. They devise trade policies that may redistribute revenues among domestic groups by crafting trade rules which are apt to change the domestic relative prices in favor of SIG sponsors. Every trade policy decision then has an effect on domestic general welfare, but also on the well-being of influential SIGs, and therewith on the decision-maker’s political support.

---

264 This statement, of course, is another simplifying assumption.
265 A policymaker is not necessarily “auctioning” herself off to the highest bidder. Politicians’ personal convictions, intentions, and standpoints can be reflected in the respective objective functions. They enter according to the weight a policymaker attaches to general welfare, specific groups, or by the choice of SIGs that she strives to be supported by.
266 In a Heckscher-Ohlin world sector-specific SIGs are important, whereas under Stolper-Samuelson assumptions factor-specific SIGs (representing labor, land, and capital) are pivotal. In conformity with contemporary literature we shall assume a Heckscher-Ohlin world of trade.
267 Baldwin (1987) was the first to show that a model in which SIGs lobby for protection (or trade liberalization) can equivalently be obtained from a model in which governments give more weight to producers’ than to consumers’ interests. The author demonstrates that policymakers care about the domestic income distribution more than the maximization of total domestic income. They achieve their preferred domestic income distribution by means of tampering with the domestic trade structure.
For our purposes, it is fully satisfactory to demand a *generic* politically realistic government objective function in trade affairs which is geared towards a redistribution of funds. In order to maintain the highest level of generality, we are not concerned with micro-foundations of trade policymaking. We are agnostic when it comes to the specifics of governments’ PROF, and would like to avoid the pitfalls that “endogenous trade models” have grappled with.\(^{268}\)

Taking self-interested individuals/governments as the principle actor of national trade policymaking brings with it a string of important implications, formulated below as corollaries to the above assumption.

**Corollary 1: The State is not a unitary actor.** A political economy view on trade-making is basically a rejection of “statist” perceptions that assume nations are unitary actors with a fixed utility or set of norms and values. Domestic interaction shapes the preferences and norms of policymakers in charge of formulating national trade policy and of negotiating international trade agreements. Assessing trade policy on the sub-state

\(^{268}\) Formal political-economic models of endogenous trade-policymaking and the *structure of protection* supply micro-foundations of the domestic trade-making process. To that end, they specify in more detail governments’ PROF, and SIGs’ channels of influence-taking. This requires researchers to open the “black box” of domestic decision-making, and to supply the necessary formal specifications of trade-policymaking. In that respect Grossman and Helpman’s 1994 *AER*-article “Protection for Sale” must be lauded for presenting the first model to give the lobbying game of trade policymaking a convincing microeconomic foundation. Theirs is generally acknowledged as the standard workhorse model of endogenous tariff policymaking (Rodrik 1995). In their model, Grossman and Helpman assume that trade policymakers maximize a weighted sum of general social welfare and special interest group income (securing political support in the form of financial campaign contributions). SIGs welfare thereby enters policymakers’ PROF with a larger specific weight, signaling their higher importance relative to the general electorate (consumers). Policymakers auction off their domestic trade policy (a vector of trade protection of all industries) to special interest groups. SIGs simultaneously and competitively bid for their preferred trade policy vector. The policymaker maximizes over the competitive bid and directs a unique trade policy accordingly.

Elegant as it may be, the “Protection for Sale” model is neither all-encompassing nor unproblematic in its assumptions, nor uncontested: (i) Trade scholars have criticized the shape of the government objective function (Ethier 2004b; Levy 2003), and have instead proposed different objective functions of political support (such as a conservative social welfare function, cf. Corden 1997; Hillman 1982; Hillman and Ursprung 1994); (ii) The assumption that SIGs take influence via *campaign contributions*, as maintained by Grossman and Helpman, is equally critical: SIGs can exercise influence on policymakers via the provision of market-specific information (Milner 1997; Ethier 2004b; Ethier 2006; Grossman and Helpman 2001, Part II), or via electoral power (e.g. Dixit and Londregan 1986; Mayer 1984; Mayer and Riezman 1990); (iii) The assumption that each SIG lobbies for every product is highly unrealistic (Baldwin and Robert-Nicaud 2006). (iv) Grossman and Helpman assume SIG-creation and their political sway to be *exogenous*; this leaves out the important issue of whether export or import-competing SIGs have more political clout (see Mitra 1999 for an interesting attempt to endogenize SIG creation). (v) Foreign SIGs per assumption are excluded from the domestic lobbying game (Grossman and Helpman 1995b, 1995a). In the same vein, domestic special interests cannot influence foreign policymakers. (vi) Finally, as a general critique of Grossman and Helpman and practically all models of endogenous trade policy, Rodrik (1995, p. 1476) submits two salient points: First, theirs is a model of *tariff* protection and not of *trade* protection per se. Second, the authors do not explain why governments should use such an *inefficient instrument* as tariffs to redistribute funds and to manipulate domestic prices. Goldberg and Maggi’s (1999)
level opens up the “black box” of domestic decision-making, and gives an important role to various non-signatory actors, such as consumers, lobby groups, or NGOs.

**Corollary 2: Self-interest is the driving force behind trade policymaking.** An important implication is that trade agreements are not concluded for the sake of the general welfare. This is a direct rejection of classical economic thinking which conjectures that welfare-oriented social planners, or benevolent dictators, conclude trade agreements to avoid negative economic externalities, or to tie the hands of future generations of policymakers. General welfare considerations may play an important role in trade decision-makers’ trade policy considerations – these considerations just enter the equation indirectly via policymakers’ own utility function.

Much in the same vein, the assumption that self-interested policymakers represent countries in trade negotiations and consequently draft contracts to their liking, is in contrast to some lawyers’ view that the WTO was created first and foremost so as to protect the needs, expectations, and rights of non-contracting parties such as consumers, farmers in LDCs, exporters or NGOs. From the viewpoint of methodological individualism these commentators’ notion of the WTO is noble, but eventually a myopic conception of reality: Following political-economic methodology, trade negotiators a priori have no interest in ameliorating the situation of sub-state factions – as long as they are not powerful – if that comes at the cost of depriving themselves of political discretion. If governments decide to do so, it happens out of (“selfish”) re-election reasons and not out of altruism vis-à-vis these uninvolved third parties.

Of course, non-contracting parties are (or at least can be) influential in shaping trade policy decisions, but they just do so indirectly via the political support function of trade policymakers. As stated above, SIGs can influence domestic trade policy outcomes by “purchasing” acts of trade intervention (be it protection, subsidization, market liberalization, or some other kind of market intervention) in return for political support.

---

269 The latter view seems to prevail in various contributions (e.g. Charnovitz 2001; Jackson 2004; Pauwelyn 2001; Petersmann 2002, 2003; Tumlir 1985; Vazquez and Jackson 2002).

270 Political economy quite accurately presents trade policy as a “two-level game” in which trade decision-makers have to accommodate domestic constituents when cooperating internationally (Milner 1997; Milner and Rosendorff 1997; Putnam 1988; Ruggie 1982; Ethier 2006; Grossman and Helpman 1995b): In the first-level game domestic special interest groups lobby for a policymaker’s support, and thereby significantly shape the
4 The WTO as an incomplete contract

**Corollary 3: Trade agreements are member-made legal orders.** WTO Members make their own law. Although the GATT/WTO “… does not exist in a vacuum, but [is] an integral part of the wider structure of international law …”, WTO Members will contract freely, as long as basic tenets and peremptory norms of international law are not violated by the treaty.

This is consequential, since it means that the frame of reference for the analysis is explicitly *endogenous*. Self-made rules of the game are in strict accordance with policymakers’ short-, medium-, and long-term interests. The frame of reference for WTO Members thus consists of the norms, rules, and regulations *intrinsic* to the contractual relationship itself – it is not some exogenous legal or normative codex.

**Corollary 4: Trade agreements pursue political, not economic, goals.** The above political economy assumption makes any type of agreement between self-interested policymakers an inherently *political* contract. The utility function of policymakers in charge of trade negotiations is such that it does not directly maximize general economic welfare, but instead pursues distinct re-distributional objectives. Thus, there is no reason to believe *a priori* that trade agreements are concluded for purely economic efficiency reasons. The relevant metric is *political* well-being, not *general economic* welfare.

**Corollary 5: The natural state of affairs is protection, not free trade.** Ever since the writings of David Ricardo and John Stuart Mill, the welfare-enhancing qualities of international trade have been among economists’ most cherished beliefs (Irwin 1996). If it were for economists, free trade should be the natural state of affairs in the world, since in absence of trade barriers global consumption, production, and resource allocation are optimal. Reality, however, paints a different picture: Countries are customarily reluctant to open up their borders to foreign products, and are often unwilling to liberalize trade unilaterally. What Krugman (1991; 1997) termed the “GATT-think” may seem like

---

*second-level* game outcome, in which two or more policymakers carve out the details of an international trade agreement amongst each other.

271 The term “WTO Members” is shorthand for “trade decision-makers representing a WTO Member in international trade negotiations”.

272 Stated in the GATT panel *US – Nicaraguan Trade* case (at §4.5).

273 On this account, see our discussion in Box 3.2 (subsection 3.2.3) and subsection 3.2.4 *supra*, where doubt is cast on theories of *external* relational contracting.

274 Although both SIGs and policymakers individually strive for well-being and hence wish for the most efficient agreement to be concluded, the resulting contract does not necessarily lead to an outcome that is globally efficient in terms of general welfare.

275 “1. Exports are good. 2. Imports are bad. 3. Other things equal, an equal increase in imports and exports is good” (Krugman 1991, p. 5).
economic illiteracy, but is absolutely logical, once the paradigmatic self-interested policymaker enters the scene of trade decisions. Under a set of political economy assumptions, mercantilism – but not free trade – becomes the dominant explanation for the domestic structure of protection in practically all countries: Obviously, there exist some domestic interest groups that benefit from trade protectionism in some sectors, and which can somehow convince domestic policymakers to discriminate against foreign exports.\footnote{An alternative but essentially congruous argument is that policymakers have a preferred domestic distribution of production and consumption rents, and that they choose border measures (tariffs, red-tape measures, quotas) to achieve this goal.} Free trade as a domestic equilibrium of the trade-setting game can be expected to be a result of the domestic tariff-setting game in the rarest circumstances only.\footnote{Indeed, the situation in which a government only cares about social welfare is treated as an improbable “special case” in contemporary models of trade agreements: Either the head of a government in an economically insignificant country is motivated by an autonomous ideological concern (is a “free-trader”, cf. Baldwin 1989). Or no special interest groups exist in a country. As a final alternative, rival special interest groups’ efforts to exercise political influence on the government cancel each other out (Grossman and Helpman 1995b at footnote 11).}

If we assume that policymakers have a preferred domestic distribution of rents, the question is effectively why they do not content themselves with pursuing these mercantilist aims, but instead may wish to conclude a trade agreement with other States or Territories. The short answer is that trade liberalization – just like technical progress and technological innovation – produces consumption and production efficiencies.\footnote{In fact, conceptually, the presence of cross-country trade and international division of labor is little more than in-sourcing technological progress from abroad.} Gains from trade enhance welfare and foster economic growth, which in turn may boost policymakers’ political support, and thus their re-election chances. Yet this answer is too simplistic: It sketches a drive towards free trade, but neglects opposition thereto. In addition, the explanation is apt to describe why governments may unilaterally liberalize, but insufficient for elucidating why countries demand trade cooperation to be a mutual commitment. Lastly, it does not explain what motivates sovereign States to organize their cooperation in the form of a formal written contract.

As a matter of logic, international trade agreements must help trade decision-makers achieve their self-interested political economy aims better, faster and more effectively; otherwise they would not conclude them in the first place, and prefer non-cooperative Nash-settings. Finding convincing explanations for why policymakers want to conclude a binding trade contract is the topic of the next subsection.
4 The WTO as an incomplete contract

4.1.2 Contractual intent: What is the rationale for trade cooperation?

Why do sovereign countries cooperate in trade affairs? Starting with the seminal work of Harry Johnson (1953), economists have strived to formally address this question. In this subsection we will review the available economic rationales for trade policy cooperation. This will be followed by a short critique of the existing approaches – where they fail to convince and what aspects they neglect. Consequently, we point to the rich field of non-economic literature on trade cooperation.

4.1.2.1 An overview of economic approaches to trade agreements

Charts 4.1a and b (next page) provide for an overview of the existing economic explanations for why countries conclude trade contracts.

International economics has come up with two generic answers as to why countries cooperate in trade: One strand of the literature, the so-called commitment school, alleges a strictly internal, domestic problem that a trade agreement can overcome. Another, by far more widely accepted, branch of the literature contends that it is international spillovers and policymakers’ preference for market access that motivate countries to engage in mutual trade cooperation.

As Chart 4.1a (top panel) shows, the commitment approach subdivides into three branches of literature, namely into the time-inconsistency, the hand-tying, and the constitutional approach (numbers 1-3). The market access branch breaks down into two broad categories: the terms-of-trade school (which separates into the optimal tariff and the politically enhanced terms-of-trade perspectives; numbers 4 and 5), and the political externalities view (number 6). As we will show below, the terms-of-trade (TOT) and the political externality variants differ vastly in their assessment of the nature of the international externality that causes countries to cooperate. Some authors have set out to combine two rationales for contracting: Ethier (2006; 2004c), and Grossman and Helpman (1995b) assume the presence of terms-of-trade and political externalities (number 7), while Maggi and Rodriguez-Clare (2005), and Bagwell and Staiger (2005b) combine a domestic commitment with a terms-of-trade argument (number 8).
Chart 4.1 (a; b) Overview of economic rationales for trade agreements

<table>
<thead>
<tr>
<th>Objective</th>
<th>Rationale for contracting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particularistic (political economy)</td>
<td>Domestic problem</td>
</tr>
<tr>
<td></td>
<td>Hand-tying (1)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>General welfare</td>
<td>Time-inconsistency (2)</td>
</tr>
<tr>
<td></td>
<td>Constitutionalism (3)</td>
</tr>
</tbody>
</table>

Source: author
Notes: Charts 4.1a and b offer a categorization of different economic rationales for trade agreements that can be found in international economics literature. Eight distinct approaches can be distinguished (numbered 1-8). Panel “a” (top; in tree form) shows that economics has come up with two generic reasons for the existence of trade contracts: The commitment literature alleges a purely domestic problem that an international contact can help overcome, while the market access externalities approach contends that international spillovers are at the source of the trade contract. Two distinct market access strands differ in their assessment of the nature of international externalities. Some hybrid approaches exist which seek to combine two literature strands. The shaded area shows approaches that are based on political economy assumptions which we subscribe to. TOT stands for “terms-of-trade”. Panel “b” (bottom; in matrix form) is another way of categorizing the existing economic literature: The horizontal axis plots the rationale for contracting, while the vertical axis marks the maximand (objective), which can be general welfare, or particularistic welfare of the self-interested policymaker in charge.
Chart 4.1b puts the information of panel 4.1a in a matrix along two axes: The horizontal axis plots the *rationale for contracting*, according to the underlying problem that the conclusion of a contract can solve. It distinguishes between a purely domestic and an international problem. The vertical axis depicts the *nature of objective* pursued, or the function that is to be maximized. This can either be general welfare or self-interest of the policymakers.  

Without going into great detail, we briefly characterize the existing economic rationales for trade agreements.  

### 4.1.2.2 The commitment approach to trade agreements

The commitment approach to trade agreements alleges a distinct inward-oriented domestic problem that can be solved by international trade covenants. Proponents of this approach argue that policymakers merely utilize *external* pressure generated by the binding conclusion of an international contract to overcome a domestic inefficiency resulting from the strategic interaction between government and the private sector. A trade agreement is concluded so as to deliberately restrict the future discretion of governments over trade policy. By exposing themselves to sanctions in case of contractual defection (protectionist backtracking from previously made trade liberalization concessions), domestic trade policymakers can credibly commit to welfare-superior trade liberalization (lock-in effect). Hence, the commitment to an international agreement is used as a signal against domestic actors that the government cannot afford to renege on its initial contractual commitment, and that *ex post* backtracking is not an option. This external threat makes the policy announcement *ex ante* credible vis-à-vis domestic agents. The commitment argument comes in three flavors (denoted by numbers 1-3 in Chart 4.1):

The first variant is an adaptation of the well-known *time-inconsistency* literature in macro-economics. Domestic problems between a benevolent-dictator-type
government and the private sector can arise when the government’s decision to implement a domestic trade liberalization policy at some future time is no longer optimal when that time arrives. Certain domestic groups engage in strategic actions which render the policy implementation as planned futile. Thus, the announcement that a certain policy will be implemented at a later time is not credible \textit{ex ante}.\footnote{One example of time-inconsistency in trade policy is offered by Matsuyama (1990): Suppose that country \(A\) protects a large yet inefficient industrial sector behind high tariff barriers. \(A\)’s government realizes that at present the costs of maintaining this sector are too high. It therefore announces that at a future date it will open up the sector to international competition. If the announcement is credible, the industry will decide to restructure and invest in cost-saving technologies. The problem of time-inconsistency, now, arises if the industry foresees that the government will not liberalize, should the sector prove to not yet be ready for international competition. Acting strategically, the industry will \textit{not} undertake the required restructuring after the announcement of trade liberalization. At the stage of policy implementation, the government has no chance but to postpone (or refrain from) liberalization, because if it did push reforms through it would face the costs of a crisis in the sector. The government is trapped in a situation in which it cannot credibly liberalize.} As a solution to this conundrum, the benevolent policymaker deliberately “tie[s] her hands to the mast of freer trade” (McGinnis and Movsesian 2000 at p. 515) in order to maximize long-term general welfare. She hopes that an international trade agreement will help her make credible policy commitments affecting the private sector that she would not be able to maintain without the agreement.\footnote{Proponents of this time-inconsistency view in the field of international trade include Staiger and Tabellini (1987; 1999; 1989), Gruenspecht (1988), Lapan (1988), Maskin and Newberry (1990), Tornell (1991), Mayer (1994), McLaren (1997; 2002), Grossman and Maggi (1998).}

The second variant of the commitment school shall be called the \textit{hand-tying approach}. Literature on hand-tying adds a political economy element to the domestic trade-making story: Trade agreements offer a way for a weak government to foreclose political pressure in the face of powerful domestic special interest groups which are lobbying for trade protection.\footnote{The case of hand-tying is different from the time-inconsistency story. The government does not lack credibility, and there is no time-inconsistency problem present. Nevertheless, a government may wish to lock in its policy to diminish the likelihood that its current policies get reversed in the future.} At first sight, this argument may appear contradictory to political economy models that stress the role of lobbies in maximizing a government’s political support: After all, why would a government want to commit to a trade agreement that isolates itself from powerful lobby groups if it receives electoral contributions from them? Maggi and Rodriguez-Clare (1998) describe situations in which short-term benefits in the form of SIGs’ political support are outweighed by long-term costs of protection.\footnote{Think of a situation, wherein a government is subject to lobbying pressure by a well-organized, powerful, but eventually uneconomical import-competing sector. If the country does not have (nor may develop over time) a comparative advantage in that industry, protection will distort investment and lead to an oversized sector. Although letting the SIG have its way may be better for the self-interested policymaker \textit{in the short run}, the costs of these distortions may prove to be too large \textit{in the long run}. The government may thus seek to enter into}
governments seek to minimize unwanted distortions in the present that may arise in the future.\textsuperscript{286}

The third variant of the commitment approach to trade agreements is an extension of the economic literature of \textit{constitutionalism}.\textsuperscript{287} It is based on two central insights, namely that the individual citizen is the legitimate principal in all domestic and world affairs (including trade policy), and that government failure and rent-seeking behavior by public officials are rampant and need to be curbed by means of an adequate constitutional framework. In the presence of overwhelming SIG pressure for trade protectionism, a trade agreement with another country acts as an additional constitutional constraint, a “second line of constitutional entrenchment of personal rights” (Tumlir 1985, p. 87). The conclusion of a trade agreement may hence be seen as a logical extension of the national constitution, aimed at safeguarding the latter’s functioning. For the citizenry of a country an international trade contract operates as an international peg (or anchor) against government misdemeanor and lobby influence.\textsuperscript{288}

\textbf{4.1.2.3 The market access externalities approach to trade agreements}

Most economists tend to prefer another explanation for the existence of international trade agreements, namely that of market-access-related externalities.\textsuperscript{289} The argumentation of this branch of trade cooperation literature runs like this:

Governments want to gain market access to foreign countries for their export sectors. At the same time, they have a preference for high tariff barriers at home (which reduces other countries’ market access to the home market). Unilateral trade protection of one country affects the welfare of trade partners negatively; it provokes harmful market

\begin{footnotesize}
\begin{enumerate}
\item For variants of this political-economic hand-tying argument, cf. Krishna and Mitra (1999), Maggi and Rodriguez-Clare (2005), and Mitra (1999).
\item Early proponents of the literature of “constitutionalism”, such as James Buchanan, Gordon Tullock, Anthony Downs, and Mancur Olsen have shown how citizens – by writing a national constitution – overcome collective action problems, tie down policymakers’ discretion, and curb the influence of government and private actors on fundamental freedoms and civil liberties.
\item Contributions to trade scholarship in the constitutionalist vein include Regan (2006); McGinnis and Movsesian (2000); Petersmann (1986; 2003; 2002), and Tumlir (1985). Hauser and Roitinger (2004, pp. 642) provide for an explanation and overview of the constitutional approach to the WTO.
\item “[A] cross-country externality, such as the terms-of-trade externality, lies at the heart of all of the major theoretical approaches to the study of trade agreements” (Bagwell and Staiger 2002b, chapter 1 at footnote 4).
\end{enumerate}
\end{footnotesize}
access externalities, or negative spillovers. The strategic set-up of a prisoners’ dilemma (PD) emerges. Excessive trade protection, albeit inefficient, becomes the dominant strategy for importing countries. International trade agreements can help overcome these inefficient economic and political market access-related externalities. A trade contract is apt to constrain unilateral “beggar-thy-neighbor” policies by eliminating those inefficient restrictions in trade volumes which arise when policies are set unilaterally. Thus, trade agreements offer governments an opportunity to escape a PD.

Theories for trade agreements based on market access externalities come in two general categories: The first variation is the terms-of-trade (TOT) school which further subdivides into the optimal tariff approach (number 4 in Chart 4.1 above) and the politically enhanced TOT approach (number 5). The second variation is the political externalities school (number 6). A mixed externalities school tries to combine both categories (number 7).

The optimal tariff approach: The oldest theory of trade cooperation alleges a terms-of-trade-driven PD between two benevolent welfare-maximizing governments. The dilemma can be overcome by means of concluding an international trade accord. The literature on optimal tariff policy has a long history dating back to John Stuart Mill, and was formalized by Harry Johnson in 1953. Its basic insights remain influential through the work of Bagwell and Staiger and their politically enhanced TOT theory (infra). In brief, the optimal tariff approach recognizes that large, economically powerful countries can shift some of the costs of a domestic protectionist measure (the erection of an import tariff) on foreign trade partners through the depression of world prices for that import good. Trade protection can improve the TOT of large countries, and therewith enhance general welfare of its citizenship – at the expense of trade partners’ welfare: Beggar-thy-neighbor behavior depresses world prices for foreign export goods.

According to proponents of the optimal tariff view, large countries realize that a reciprocal selection of import tariffs produces significant global inefficiencies. These inefficiencies can be avoided by mutual trade cooperation in the form of a trade agreement which fixes worldwide terms of trade, and thereby locks in world prices.

---

290 A State’s government sets its import barriers in order to maximize its welfare. It recognizes that some of the cost of this measure will fall upon foreign exporters whose products sell less. This externality leads all rational governments to set unilateral trade barriers that are higher than would be efficient.

291 By taxing its imports, a large country has the means to inflict TOT inefficiencies on its trade partner(s). In return however, each country is prone to TOT-deteriorations inflicted upon its own export sectors through large countries’ erection of import tariffs.
So far, so good. But what is the intellectual link between the rather abstract concept of TOT consequences of trade policy choices and trade policymakers’ proven appreciation for market access? After all, trade negotiators are always eager to point out the additional market access commitments they have secured over the course of trade talks. Yet they rarely mention the TOT implications of their negotiations. Bagwell and Staiger (2002b, pp. 28) claim to have found a surprisingly simple answer. “The terms-of-trade consequences of trade policy choices can be expressed equivalently in the language of market access, and so the terms-of-trade consequences and the market-access implications of trade policy choices are different ways of expressing the same thing” (ibid. p. 5). According to the authors, any price effect (say, a TOT deterioration for country A through B’s imposition of a trade impediment) necessarily has a corresponding volume effect (country A’s reduction of access to B’s markets).292 Countries have a (TOT-induced) appreciation for unequal, imbalanced market access, and are wary of granting unilateral liberalization for that same reason. The conclusion of a reciprocal free trade deal contractually fixes mutual market access via common, undistorted world prices for all goods and services.

The politically enhanced TOT approach: Influenced by the advances of political economy models in international economics, Bagwell and Staiger have enhanced the neoclassical optimal tariff approach by giving it a political-economic twist (see approach 5 in Chart 4.1).293 The distinguishing feature between the traditional optimal tariff school and the politically enhanced approach is that governments, in addition to caring about economic efficiency consequences of local-price movements implied by their tariff selections, may also be motivated by political (i.e. distributional) objectives.294

---

292 Bagwell and Staiger (ibid., p. 29) contend: “[W]e may interpret ‘cost shifting’, ‘terms-of-trade gain’, and ‘market-access restriction’ as three phrases that describe the single economic experience that occurs when the domestic government raises its import tariff and restricts foreign access to its market […] We may now say that a government secures additional market access from its trading partner through negotiations if there exists a world price such that the trading partner’s negotiated policy changes provide additional access to the trading partner’s market” (emphases in original). See also Bagwell and Staiger (2002a; 1999), and Bagwell et al. (2002).

293 The politically enhanced TOT school is championed by Kyle Bagwell and Robert Staiger, but was taken up by a multitude of WTO scholars (Bagwell and Staiger 1990, 1997, 1999, 2002a, 2002b; Bagwell, et al. 2002; Bown 2002a; 2003; Grossman and Helpman 1995b; Ethier 2001a; 2004a).

294 Self-interested policymakers are concerned primarily with a number of internal (political-economic) objectives that relate to the domestic relative price of imports in terms of exports. Bagwell and Staiger formally model this concern by assuming the most general government objective function possible: The only structure placed on policymakers’ PROF is that, holding its local price fixed, a government is assumed to achieve higher welfare when its TOT improves (2002b, pp. 19).
The outcome of the politically enhanced economic externality school à la Bagwell and Staiger is that eliminating the TOT externality is still the sole rationale for trade agreements – even under a political-economic perspective. Abandoning the assumption of benevolent policymakers does not decrease the significance of purely economic TOT externalities as the core motivation for the conclusion of international trade agreements.\(^{295}\) The only outcome that is truly different from the optimal tariff approach is that the resulting reciprocal trade accord does not necessarily prescribe free trade, but displays positive, politically optimal reciprocal tariffs in the equilibrium.

**The political externality school of thought:** The political externalities strand of the trade literature (number 6 in Chart 4.1) gained prominence through Ethier (2004b; 2004c). It is critical of the terms-of-trade approach, alleging that “a trade agreement serves governments to get credit for the reduction in foreign trade barriers. International cooperation is not about the elimination of economic (world-price) externalities, but about political externalities. The latter arise when politicians in one country believe that their political status is directly affected by actions of politicians of another country” (Hauser and Roitinger 2004, p. 652, emphasis in the original; see also Levy 1999; 1997).

In a model of two small countries (where the established TOT theory would postulate reciprocal free trade as a Nash solution, and therefore an absence of a trade contract), Ethier (2004b) detects a motivation for trade agreements based on international political spillovers. The model features self-interested policymakers,\(^{296}\) unsophisticated factor owners, and multiple sectors. Factor owners (import-competing and exporting interests) care about trade volumes independently of factor rewards, and give trade decision-makers extra (mis-)credit for the direct market access effects of government actions.\(^{297}\)

\(^{295}\) “While the inclusion of political concerns enhances the realism of the model, we show that it does not offer any separate purpose for trade agreements. Whether or not governments have political motivations, it is their ability to shift the costs of protection onto one another through terms-of-trade movements that create an inefficiency when tariffs are selected unilaterally […] the purpose of a trade agreement is to offer a means of escape from a terms-of-trade-driven prisoners’ dilemma” (Bagwell and Staiger 2002b, p. 3).

\(^{296}\) Ethier’s underlying government PROF follows the tradition of earlier political economy contributions which have stressed the importance of political motives behind reciprocal market access exchange (such as Finger 1991; Hillman et al. 1995; Hillman and Moser 1996; Hillman and Ursprung 1994, 1988; Hauser 1986; Moser 1990).

\(^{297}\) Exporters appreciate direct market access gains from trade cooperation, while import-competing interests feel hurt by lower domestic trade barriers (see Axioms 1 and 2 in Ethier 2004b). Technically, factor owners are bounded rational in that they are ignorant of the “Lerner Symmetry”, which postulates balanced trade volumes. This does not seem an unrealistic hypothesis in a world with more than 2 sectors (let alone more than 2 countries), yet must be seen as an additional confining assumption.
The same mercantilist logic as held by the traditional models applies: Governments benefit from being granted market access by other States, but dislike ceding market access. Unilateral liberalization is thus out of the question for self-interested policymakers. Hence, some level of *mutual* market access cooperation will occur between the policymakers of the two small trading nations. Governments need *direct* market-opening achievements vis-à-vis their exporters in order to counterbalance the ire of import-competing sectors over *direct* losses in domestic sales. A trade agreement can remedy this situation. Governments exchange market access commitments in the form of tariff cuts, since each signatory party is dependant on the goodwill of the other. This is the nature of political externalities. Any trade agreement must be based on reciprocity. This results in additional market access (and therewith extra-profits for exports), but also in a *neutralized* *direct market access balance* (which could be called a policymaker’s *balance of blame*). A balanced market access score brings about sufficient political support from exporters so as to outweigh protectionist pressure from import-competing industries.

Ethier’s model yields results that are quite distinct from the classical TOT-induced rationale for trade agreements. In contrast to Bagwell and Staiger’s politically enhanced TOT approach – international market access is appreciated for purely general welfare reasons – political market access externalities are part and parcel of the government’s domestic objective function. The international market access balance *directly* influences the domestic political support balance, since the support of governments by specific factor owners is directly linked to market access considerations. Without the trade agreements that secure market access cooperation, policymakers would not be able to achieve their goal of maximizing their political support function. In other words, political externalities are *political-support reverberations* of the unilateral market-closing behavior of trade partners.

**The mixed externality school of thought:** In Grossman and Helpman (1995b), and Ethier (2006; 2004c), trade scholars link an endogenous trade policy model of domestic tariff-setting with a mixed model of trade agreements based on political and economic

---

298 Also, since export subsidies are assumed away, a government is not able to simply tax imports and use the tariff revenue to subsidize exports.

299 Due to their comparative advantage, exporters are assumed to have bigger clout with domestic policymakers.

300 In contrast to that, in the Bagwell/Staiger-model world the preferred domestic income distribution can be achieved through *any* set of internal policy measures, which do not even have to be connected to trade policy. The trade agreement is solely concluded to escape a welfare-economic negative-externality PD.
externalities (cf. number 7 in Chart 4.1). These contributions must be lauded for opening the “black box” of the domestic trade policy structure: The models in question present explicit microeconomic foundations for the interaction between self-interested policymakers and domestic SIGs.\textsuperscript{301} The authors use an extended version of Grossman and Helpman’s 1994 “Protection for Sale” model (see above footnote 268). The domestic trade-setting game is hereby supplemented with an additional stage of international trade cooperation: Self-interested policymakers in two large countries are assumed to maximize a weighted sum of consumers’ and industry-SIGs’ welfare. Again, governments’ desires to contract is driven by market access considerations. Parties enjoy additional access, avoid unilateral market opening, and are happy with a reciprocal market access balance that is evened out. In contrast to models of the two previous approaches to trade agreements described above, both political support and TOT externalities emerge from the structure of the domestic lobbying game. Notably, their presence is not assumed \textit{ex ante}. Grossman and Helpman’s 1995 article confirms findings of the politically-enhanced TOT approach, namely that the only relevant externalities travel through TOT movements. Applying a slight modification of the Grossman/Helpman (1995b) framework, Ethier (2006; 2004c) reaches a different conclusion: If self-interested policymakers only care about political externalities and not at all about TOT, the resulting trade agreement will look like the GATT/WTO, viz. a cross-sectoral, multilateral agreement that ousts export subsidies.

\textbf{4.1.2.4 A brief evaluation of rationales for trade contracts}

Why do sovereign countries conclude trade liberalization contracts? Above we have reviewed eight answers from the discipline of international economics. In this section, we would like to bring forth a short evaluation of these theories. This will be followed by a more general critique of economic approaches.\textsuperscript{302}

---

\textsuperscript{301} The mixed externality approach does not only present solid micro-foundations for the domestic lobbying game. It is also free from the arbitrary assumptions that were criticized in the original political externality-approach (Bagwell and Staiger 2002b, p. 31): SIGs are \textit{not} assumed to be ignorant of the Lerner Symmetry, and export subsidies are allowed as feasible domestic trade policy (they are \textit{not} assumed away).

\textsuperscript{302} We do not engage in a full-blown evaluation of all economic approaches here, but rather choose to highlight certain criticism is most relevant. The inclined reader is referred to the reviews in WTO (2007, section II.B), Hauser and Roitinger (2004, sections 1.2 and 1.3), and Bagwell and Staiger (2002b, section 2.1.5). A general critique of economic approaches to trade cooperation and an overview of non-economic theories can be found in WTO (2007, sections II. B, C, D, and F).
A The commitment approach

It is often argued that the commitment approach to trade agreements is still in its infancy (Bagwell and Staiger 2002b, p. 35; Hauser and Roitinger 2004, p. 649). We concur with this point of view. Although we do not deny that international trade agreements may indeed help governments justify awkward trade liberalization policies to domestic constituencies (cf. Staiger and Tabellini 1999), it is hard to accept the commitment approach as a full-scale motivation for policymakers to conclude a trade agreement. We summarize briefly what we take to be the most pressing counter-arguments against the commitment approach.

1. A “commitment contract” does not solve a problem. Commitment scholars suggest that international trade agreements are concluded for entirely (or largely) domestic reasons. The trade contract is therewith merely a means to an end. All contracting parties subscribe to the same end of trade liberalization, employing the means of punishment threats. This reduces trade negotiation dynamics from a PD-type collaboration game to a simple coordination game. A synergetic coordination agreement, where every party has congruent objectives, does not solve a problem, since one does not exist – the risk of opportunism is notably absent (see our discussion in subsection 2.3.1 above). This begs the question of why parties would feel the need to sit down, negotiate, and put into writing a substantial body of rights and obligations at all. A simple handshake and the common accord to apply infinite (coercive) punishments for defections would suffice to achieve policymakers’ objective of tying their hands.

2. A trade contract solely based on the commitment rationale has a counterfactual institutional design. A hypothetical “commitment contract” (i.e. one that follows the strict logic that commitment theorists allege) would have characteristics quite distinct from any trade agreement observed in the real world. If policymakers really needed the trade agreement as a mere commitment device for binding their future discretion, any sort of ex post discretion should consequently be completely prohibited. Contractual flexibility mechanisms, as well as non-coercive enforcement remedies, clash with this

---

303 Commitment- and externality-based theories are argued to be “possibly complementary” (Bagwell and Staiger 2002b, p.14). Various commitment-cum-TOT (number 8 in Chart 4.1) approaches seem to support this contention (e.g. Bagwell and Staiger 2005b; Maggi and Rodriguez-Clare 2005).

304 On this note, see Mavroidis (2007, chapter 1), and Srinivasan (2005) in addition to the sources mentioned in footnote 302.
commitment logic. Yet escape and less-than-prohibitive punishment for contractual deviation are a key feature of virtually every trade agreement in existence today.

Another counterfactual of a hypothetical trade agreement based on commitment problems is that we should see a very low level of trade concessions: Intuitively, a high level of rigidity (bindingness of trade obligations) in a trade agreement significantly reduces a country’s readiness to commit to extensive liberalization. Hence, the depth and breadth of liberalization should be expected to be relatively modest and commitments rather static in nature. This, however, collides with the successful trade liberalization efforts the world trading system has witnessed over the last 60 years.

3. Why would governments choose an external agreement to lock in their commitments vis-à-vis influential special interest groups? The commitment approach posits a domestic problem, but proposes an international hand-tying solution. A trade agreement is probably not the most intuitive and definitely not the most straight-forward solution to a domestic commitment problem. Isolating trade policymakers from the domestic political process, or delegating authority to politically less exposed actors (such as the executive or a national trade board) may be an equally efficient national means to achieve this commitment end. Giving a domestic trade decision-maker an independent third-party status akin to a Supreme Court judge or a central bank could yield better outcomes.

4. Why would self-centered policymakers forestall their future policy discretion? The political economy approach to international trade replaces the maximization of social welfare by a PROF that reflects the self-interest of the incumbent. Under this notion, can hand-tying really be a viable long-term strategy for trade negotiators, and if so, under what circumstances? Intuitively, it is hard to see why trade policymakers

---

305 If we were to explain the role of trade policy flexibility mechanisms in a “commitment world”, we would need convincing explanations as to how non-performance clauses and policymakers’ commitments can co-exist. One way could be along the lines of Ethier (1998; 2002), who models trade negotiators as individuals different from those deciding on subsequent protectionist policy. But this altruistic behavior on the part of trade negotiators clashes with the precept of policymakers’ self-interest. Another possible way is to assume that governments try to trick their partners after the conclusion of the agreement by secretly engaging in unobservable non-tariff barriers (Copeland 1990; Copeland et al. 1989). But this is also hardly an explanation for formal trade flexibility rules that exist in almost every trade agreement.

306 Fierce interest group battles in the run-up to the conclusion of such a rigid agreement can be expected. These turf wars should prevent governments from making far-reaching commitments (see Ethier 2004b; 2001b for economic underpinnings of this assertion).
would give away their most important policy tool – trade policy flexibility – once and for all in return for the straightjacket of a trade agreement.\textsuperscript{307}

B The politically enhanced terms-of-trade approach

Turning to the politically enhanced terms-of-trade approach to trade agreements, it seems important to note that this is the only research program that formally integrates an explanation of why countries cooperate, with an explanation of how they can do this. So far, it is the only available economic approach to explain both the existence of multilateral trade agreements and their architecture (Keck and Schropp 2007, p. 5).

However, there is some conceptual criticism to be brought against the Bagwell/Staiger school: First, their approach can neither explain why small countries join trade agreements, nor why large countries would allow small countries to accede.\textsuperscript{308} Second, the politically enhanced TOT school fails to elicit why export tariffs (which are apt to influence world prices just as well as import taxes do) are not banned in most trade agreements, and why export subsidies on the other hand are prohibited. Third, this theory does not explain why governments would use a relatively inefficient policy tool, such as import tariffs, to manipulate domestic prices in order to maximize their political support (Rodrik 1995, p. 1476). Fourth, the politically enhanced TOT theory assumes a neoclassical world characterized by perfect competition, constant returns to scale, and constant demand elasticities. These are consequential and indeed confining assumptions. For example, in a Bagwell/Staiger world, producers always operate along their production-possibility frontier and can always sell all their goods and services. Under this notion, policymakers cherish additional market access \textit{not} because of additional export sales, but solely for its TOT impact on world prices. But reality paints a different picture in this respect. Exporters cherish additional market access precisely for heralding additional sales and larger capacity utilization. This raises some doubt as to the general applicability of the Bagwell/Staiger world.

\textsuperscript{307} Maggi and Rodriguez-Clare (1998) examine the conditions under which the commitment approach would lead self-interested policymakers to sign a trade agreement. The counterintuitive results of their model are that countries with strong import-competing special interest groups and weak government are more likely to enter an international trade agreement. In addition, only if policymakers care significantly about both social welfare and campaign contributions will they engage in trade negotiations. Notably, too much responsiveness of policymakers to political contributions, as well as too much benevolence (low responsiveness to special interests) will render efforts to engage in multilateral trade negotiations futile.

\textsuperscript{308} See WTO (2007, chapter II.B.5) on the issue of \textit{asymmetrical} trade cooperation between countries of different economic sizes.
As a final criticism, the politically enhanced TOT approach fails to explain the persistent anti-trade bias which permeates real-life domestic trade policy (Ethier 2004c, p. 3): Why is TOT-induced protectionism more valuable to large-country governments than political pressure for free markets on part of export SIGs, of foreign SIGs and of downstream industries lobbies, or than consumer welfare? (See Hauser and Roitinger 2004; Levy 2003; Ethier 2004c; Ethier 2004b; Rodrik 1995; Roitinger 2004 reviews these points of criticism in more depth.)

C The political externalities approach

Ethier’s (2004b) small-country approach to trade agreements, featuring pure political externalities, must be reproached for deploying a fairly unusual and arguably “hand-waiving” government PROF. Many of the author’s findings are a direct consequence of his Axioms 1 and 2 (ibid., p. 306), which present a rather rigid corset of assumptions: As Bagwell and Staiger (2002b, pp. 30) contend, Ethier’s argumentation is driven by the external restrictions of (i) bounded SIG rationality, and (ii) the prohibition of export subsidies. Also, Ethier is notoriously ambiguous about the real nature of his alleged political externalities.

D The mixed externalities approach

This approach to trade agreements is a promising path, but still in working paper stage. Ethier’s (2006; 2004c) contributions in particular are extremely difficult to follow, and it is almost impossible to present a sensible economic interpretation of his outcomes. Also, the author bases his analysis on a modified Grossman/Helpman (1994) model of domestic trade protection, which itself uses quite special and rigid assumptions. On the whole, Ethier’s mixed externality model is good in terms of its general intuition, but far from a full-blown theory of political market-access externalities.

E A general critique of economic models

As a more general critique that concerns economic theories of trade agreements in general, we would like to point out three shortcomings: First, we take issue with those approaches based on the assumption that trade policymakers try to maximize a general welfare objective function (this goes for the optimal tariff, the time-inconsistency, and

---

309 See footnote 268 above for the limitations of Grossman/Helpman’s model of endogenous trade policy generation.
the constitutionalism theories, as Chart 4.1 illustrates). Not only does general welfare maximization collide with our political economy assumption of politically realistic, government objective functions. The assumption of benevolent dictators also defies realism: Governments, no matter whether democratically elected or dictatorial, always have domestic distributional goals that they address by means of trade instruments. Practically no small country in the world (except perhaps Singapore) applies anything remotely close to a zero tariff-level on imports.

Second, another important shortcoming of all eight reviewed approaches is that they are mainly theories of trade in goods. Little consideration is given to trade in services, intellectual property rights, investment measures, issues of competition or government procurement, etc.

Third, even theories about trade agreements based on goods are in effect models of tariff protection, but not ones of trade protection in general. They explain trade cooperation only in terms of import taxes; non-tariff barriers are usually completely omitted from these models.310

### 4.1.2.5 Additional rationales for trade contracts

We have contended that none of the reviewed economic rationales for trade cooperation is flawless in its argumentation. Further, it seems that economic scholarship has overlooked additional motivations for the conclusion of trade agreements. We can detect two areas of omission by trade economics: First, economists have excluded from consideration important non-economic objectives of cooperating in trade affairs. Second, they have neglected an additional economic contracting rationale: Trade agreements may be concluded to solve a multilateral coordination problem, and to establish and safeguard minimum global trading standards. Consider Chart 4.2 as an illustration:

Chart 4.2 is an extension of Chart 4.1.b, and is plotted along the same two axes. The boxes of the matrix are filled with various rationales for trade agreements brought forth by the disciplines of economics (as reviewed above), IR, and international law.311 Interestingly, the eight economic explanations for trade agreements discussed above are scrambled into four cells in the upper left-hand corner of the matrix.

---

310 Notable exceptions include Hungerford (1991), Copeland (1990), Levy (2003), and Horn (2006).
311 See WTO (2007, chapter II.B.3 and 4) for detailed explanations and literature resources of IR and legal approaches to trade cooperation.
Chart 4.2 Economic and non-economic rationales for trade agreements

The vertical axis distinguishes between different “natures of objective”. The first two rows plot theories of trade agreements that place economic efficiency at center stage: Governments may wish to maximize the general welfare of their citizenship, or self-interested policymakers maximize their own PROF. Economic models of trade cooperation assume that agents cooperate if it serves their rational interest. By combining forces, signatories seek to “maximize the size of the pie” by inducing efficiency.

Despite this, (politico-) economic utility maximization and absolute efficiency (that we also subscribe to) may not be the only motives that drive countries into trade cooperation. Players may try to maximize their relative power position vis-à-vis other countries (see the bottom row of Chart 4.2). To proponents of this view (mostly IR scholars), international cooperation is akin to a zero-sum game, where gains by one party necessarily come at a loss to another one (the “size of the pie” is thought to be fixed).
Cooperating parties anxiously watch over their power rank within the international system, aiming to maximize their “slice of the pie”.312

Non-rationalist approaches to decision-making (second row from the bottom) reject the assumption that agents are only driven by a narrow pursuit of selfish well-being and rigid cost-benefit considerations, and instead subscribe to fundamental norms, shared ideas, inter-subjective beliefs, traditions, and habits as drivers of human action (see footnote 260 above). In the trade realm, the establishment of an “international society” and the peace-promoting quality of trade are said to be important norm drivers for the conclusion of trade pacts. The idea that multilateral, non-discriminatory trade would contribute significantly to global peace was reportedly a central motive for English and American policymakers designing the post-World War II trading system.313

On the horizontal axis, the distinguishing criterion is “rationale for contracting”, or the kind of problem that the conclusion of a trade agreement can help overcome. As was shown in Chart 4.1.b above, economic trade theorists allege that trade contracts can solve a distinctly domestic commitment issue or problems connected to international market access externalities. There may, however, be additional reasons for contracting:

First, theories adhering to methodological collectivism (also called structuralism, or holism) contend that it is the system that shapes the actors and not vice-versa. The system, according to structuralists, is more than the sum of its constituent parts, and is therefore apt to shape how agents think, act, and interrelate (see footnote 260 above). The international context, or the nature of the system, impels States to act cooperatively in trade affairs. Some theorists from the disciplines of IR and international law thus allege some form of superordinate systemic imperative for contracting.314

A second rationale for contracting that largely went unnoticed by trade economists concerns the strategic set-up of the cooperation game which a trade contract is able to overcome. All theories of trade agreements based on market externalities reviewed

312 Under this conception of distributive efficiency, parties worry about relative gains, and largely ignore (or take for granted) absolute gains (see our discussion in footnote 51 in subsection 2.1.2 above).

313 See Penrose (1953) and Meade (1942). The genuine belief that trade “dovetailed with peace” (Hull 1948, p. 81) seemed to have been a principal driver among Western Allies for establishing a world trade order in the Post-World War II era (see Irwin 1996, 2005). The historical record clearly shows that the ITO/GATT was envisioned by the Allied powers to be part of a bigger cooperation scheme that included the establishment of the United Nations, the World Bank, and the International Monetary Fund (see e.g. Jackson 1969; Dam 1970; Gardner 1980).

314 For details and references of constructivist approaches to trade agreements, see WTO (2007, section II.B.3.b.ii).
above recognize the presence of an international PD set-up in the interaction between countries. In other words, it is assumed that international trade is akin to a collaboration game. However, as we stated in subsection 2.3.1 above, contracts can also be concluded with the aim of solving complicated coordination games. Unfortunately, as Chart 4.2 indicates, the issue is largely unstudied in the trade literature, and specifically in what respects trade agreements can be conceptualized as coordination games, and the impact this would have on the design of trade agreements (see also WTO 2007, chapter II.B.3.b). As we will argue infra, trade agreements based on “minimum standards” or “positive integration”, such as the WTO Agreements TRIPS or TRIMs, may be properly conceptualized as collaboration games.

4.1.3 A tentative conclusion: Trade agreements based on market access externalities and minimum standards

We opened subsection 4.1.2 looking for the contractual intent of the WTO. What can be learned from this brief literature review of trade cooperation theories? We draw two conclusions.

Firstly, trade agreements are probably best characterized as “mixed-motive games” (WTO 2007, section II.B.6; Mavroidis 2007, subsections 1.2.4 and 1.2.5). A mixed-motive game has an internal and an external dimension to it: Internally, any country is likely to be motivated by an array of (partially conflicting) economic and non-economic objectives that are pursued by signing up to a trade agreement. The government of a country may wish to promote peace and stability in its region, propel its power rank in the international system, attract foreign direct investment, fight unemployment, mitigate the influence of special interest groups, or stop trade partners from engaging in excessive beggar-thy-neighbor policies.

315 “Different as the two rationales for trade agreements [i.e. the TOT- and the political externalities approach] may be in substance […], their basic intuition is identical: Both schools allege a simple game set-up where two or more rational players are faced with a PD situation. A trade contract can help overcome the inherent inefficiencies – given that two fundamental conditions hold: infinite repetition and the self-enforcement property. Infinite repetition avoids an immediate breakdown of the trade game, and self-enforcement means that any punishment (or rather: threat thereof) can successfully be enacted by the membership of the agreement itself” (Keck and Schropp 2007, p. 6).

316 To recapitulate: In coordination games signatories join forces so as to reap mutual transaction efficiencies (synergies). All contracting parties know that conceding to some level of regulation is mutually efficiency-enhancing. However, doing so requires them to contractually fix the mutual level of cooperation ex ante. Disagreement may arise among the signatories over finding, and maintaining, the optimal level and composition of cooperation.
4 The WTO as an incomplete contract

The *external* dimension of a mixed motive game is that signatory countries can be expected to be quite heterogeneous in their contracting goals: By concluding a trade agreement, small, democratic, developing countries can be expected to pursue different sets of core objectives than large, dictatorial, developed regimes. In other words, each country probably possesses a unique set/bundle of trade cooperation objectives.

Our second conclusion is that trade scholarship is still far from establishing a convincing answer as to why sovereign countries engage in trade cooperation. More work needs to be done to produce testable results as to which of the discussed approaches (or which combination thereof) best manages to explain countries’ cooperation motivations. Whilst current economic, political, and legal approaches seem able to elucidate facets of the cooperative drive, each one is far from capturing the whole picture. Cross-disciplinary work seems a fruitful and promising avenue for future research.

Where does that leave us in our quest to identify contractual intent of the WTO? In lieu of a *conclusion* we offer a *convention*: We largely side with Mavroidis (2007, chapter 1.2.5) who contends that the GATT, although a mixed-motive game, was probably mainly concluded so as to constrain excessive unilateralism. In other words, countries – for whichever underlying reason – are driven by the objective of gaining access for domestic goods and services to foreign markets. Trade contracts are thus concluded with the predominant aim of achieving and safeguarding a mutually acceptable *reciprocal market access balance*. We add a second, albeit subordinate, rationale for contracting: The provision of general minimum standards in important trade areas such as intellectual property, investment, or import licensing.

4.1.3.1 *The primacy of market access externalities*

Our argument in favor of the primacy of the market access motivation is nurtured by the following deliberations: Firstly, it is probably fair to say that economic rationale is at the core of the WTO. After all, trade is about exchange of goods and services – a profoundly economic enterprise. Had peace, security, development or social philanthropy been the central objectives for contracting, parties could have concluded

---

317 “The various GATT instruments are there to guarantee that the multiple negotiation effects stemming from unilateral definition of trade policies will be addressed” (Mavroidis 2007, p. 26).

318 The actual content of the contracts, preparatory work of the ITO/GATT, recollections of contemporaries (e.g. Hull 1948; Meade 1942; Penrose 1953), and findings by economic historians (e.g. Irwin 1996, 2005; Miller 2000) may count as arguments in favor of the profoundly economic nature of ITO, GATT, and WTO.
treaties with exactly these aims. In fact, countries – by establishing the United Nations with all its adjunct organizations – have proven to be willing to contract for these reasons. This contention rules out non-economic rationales as the main drivers of cooperation in trade. It is sufficient to presume that trade should not interfere with concerns of global peace and stability.

Secondly, we argued above that trade policy is likely to be shaped by self-interested policymakers who are directly influenced by domestic stakeholder groups. This political economy tenet leaves little room for orthodox state-centric approaches to trade cooperation, which assume that governments are monolithic entities that strictly pursue general welfare objectives. In the same vein, constructivism and all other theories adhering to methodological collectivism are largely excluded from the explanatory ambit that we cherish.

Thirdly, we adhere to a Paretian logic of contracting: Policymakers are inclined to conclude trade contracts that are good for themselves. This is a rejection of those IR approaches that assume trade agreements to be a continuation of diplomacy with other means (to paraphrase von Clausewitz’ famous dictum), geared towards forcing other countries into submission, and securing the nation’s power rank in the international system. Although the issue of “trade as a geopolitical instrument” may constitute an interesting topic, we do not consider it in this study (see discussion in footnote 51 above).

Fourthly, we deem the theories based on a purely domestic commitment rationale unconvincing. As discussed supra (subsection 4.1.2.4.A), we take the three variants of the commitment approach to trade cooperation to be weak rationales for contracting, not least because they construct counterfactual outcomes.

Finally, we note that trade negotiators emphasize the market access implications and opportunities of trade agreements in their communication with the public (see Bagwell and Staiger 2002b at p. 182).

In summary, we regard externality-based theories of trade agreements most fitting to explain why sovereign countries may want to conclude multilateral trade contracts: A trade contract based on market access externalities solves a problem and thereby displays all the ingredients of a contract. It displays a clear motivation for a trade agreement – market access potential to foreign markets. It also assigns an assurance function to the
contract, i.e. the goal of constraining excessive unilateral behavior by repetition of the prisoners’ dilemmatic trade game.\footnote{Motivation and assurance are the two most decisive components of every contract, as stated in Chapter 2.1.2 above.}

\subsection*{4.1.3.2 Trade contracts based on market access externalities}

Our view of trade contracts based on the market access externality motivation is the following: We are agnostic as to the micro-foundations, character, shape and composition of the PROF of trade policymakers (see above footnote 268 and accompanying text). The aim of policymakers is to get re-elected. In pursuit of this goal, they strive to maximize their political support. Depending on its degree of responsiveness to lobby influence, each government’s objective function is biased more or less in favor of organized interests.

Domestic special interest groups are keen on maximizing the well-being (income) of their members, and hence have a profound stake in the structure of domestic trade policy. SIGs mainly fall into two groups: pro-protection (import-competing) and pro-export interests. Exporters have a genuine interest in export subsidies as well as in open international markets, since those measures promise them additional access to foreign markets and cheaper sourcing, respectively. Import-competing interest groups also cherish subsidies, but favor trade protection (at least in their respective sectors) over open markets. The general electorate (consumers) prefers cheaper products, more efficient resource allocation, employment opportunities, and higher wage levels.

Accordingly, the self-interested policymaker has at least four – partially countervailing – interests to consider when defining national trade policy. The components of her domestic political support balance are (i) consumer-, (ii) export-, (iii) import-competing- and (iv) TOT interests.\footnote{Consumer and exporter interests presumably work at cross-purposes with import-competing interests and TOT motives.} Any trade policy decision may produce complex domestic reverberations in the form of \textit{direct} – political – and \textit{indirect} – economic – implications for the initiating government, namely in the form of TOT- and political support effects.\footnote{For example, fixing world prices with the aim of neutralizing TOT externalities may not go down well with import-competing industries which had hoped for extra protection, but is popular with consumers, importers, consuming industries, and exporters. Depending on the political or financial clout that import-competing lobbies bring to bear, a trade decision-maker may refrain from liberalizing that sector, which in turn may upset the TOT balance and enrage export interests and consumers.}
Domestic decisions aside, it is certain that protectionist trade measures taken by a trade partner will affect general welfare (through TOT deterioration and reduced gains from trade) as well as local lobbies’ income. The political status of policymakers is thus damaged by opportunist market-closing policies by foreign countries. Yet in the same vein, the home government’s political status increases if it manages to “snatch” unilateral market access from its trading partner(s). The familiar prisoners’ dilemmatic situation is the result: Anticipating the risk of excessive unilateral action by trade partners, both (or all) countries will revert to non-cooperative Nash behavior, in which trade barriers are high, trade in goods and services inefficiently low, and political support suboptimal (voters, exporters, and consuming industries punish the government for the high trade barriers and promise more support if international trade is promoted). A mutual trade agreement can help trade policymakers reap higher political support levels from the general electorate and domestic SIGs: Countries agree to cut down on their trade barriers with each other (be they tariffs, quotas, red-tape measures or other market-closing policies). As long as the political loss of doing so (in terms of foregone TOT gains and political support losses of import-competing lobbies) is outweighed by positive externalities (political support gain and general trade efficiencies), a trade agreement is mutually beneficial to self-interested policymakers.

Two aspects are important for our characterization of trade agreements based on market access externalities: First, the resulting market access deal between signatories is profoundly political (as per policymakers’ utility functions), and it must be reciprocal. Reciprocity is an imperative of the Paretian logic of contracting. Two signatory countries engage in market access negotiations until the first party (the most reluctant liberalizer) hits its political optimum, i.e. its preferred domestic political support balance. That country (rather: its political representative) has then successfully optimized its preferred bundle, consisting of sector-specific market-access-commitments received and commitments granted. It has successfully exchanged the biggest possible set of concessions for the politically lowest price. Once at its optimum, the government is

---

322 The politically optimal level of liberalization implies welfare-maximal breadth and depth of the agreement – as seen from the point of view of the trade policymakers in charge. Governments are inclined to include an optimal set of sectors in the trade deal (breadth), and are willing to liberalize each sector up to its politically optimal level (depth).

323 Ideally, a trade negotiator would trade off the access to her domestic market as “cheaply” as possible: She would like to liberalize those domestic sectors that are economically unimportant, where consumer- and/or downstream industry interests are powerful, or sectors which are weakly represented by special interests (e.g. truck drivers, who in many countries are notorious for failing to organize their interests). Naturally, each policymaker urges trade partners to liberalize as many important markets as freely as possible (especially those where domestic export interests are strong).
not willing to change the terms of the deal anymore. Any further liberalization would lead to a deterioration of the trade negotiator’s domestic political support balance.\textsuperscript{324} Free trade, notably, is rarely the outcome of these reciprocal market access negotiations.\textsuperscript{325}

The second point we wish to emphasize is that multilateral trade contracts are essentially webs of bilateral trade deals. The existence of non-discrimination provisions notwithstanding, every signatory engages in bilateral and reciprocal promises of market access.\textsuperscript{326} Trade liberalization commitments are owed on a country-to-country basis, not towards the membership as a whole.\textsuperscript{327}

\textbf{4.1.3.3 Minimum standards as a second(ary) rationale for trade agreements}

Above, we criticized mainstream economics for exclusively rooting the rationale for trade agreements in the prisoners’ dilemma problématique. Theories of trade agreements based solely on market access externalities fail to explain why and how the trade community integrated relatively new issues into the world trade regime with the inception of the WTO in 1995. Consider TRIPS, TRIMs, or the Agreement on import licensing procedures (ILP): The nature of these new agreements has little to do with any market access motivation. GATT and GATS are bilateral market access deals. By contrast, these new Agreements are positive integration contracts geared towards enhancing interaction efficiency of the world trading system.\textsuperscript{328} They contain minimum

\textsuperscript{324} In the special case of interaction between completely symmetrical countries (a convenient but hardly realistic assumption often used in formal models, cf. Bagwell and Staiger 2002b at p. 61), all participating country representatives achieve their political optimum at the same level of market access.

\textsuperscript{325} Policymakers consent to tariff levels, GATS concessions levels, or other commitments in various sectors so as to maximize their selfish, subjective utility functions in the respective arguments. Since liberalization in each sector inevitably pits exporter- and using-industry welfare against that of import-competing interests, the optimal level of liberalization varies depending on the private-sector power balance in each industry. That the politically optimal depth of this liberalization would be free trade is highly unlikely, and if so, of a purely coincidental nature (confer footnote 277 above).

\textsuperscript{326} Non-discrimination provisions are transaction cost efficiency-enhancing tools, but they do not change the inherently bilateral logic of market access concessions.

\textsuperscript{327} That trade concessions are exchanged (and therefore owed) on a bilateral basis is a consequential finding, which warrants a bit of elaboration: It is in the self-interest of every policymaker to extract the most extensive trade concessions possible from other contracting parties. She therefore only cares about how many concessions she has to “give away” in return. This logic is inherently bilateral. Whether or not a party \(B\) frustrates previous market access commitments of another third party \(C\) is inconsequential to the policymaker who represents party \(A\), as long as this measure does not infringe on her own utility. The right to trade is not owed to the collective membership, but directly to every signatory party. The market access obligation is consequently organized in a web of bilateral deals.

\textsuperscript{328} Positive integration norms are based on a “thou shalt...” (prescriptive) logic, whereas negative integration norms are based on a “thou shalt not...” (prohibitive) logic. Positive norms mandate the establishment of a certain result or effect, while negative norms prohibit certain behavior or outcomes.
standard obligations that mandate the quasi-ubiquitous introduction of certain institutional features and procedures – independent of the state of implementation in other countries. Positive integration contracts aim at creating new markets, or to enhance the general efficiency of the existing ones. TRIPS concessions, for example, are far-reaching promises to have in place a set of minimum standard rules.\footnote{329}

Trade agreements based on minimum standards differ from a market access explanation of the WTO in three important aspects:

(i) the contractual intent;

(ii) the underlying problem; and

(iii) the nature of commitments.

(i) \textbf{Different contractual intent.} Whereas trade agreements based on market access externalities are concluded so as to achieve \textit{transaction efficiency} (\textit{ex post} efficiencies from exchanging goods or services), these new trade topics in the WTO can be traced back to a desire to ameliorate the general trading environment. “Law-harmonizing” efforts by Members (Petersmann 2002, p. 51) in the areas of intellectual property rights or transnational investment are geared towards setting core standards for international trade. Minimum standards are apt to ameliorate the underlying conditions in which world trade takes place, and to raise the general efficiency of the world trading system.\footnote{330} Thus, minimum standard agreements are geared towards achieving \textit{transaction cost efficiencies} in a repeated-interaction setting.\footnote{331}

(ii) \textbf{Different underlying problem.} The basic problématique which a minimum standard contract is able to solve may not be found in a PD, or any other \textit{collaboration} game, but effectively in a \textit{coordination} game: Every signatory understands that \textit{some} sorts of legal harmonization and \textit{some} minimum level of policy coordination is beneficial to

\footnote{329} Other “new WTO issues”, such as competition, investment, trade facilitation, labor and environment, belong to the same category of positive integration rules. These new issues have not yet been cast into WTO Agreements, but have been on the bargaining table for some time (for in-depth introduction into new trade policy issues, see Hoekman and Kostecki 1995; Jackson 1997a, pp. 305-18; Trebilcock and Howse 2006).

\footnote{330} By concluding trade contracts based on minimum standards, policymakers are far from acting altruistically. Rather, self-interested trade decision-makers can be assumed to act under duress from important special interest groups. According to some authors (e.g. Harms et al. 2003; Odell and Sell 2006) this is exactly the dynamic that led to the conclusion of some multilateral trade agreements, such as GATS, TRIPS and TRIMs: Self-interested private firms or lobby groups formed cross-country coalitions, negotiated with self-interested public entities in affected countries, and so pushed the formation of these accords.

\footnote{331} As shown in Chapter 2.2.2, \textit{transaction cost efficiency} can be achieved by the conclusion of a contract. Parties avail themselves from having to re-bargain the terms of a contractual exchange over and over again. The contract sets binding standard operating procedures, which reduces the costs of interaction.
everybody. Every transactor thus has an inherent interest in a binding agreement on general efficiency-enhancing measures. The contentious issue, however, over which parties are in disagreement, is the optimal level of positive integration. The contract thus fixes this negotiated level.332

(iii) Different nature of commitments. Finally, the nature of commitments in trade agreements based on market access externalities is different from the kinds of promises that are made pursuant to TRIPS, TRIMs, or IPC. These Agreements are based on prescriptive norms that mandate WTO Members to prepare the legal institutional grounds for frictionless and orderly interaction in the fields of intellectual property, investment measures, government procurement, etc. As will be argued in more detail infra, minimum standard promises are not bilateral, but multilateral, because if one country fails to live up to its promises, the efficiency of the entire system suffers. Positive multilateral integration is probably more difficult to achieve than the negative integration used in GATT and GATS: Negative integration in the trade liberalization agreements of GATT and GATS proceeds via the removal of trade barriers in bilateral deals that are then applied in a non-discriminatory way to all other Member States. Bargaining is not constrained by the decision rule of unanimity, since no Member is under an explicit obligation to lower its tariffs. Positive integration agreements, by contrast, can only be achieved with the formal approval of all WTO Member States (de Bièvre 2004, p. 4).

Our contention that minimum standard agreements are akin to coordination games may be contested. Consider the case of IP protection: It may be argued that some countries profit from IP infringements (e.g. China’s repeated copyright encroachments), and that TRIPS was concluded precisely to forestall those kinds of opportunistic behavior. Critics may prefer the “stag hunt” (footnote 102 in subsection 2.3.1) to the “battle-of-sexes” metaphor to explain the rationale for TRIPS. We do not contest the fact that countries may gain from temporary opportunistic IP infringements. Our point is rather that every WTO Member has an interest in some common level of IP protection. That China’s optimal level is potentially lower than the United States’ is another, separate, story.
4.2 Primary rules of contracting: Basic entitlements in the WTO

In the last section it was argued that the WTO as we know it today is best understood as a multiple-objective, political agreement among self-interested policymakers. The two main motivations for contracting are:

- market access externalities; and
- minimum standards in international trade.

This pair of objectives for entering into a trade agreement is converted into treaty language by putting in place two families of primary rules of contracting: substantive and auxiliary entitlements (see subsection 2.2.1). The contract’s paramount entitlement, which we call the “market access”, or “trade” entitlement, is exchanged bilaterally between contracting party dyads. The market access entitlement consists of mixture of a substantive trade liberalization entitlements and dependant auxiliary entitlements. Other substantive entitlements are multilateral in nature and shall be called “minimum standard” or “positive integration” entitlements. A group of basic auxiliary entitlements geared towards rendering international trade more efficient shall also be introduced. They, too, are owed multilaterally.

4.2.1 The bilateral market access entitlement

The market access entitlement is a WTO Member’s right to compete fairly in its trade partner’s or partners’ market up to the degree granted by each of those countries. The entitlement is composed of substantive and contracting obligations. Substantive obligations define the reciprocal trade liberalization commitments in the form of tariff cuts and service concessions in the four GATS service modes of supply. These substantive entitlements are accompanied by contracting provisions (auxiliary entitlements) to maintain and safeguard the agreed-upon level of bilateral cooperation.333

- Substantive trade liberalization concessions are the currency of any trade contract. It lies in the contractual nature of the WTO that the trade entitlement is reciprocally owed. Each contracting party enters into a bilateral trade liberalization deal with every other WTO Member, and fixes the level of market access it is willing to grant in
The WTO as an incomplete contract

return for access to the foreign market.\footnote{Some WTO scholars, particularly those doing formal modeling, regularly portray the WTO/GATT as a \textit{tariff}-exchange contract. This is inaccurate. The market access entitlement is not just an entitlement to a bilateral exchange of \textit{tariffs}, but one of reciprocal market access.} The level of market access is equivalent to the \textit{size of the promise} – the number of sectors signatories are willing to liberalize, and the degree of market-opening they agree to be bound to. Trade liberalization concessions are constituted by the compulsory tariff bindings (the \textquotedblleft schedules of commitments\textquotedblright{} regulated in Arts. II and XXVIII\textit{bis} GATT) and by positive GATS concessions in the four service modes.\footnote{See our discussion in footnote 327 above.}

- However, the market access entitlement is more than just the substantive commitment to trade liberalization. It also covers \textit{every} modality that may influence the mutually agreed trade balance. Integrated into the trade entitlement are auxiliary entitlements in the form of:
  - non-discrimination stipulations (e.g. Art. I and III GATT);
  - a prohibition of quantitative restrictions (Art. XI GATT);
  - codes of conduct detailing how to deal with non-tariff barriers (e.g. Art. III GATT, SCM, TBT, SPS, GPA, ROO\footnote{The acronyms stand for the \textquotedblleft Agreement on Subsidies and Countervailing Measures\textquotedblright{}, the \textquotedblleft Agreement on Technical Barriers to Trade\textquotedblright{}, and the \textquotedblleft Agreement on Sanitary and Phytosanitary Measures\textquotedblright{}, the \textquotedblleft Agreement on Government Procurement\textquotedblright{}, and the \textquotedblleft Agreement on Rules of Origin\textquotedblright{}, respectively.});
  - explicit exceptions to the right to compete in foreign markets also form part of the trade entitlement: Examples of exceptions include GATT Arts. IV (on cinematography), XVII (on state trading), Art. XXIV GATT (on preferential trade agreements), XXI (on national security), the \textquotedblleft Enabling Clause\textquotedblright{} (based on GATT Arts. XXXVI-XXXVIII), and waivers (Art. IX of the WTO Agreement).

\subsection*{4.2.2 Minimum standard entitlements}

The bilaterally owed trade entitlement is crucial, though of course not the only entitlement exchanged in the WTO contract. As Pauwelyn (2006 at footnotes 91 and 93) aptly states, there are other, \textit{non-market access-related} entitlements in the WTO which are not explainable by the logic of reciprocal market access concessions. As pointed out above, the so-called new issue areas of the WTO, as laid down in the multilateral TRIMs...
and TRIPS, the ILP, and other plurilateral WTO Agreements are based on a motivation quite distinct from reciprocal market access exchange. Their objective is to design positive concessions which prescribe every participating Member to adhere to an agreed set of minimum legal standards. Positive integration rules mandate the introduction of certain institutional features and procedures – independently from the state of implementation in other countries.

What is the nature of minimum standard entitlements? What sets them apart from the bilaterally owed market access-based trade entitlements that form the backbone of GATT and GATS? Minimum standard entitlements are owed to the contracting community as a whole, that is, they have a multilateral ambit. Their \textit{erga omnes partes} logic is distinct from a bilateral logic.\footnote{Multilateral obligations are sometimes called \textit{obligations erga omnes partes} (cf. Pauwelyn 2001), since they are owed to all Members alike. This nomenclature is, however, slightly inaccurate and misleading, because \textit{as a matter of positive law} (by virtue of Art. 3.4 DSU), \textit{all} WTO rights and obligations are \textit{de iure} applicable to the whole membership. Based on this insight, some scholars have argued that reciprocity and bilateral market access-related rights and obligations are a thing of the past, and that all contemporary WTO-legal obligations have a multilateral ambit (Charnovitz 2001, 2002a; Pauwelyn 2000; Jackson 2004). We disagree. As we have shown, the \textit{logic} of the market access entitlement is essentially bilateral (cf. footnote 76/327), even though it is \textit{de iure} owed to the entire membership. As a convention, we will use the term \textit{erga omnes partes} only for those entitlements, whose contractual and systemic logic is inherently multilateral in nature, i.e. that are \textit{de facto} owed to all Members at the same time.} For reciprocal entitlements the \textit{rights} of one contracting party constitute the \textit{obligations} of the other. Transactors exchange entitlements by securing commitments in return for the ones given. A multilateral entitlement mandates every contracting party to behave in a certain manner (for example that every WTO Member establish a national patent office): The entitlement is not exchanged, it is \textit{owned by} the entire community of signatories, as well as \textit{owed to} the entire WTO community. Multilateral entitlements oblige every signatory to abide by the same “rules of the game”. If a country violates its multilateral entitlement to, say, supply patent protection by refusing to establish a patent office, that country impairs the competitive opportunities of all other Members. It brings down the general level of cooperation and harms the system as a whole.

\subsection{4.2.3 Basic auxiliary rules of entitlement}

Basic, or multilateral, \textit{auxiliary} rules of entitlements are phrased as \textit{positive norms}, i.e. as contractual prescriptions that rarely leave any degrees of freedom or discretion to transactors.\footnote{Focus is put on important, non-dependant \textit{primary} auxiliary rules of contracting here. Dependent auxiliary rules of entitlement protection are discussed in the next chapter.} They, too, are multilateral by nature in that they oblige all contracting
parties to the same degree. Multilateral auxiliary entitlements are social ordering devices that supply the trade agreement with a fundamental structure necessary to facilitate the underlying exchange that the substantive entitlements ensure.

Yet in contrast to substantive entitlements, basic auxiliary entitlements carry no proper transactional motivation or contracting rationale. They do not change the net balance of contractual exchanges laid down by substantive commitments. They are merely guided by a desire to ameliorate the general trading environment, the basic conditions in which world trade takes place. We propose to distinguish four types of basic auxiliary entitlements in the WTO:

**Procedural rules.** The purpose of procedural rules is to organize and structure a contractual relationship in a sensible manner. The WTO contains a whole range of rules that delineate organizational issues, such as time-lines (for example those laid down in Arts. 4, 8, 16, 17, 20 and 22 of the DSU), rules of decision-making (Art. IX of the WTO Agreement), voting and selection procedures (e.g. Art. 8.4 DSU on dispute panel composition), special procedures (e.g. on interpretations laid down in Art. IX.2 of the WTO Agreement), disclosure requirements (such as those in Art. XXIV.7 GATT, or Arts. 3, 4, 10, 25 of the DSU), or general provisions of *modus operandi* (such as Art. XXVIII*bis* GATT on initial tariff negotiations, Arts. XXII, XIV and XV WTO Agreement on “accession”, “acceptance, entry into force and deposit”, and “withdrawal”, respectively).

**Transparency entitlements.** Transparency entitlements are obligations spread across the range of WTO Agreements. They safeguard the observation of basic rules of the game, and are aimed at reducing unnecessary search and error costs in connection with economic exchange.\(^{339}\) We find Wolfe’s (2003) classification helpful, which groups transparency provisions into: (i) tariff and services schedules which codify Members’ commitments (e.g. Art. II GATT or Art. XX GATS); (ii) the Trade Policy Review Mechanism; (iii) publication and notification (e.g. Art. X GATT on “publication and notification”).

---

\(^{339}\) While a transparency provision in a particular WTO Agreement may be concerned only with the narrow range of measures covered under that Agreement, such as subsidies or SPS regulations, the cumulative effect of these provisions is to diminish the opaqueness of a Member’s trade regime and trade policymaking process. Transparency in the WTO helps achieve two objectives: “The first is improved compliance by the parties to the commitments they have made under the trade agreement. Shining a spotlight on the parties keeps them on the straight and narrow. Second, going beyond the parties to the agreement, transparency should help private economic agents better understand the environment in which they operate and enable them to make better decisions. Markets will function more efficiently when economic agents have better information” (WTO 2007, chapter II.C.6.b).
The WTO as an incomplete contract

administration of trade regulation
(iv) internal transparency which ascertains transparency of the institution to its Members, (v) and – less interesting for our purposes – external transparency to civil society.

Obligations owed to the institution. Another kind of basic auxiliary entitlements are Member obligations owed to the institution itself. Yearly financial contributions by Members to the WTO may count as an example, or the obligation to assign trade experts to serve on dispute panels.

“External” entitlements. International trade has crucial ties to many other activities of international concern (Pauwelyn 2003). Therefore, public international law is the relevant “playground”, or domain, in which WTO Members are free to contract within. Multilateral external entitlements constitute and delineate the limits of contract freedom in international trade law. Of special importance are peremptory norms of international law (ius cogens). Peremptory norms are fundamental principles of public international law. They have acceptance among the international community of States as a whole, not only for WTO Members. Unlike norms of customary international law that can be modified/changed by mutual consent, WTO Members must not “contract around” ius cogens norms. Under the Vienna Convention on the Law of Treaties, any treaty in violation of a peremptory norm is null and void (Art. 53 VCLT).

4.2.4 The prominent role of the market access entitlement

By means of a summary, Chart 4.3 categorizes the nature of WTO entitlements along the bilateral/multilateral dichotomy and along the substantive/auxiliary divide. Three types of entitlements can be distinguished: First, the market access, or trade entitlement, consisting of substantive reciprocal tariff liberalization commitments, and dependant

340 Notification provisions are, in terms of their number, the most commonly found transparency mechanism in the WTO. Notifications mandate signatories to inform other Members about the enactment of a new legislation, the adoption of new measures, or the progress made in the implementation of commitments. A number of Agreements also contain obligations to publish information on a relevant measure, or establish official points of contract so that Members know how to obtain the relevant information.

341 Internal transparency has less to do with clarification of the provisions in the WTO than with eliciting the process of decision-making in the institution or organization itself. The objective of internal transparency is to make decision-making in the WTO transparent and genuinely inclusive, especially for the smallest and poorest WTO Members.

342 External transparency refers to the WTO’s efforts to engage with civil society groups and the public at large. Much of this involves making official WTO documents, reports, submissions by WTO Members and trade-related statistics available.

343 Generally accepted ius cogens norms include prohibitions on waging aggressive war, crimes against humanity, war crimes, piracy, genocide, slavery, and torture (Malanczuk 1997, p. 375).
of WTO entitlements that specify, safeguard and maintain the mutual level of market access. Second, various minimum standard entitlements, which are substantive and multilateral in nature and apply to areas like intellectual property protection or trade-related investment measures. Third, numerous basic auxiliary entitlements, which are multilateral and auxiliary primary rules of contracting. Minimum standard- and basic auxiliary entitlements are together termed multilateral, or non-trade entitlements.

Chart 4.3 Overview of primary entitlements in the WTO

<table>
<thead>
<tr>
<th>Nature of entitlements</th>
<th>Bilateral (reciprocal)</th>
<th>Multilateral (erga omnes partes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substantive</td>
<td>Mutual trade liberalization concessions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Tariff schedules</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Service concessions in 4 GATS service modes of supply</td>
<td></td>
</tr>
<tr>
<td>Auxiliary</td>
<td>Specify/maintain/safeguard market access</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Arts. III, XI GATT, e.g.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- TBT, SPS, ROO, e.g.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Exceptions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic (or multilateral) auxiliary entitlements</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Rules owed to organization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Procedural rules</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Transparency rules</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- External rules</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Market access entitlement”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Trade entitlement”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Multilateral entitlements”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Erga omnes partes entitlements”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Non-trade entitlements”</td>
<td></td>
</tr>
</tbody>
</table>

Source: author
Notes: Chart 4.3 gives an overview of the three basic entitlement types exchanged in the WTO. Minimum standard- and basic auxiliary entitlements are hereby formed into a group termed “multilateral-” or “non-trade entitlements”.

Of the three types of WTO entitlements the reciprocal market access obligation bears the greatest significance for contracting parties – and as an object of research. The following reasons motivate this assertion:

- The trade entitlement is “self-standing” – it fully justifies a contract. Every contract has a motivation and is concluded to solve a problem (fulfills a motivational and an assurance function). The trade entitlement is a direct consequence of the most important rationale for entering into international trade agreements, namely that of
constraining market access externalities (see also subsection 4.1.3.1 above). Multilateral auxiliary entitlements, however, neither represent an objective to contract, nor do they solve a problem that would warrant the conclusion of a trade accord. No transactor concludes an agreement just to prescribe some procedure or other. Minimum standard entitlements, like those mandated by TRIPS, contain a contracting motivation, and may even solve a collective action problem (subsection 4.1.3.3). Yet minimum standards are no “stand-alone” reasons for concluding a trade agreement. The TRIPS Agreement is not fathomable without a prior market access contract like GATT or GATS: In the absence of a binding agreement on how to organize the international trade of goods and services, no intellectual property, investment, or import licensing framework would be needed.

- **The trade entitlement is an extensive commitment.** While, say, a procedural entitlement to obey a certain notification requirement is relatively limited in significance and scope, the trade entitlement is extensive and far-reaching. It is equivalent to the promise to grant trade partners a certain level of market access – across a wide range of sectors and industries. A market access commitment affects thousands of tariff lines (or service concessions), and it pertains to a great number of domestic non-tariff measures that are prone to partially undo the market access level (such as subsidization, health and safety requirements, technical standards, rules of origin). It is fair to say that keeping up the level of market access for one’s trade partners significantly reduces the policy discretion of domestic policymakers.\(^{344}\) Therefore, market access entitlements are profound, across-the-board concessions by signatories.

- **Only trade entitlement violations are enforceable.** As we showed in Chapter 2.1.3 above, an integral part of a contract is credible enforcement. Self-enforcement lies in the nature of the WTO as a treaty.\(^{345}\) Currently, the WTO has no other ultimate enforcement provisions in place than the “suspension of concessions or other obligations”. Therefore, a violation of all other entitlements traded in the WTO must

---

\(^{344}\) In fact, the trade entitlement is so pervasive that its limits are extremely difficult to delineate: Doesn’t *every* domestic policy measure directly or indirectly affect trade? Don’t policymakers, who have some domestic regulatory objective *other than trade* in mind (such as the reduction of mortality or criminality), often risk unhinging the agreed-upon market access balance through domestic regulation? We will address inadvertent market-access-infringements *infra.*
be avenged through partial withdrawal of market access vis-à-vis a culprit Member. This is yet another sign that the trade entitlement is at the core of the WTO treaty.

If we are to understand the WTO from a contractual perspective, we have to understand the fundamentals of the contract formation phase: The locus of decision-making, the general objectives of agents, the rationale for the conclusion of a contract, and the primary contractual entitlements exchanged. Sections 4.1 and 4.2 discussed the motivation of self-interested policymakers for cooperating, and why they prefer doing so in the form of an explicit international trade agreement. Examining the three fundamental entitlements that contracting parties are likely to exchange in pursuit of their contracting objectives, it was shown that some contractual obligations follow a transaction-oriented, bilateral logic, while others follow an *erga omnes partes*, multilateral logic. Reciprocal market access is the main rationale for entering into a trade agreement. Consequently, the reciprocal contractual obligation to grant market access is the most important (though not the only) promise that WTO Members exchange. Positive integration norms enhance the contracting efficiency of the global exchange of goods and services.

**4.3 Establishing the WTO as an incomplete contract**

The fact that the WTO is an incomplete contract is not new, and has been suggested by various WTO scholars. Yet few authors have really laid out the sources of the Agreement’s incompleteness, and how it affects the making and shaping of the WTO treaty. Based on the theoretical findings collected in Chapter 3, we want to examine thoroughly what types of incertitude are present, acute, and relevant for the WTO Agreement. We locate the source of the contractual incompleteness which the WTO is fraught with and which signatories need to anticipate at the point of ratification. These

---

345 There is unanimity in the literature that the WTO is a self-enforcing contract among sovereign nations (cf. Keck and Schropp 2007 at footnote 10). In absence of a *supra*-national authority, any contract among sovereign nations must necessarily rely on self-enforcement. “The WTO has no jailhouse, no bail bondsmen, no blue helmets, no truncheons or tear gas” (Bello 1996, p. 417).


347 As stated in section 3.1 above (cf. footnote 113), the WTO literature usually exogenously assumes the presence of some type of previously defined “uncertainty”, and therewith presupposes the existence of contractual incompleteness.
insights will be useful later on for verifying whether the GATT- and WTO framers have chosen the best governance structure to address and overcome these imperfections (Chapter 5). They will also come in handy when we assess the best strategy for dealing with the Agreement’s incompleteness (Chapter 7).

Following the framework for investigating incomplete contracts developed in section 3.4.1, we take a closer look at the contractual interaction in the performance phase of the contract, i.e. in the game that unfolds after the contract was concluded. By characterizing and assessing the nature of contingencies occurring during contract performance, we can extract the prevalent types of incertitude that reasonably rational trade negotiators had to anticipate (or should have anticipated) when they sat down to negotiate the WTO.

The nature of interaction in a trade agreement concluded for reasons of market access externalities and minimum standards is fairly simple. Consider Chart 4.4: At some time $t_1$ self-interested trade negotiators convene to elaborate and write down the basic terms of their trade agreement. They lay out the mutual substantive entitlements and the institutional design supporting the trade deal (auxiliary entitlements and flexibility rules). Shortly thereafter (at $t_2$), or concomitantly, the signatory Members decide on their level of cooperation, i.e. the level of reciprocal market access concessions and the generally applicable level of minimum standards. After the conclusion of the contract, the “trade setting game” plays out: Contracting parties continuously exchange goods and services under the improved conditions of the trade agreement. Domestically, each Member government is constantly engaged in various types of policy-setting. Domestic policies may have (unintended) trade-related international spillovers (more on that infra). Both trade game and domestic policymaking are infinitely repeated, since the WTO is an open-ended, long-term contract.

At a time $t_3$ a political economy contingency occurs (the nature of which will be explained below). Policymakers are reasonably rational actors.\textsuperscript{349} An important facet of

\textsuperscript{348} It is conceptually apposite to treat the negotiation of basic institutional design and the negotiation over depth and breadth of commitments as temporally separated instants: Parties must agree on fundamental contracting principles (such as the MFN principle, non-discrimination, or enforcement instruments) before being able to decide on their mutual reduction of trade barriers and collective minimum standards.

\textsuperscript{349} In subsection 3.1.4 (at footnote 144) reasonably rational players were characterized as considerate (or rationalist) utility maximizers, who seek to take into consideration all accessible information, and carefully weigh the costs and benefits of their actions. Yet they are no rational supercomputers, and sometimes have trouble processing all existing information, or running through complicated scenarios or sensitivity analyses. This makes them prone to human contracting lapses and errors in reasoning. However, reasonably rational players are not irrational: They would never deliberately act against their self-interest.
the assumption that policymakers may not be *perfectly* rational shows in the presence of a Rawlsian “veil of ignorance” (check footnote 245 in subsection 3.4.1.2). Behind this proverbial veil, parties negotiate the terms of the contract without full knowledge of the identity of acceding countries, of their economic significance in the distant future, of their role as injurers or victims, and generally of how future states of the world will impact their well-being.

**Chart 4.4 The nature of interaction in the WTO**

We can now proceed to examine the nature of contingencies, incertitude and incompleteness in the WTO contract. We do so separately for the market access entitlement (subsection 4.3.1) and multilateral obligations (minimum standard and basic auxiliary entitlements; subsection 4.3.2).
4.3.1 Contingencies and incertitude affecting the market access entitlement

The entitlement to trade is the single most important primary rule of the WTO (subsection 4.2.4). Parties bilaterally exchange market access concessions. They negotiate the liberalization of sectors and industries up to a political-equilibrium level of cooperation which is reached once the first government hits its optimal domestic political support balance. The reciprocal balance of market access concessions is not stable, but highly dynamic and volatile. In the following subsections we examine the nature of contingencies, and how they affect the initial negotiated international market access balance between countries.

4.3.1.1 The nature of market access contingencies: Political support shocks with spillover potential

Subsection 4.1.3.2 showed that trade decision-makers, in an effort to maximize their re-election chances, must constantly balance domestic pro-export and pro-protectionism interests. As self-interested utility maximizers they strive for an optimal domestic political support balance at any point in time. However, a domestic political support balance is a fragile affair prone to considerable volatility: The future course of all those variables that directly and indirectly enter the PROF of governments determines the domestic support balance of policymakers.\(^\text{350}\) Since none of these variables are stationary, any change in relevant conditions bears the risk of unHINGING the fragile domestic political support equilibrium.\(^\text{351}\) Hence, in a dynamic environment, policymakers’ objective functions are constantly exposed to a wide range of temporary or permanent, direct and indirect shocks.

A direct shock to a policymaker is any variation of those variables that are an integral part of a policymaker’s selfish welfare function. A direct shock can be any exogenous contingency or event that affects general social welfare, because consumer welfare is assumed to enter directly into the policymakers’ utility function. An example of a direct

---

\(^{350}\) As stated above (footnote 320 and accompanying text), four – partially countervailing – interests may enter into the PROF of trade policymakers: (i) Consumer-, (ii) export-, (iii) import-competing- and (iv) TOT interests. Any state of nature that has an impact on those four interests potentially has an effect on the domestic political support balance of the policymaker.

\(^{351}\) The domestic reverberations of any new situation, or event, on a policymaker’s political support balance are likely to be complex in nature: They crucially depend on the politician’s personal preferences, the composition of special interests, as well as the political and financial clout of affected SIGs. It is almost impossible to predict with certainty how, for example, a decrease in the world price of sugar will affect a Swiss policymaker’s political support balance: Will the agri-lobby push for heightened protection? Will consumers care? Will sugar-using industries make efforts to keep the domestic price low?
exogenous shock is vast unemployment following a technological revolution. Direct shocks can however also have endogenous (or behavioral) causes: Contracting parties may – without any noticeable external change – feel differently about certain issues: A policymaker may feel more altruistic, care less for the general welfare, be more responsive towards campaign contributions and backhanders, etc.\textsuperscript{352}

An indirect shock to the domestic support balance is constituted by any exogenous change in the state of the world that affects domestic SIGs’ demand for, or ability to, lobby for special policy favors. Any external occurrence that has an impact on SIGs’ organization, composition, concentration, economic well-being, political clout, or willingness to “bribe”/influence politicians will affect a policymaker, only if it is prone to upset (improve or deteriorate) the initial domestic political support balance. Examples of indirect political support shocks are price or supply changes that intensify international competition (and may consequently stir lobbying efforts by affected domestic firms); technological breakthroughs that reduce employment in a sector or trigger an export boom; unemployment (which brings workers’ interest to the fore); macro-economic instability; constitutional changes in the country’s political institutions (e.g. a successful campaign finance reform that alters the political pressure that firms can apply); or changes in political cleavages or alignments which might make a previously pivotal sector less influential in domestic politics (for example due to changes in political majorities).\textsuperscript{353}

Yet (in-)direct domestic political support shocks of the kind explained above are not quite the contingencies that concern market access entitlements. A contractually relevant contingency is present only if, subsequent to the revelation of a state of nature, one or more policymakers are impelled to react (in the form of a trade-related policy instrument). Many political support shocks are absorbed domestically and do not create new international market access externalities. If however, upon the incidence of a domestic political support shock, a decision-maker is willing to undertake actions that are apt to upset the initially negotiated balance of market access concessions, it will give rise to a proper “market access contingency”. Market access contingencies are thus a

\textsuperscript{352} A policymaker may consider asbestos a health hazard, while feeling all right about tobacco – and change her mind next month. People may reconsider issues due to a change in perception, to successful learning, to sudden enlightenment, or because new information is processed differently. But these underlying factors are secondary here. What counts is the result: A novel political-support situation.

\textsuperscript{353} On the nature of political shocks, see e.g. Rosendorff and Milner (2001, p. 832), Rosendorff (2005, p. 392), Ethier (2001a, p. 10), Bown (2002a, pp. 295), and Hauser and Roitinger (2004, p. 654).
subset of domestic political support contingencies; they are political support shocks with spillover potential.

Some remarks are appropriate here to clarify our understanding of the concept of market access contingencies:

- **Market access contingencies can be exogenous or endogenous.** Just as political support contingencies can have an external or internal origin, market access contingencies can also be exogenous or endogenous. Parties may react in a trade-related manner to some external shock, or act due to some psychological, emotional, or behavioral contingency.

- **An event is not a contingency.** A specific incident or event (say, a technology-related demand shock in country $A$ causing an externality in country $B$) is more than a state of nature, or contingency. Events are composed of a bundle of concurring contingencies. We may think of an event as a vector of contingencies, where time, duration, location, identity of affected parties, etc. are all decisive.\(^{354}\)

- **Regret contingencies and opportunistic contingencies.**\(^{355}\) Opportunistic contingencies are those that, if acted upon, will lead to inefficient redistribution. If a signatory seizes an exogenous or strategic contingency of this kind, it enriches itself at the expense of its trading partners.\(^{356}\) Regret contingencies are those that, if considered and acted upon, may lead to general welfare-improving behavior.

- **Market access contingencies are policy situations of a certain degree of trade-relatedness.** Ethier (2001a, p. 7) refers to market access contingencies as “policy situations of a certain trade-relatedness”. This statement contains two messages: (i) The political nature of contingencies, and (ii) the concept of trade-relatedness. Trade deals are inherently political in nature, since economic concerns enter but indirectly

\(^{354}\) Hence, the occurrence of a seemingly equivalent event, for example a technological demand shock, in a different context (at a different time, in a different country, under a different political support constellation), may have not at all equivalent consequences, since the event “technology shock” really consists of a completely different set of contingencies. Sensitivity to contingencies gets more intricate the more complex the underlying contractual situation (number of actors involved, longevity, depth and breadth of promises, etc.).

\(^{355}\) See our related discussion in footnote 147.

\(^{356}\) An example may be a protectionist shock in the EU sugar industry. One or more EU Member(s), feeling under intense political pressure to increase the tariff barriers for sugar imports may want to submit to that urge, since doing so promises increased political support. However, the tariff increase would certainly provoke TOT effects, and market access losses to foreign exporters, such as Brazilian and Australia, whose policymakers would stand to lose more (measured in political support) than EU-politicians would stand to gain.
into decision-makers’ personal objective functions via the well-being of influential stakeholder groups. Thus, even if a market-access-relevant state of nature has an economic, technological, societal, or natural origin, only its effect on the policymaker’s political support function matters to her.\textsuperscript{357}

The impact of a domestic policy measure on a foreign country depends on the level of trade-relatedness of a contingency: Actions by one country might mainly be intended to address domestic (non-trade) issues, but happen to slightly affect the international trade balance. The relevant market access contingency thus displays a low trade impact. Other shocks entail a strong trade impact, such as a protectionist policy reaction to an economic depression (think of the dynamics unfolding after the Great Depression and the Smoot-Hawley tariff in 1929 and 1931).

• **Unintentional spillovers are possible.** The few economic models of the WTO that consider a non-stationary world usually model market access contingencies as protectionist “preference shocks” (Bown 2002a, pp. 295): Contingencies are assumed to be such that they provoke a direct, unambiguous, political pressure for protection at home, or for more open markets abroad (e.g. Bagwell and Staiger 2005b; Bown 2002b; Copeland 1990; Furusawa 1999; Herzing 2005; Hungerford 1991; Kovenock and Thursby 1992). In short, it is assumed that external shocks are *direct protection* market access contingencies. It should be noted, however, that governments, in pursuing their objective of achieving the best possible domestic political support balance, do not exclusively try to craft *trade* policies, but in fact shape *any kind* of domestic policy measures with the objective of maximizing political support.\textsuperscript{358} In fact the majority of a country’s policies are primarily geared towards domestic objectives. Yet some of these policies have unintentional trade impacts. Various domestic sanitary, health, environmental or consumer-safety policies may be geared towards domestic non-trade objectives, but are somehow trade-related. As

\textsuperscript{357} The fact that trade-related market access contingencies are *political* in nature is confirmed by most state-of-the-art economic models of trade agreements, and even by Bagwell and Staiger’s politically enhanced TOT approach (2002b, chapter 2): In their model, the domestic optimal tariff (τ^{PO}) will be affected by *any* contingency that is prone to disrupt a policymaker’s preferred domestic price. A change in τ^{PO} is then apt to change the global TOT-structure. For expositional convenience, however, the Bagwell and Staiger assume policymakers’ target tariff (τ^{t}) to be *static.*

\textsuperscript{358} Political economy literature on endogenous trade-making (cf. footnote 268 supra) assumes *any* government policy to be trade-related solely for the purpose of convenience.
Ethier (2001) rightly contends, any redistributive domestic policy may provoke international spillovers (cf. footnote 344 above).³⁵⁹

The incidence of market access contingencies shows that an initially negotiated market access balance is far from stable. In fact, in a non-stationary world the mutually agreed balance of market access entitlements between two self-interested policymakers is in constant flux. Many efforts undertaken by a trade policymaker to bring into balance her domestic political support structure may – intentionally, inadvertently or due to negligence – entail spillover effects on the international market access balance.

4.3.1.2 How contingencies affect the entitlement to trade: The nature of incertitude and the resulting type of incompleteness

Intuitively, in a complex contract such as the WTO – featuring many players and repeated interactions, complex domestic settings, many issue areas, and a dazzling number of liberalized sectors – it is impossible for signatories to consider in advance every possible state of the world in the contract. Based on our previously developed typology of contractual incertitude (see Chart 3.2 above), we are now prepared to lend that intuition a bit more structure and scientific rigor.

The following paragraphs elucidate the types of incompleteness that the market access entitlement is encumbered with. To that end, we detect the underlying categories of incertitude that burden the WTO contract. We follow the structure illustrated in Charts 3.2 and 3.3, searching for answers to the suite of questions: (i) can contingencies be defined/specified in advance?; (ii) can their outcome be defined?; (iii) is the probability density function known?; and (iv) is the contingency symmetrically revealed?

(i) Can market access contingencies be forecast? If WTO signatories could anticipate and specify all the exogenous and endogenous contingencies that might lead to a disruption of the mutual market access balance, they could write a complete contingent contract. To achieve that end, contracting parties would need to foresee and consider all those myriads of contingency-bundles that have a potential bearing on the market access

³⁵⁹ Applied to the context of market access contingencies, this means that inadvertent spillovers make the transnational market access balance even more volatile, since more domestic actions are apt to inadvertently provoke international spillovers.
balance – even if they occur with the slightest of probabilities. Not only would parties have to foresee the origin, nature, and domestic impact of an endogenous/exogenous market access contingency, but also anticipate the domestic reverberations on the injuring governments’ domestic support balances, policymakers’ likely reactions, and the trade-relatedness that these will cause. In addition, signatories would need to devise and prescribe the optimal response to be taken by every affected party.

We have to conclude that WTO signatories are not able to forecast all possible market access contingencies. Horn, Maggi and Staiger (2005; 2006) have convincingly shown that even fully rational and perfectly capable transactors prefer living with uncertainty over future events to having to negotiate a completely state-contingent contract. The transaction costs of adding a contingency to the WTO can easily outweigh its benefits.

(ii) Can parties predict the outcome of a market access contingency? Following the lead of subsections 3.1.2 and 3.1.3 (cf. Charts 3.2 and 3.3), the next question to tackle is whether possible outcomes can be foreseen by signatories or not. The answer is yes, because the effect can be detected. The outcome of a market access contingency is the partial closure (or opening) of trade flows. Although affected parties may not understand exactly what caused one Member to enact a market-closing (or opening) trade policy instrument, they will be quite aware of the consequences, especially if they are negative, i.e. market-closing. Affected SIGs in victim countries, in particular, will do their best to bring a loss of their competitive position in export markets to the attention of domestic trade decision-makers.

(iii) Can signatories define the probability of occurrence? Here, the answer is “maybe, but not likely”. We saw domestic political support contingencies become market access externalities only if a signatory “acts out” on the shock, that is, enacts a trade-related policy instrument. We submit that it should be quite hard for signatories to predict the probability with which certain market-opening, or rather, protectionist outcomes will occur in the future. Contracting parties may assume that market access contingencies take place along some probability distribution (e.g. normal distribution), but to actually know the probability density function of outcomes should be extremely difficult for signatories, given that they cannot foresee contingencies properly.

---

360 As shown above, there is a myriad of political, technological, economic, and environmental contingencies in a non-stationary world, all of which may have an impact on the welfare of consumers and of special interest groups – and therewith on the international political support balance.
(iv) Is the information symmetrically observable? The last question to address is whether contingencies are revealed symmetrically or asymmetrically. This question is especially important in relation to the opportunistic/regret dichotomy. Whenever market access contingencies are comprised of generally accessible information, the party affected by the domestic political support shock has little opportunity to strategically misrepresent the truth. Yet market access contingencies, it is safe to say, are privately revealed to the party affected by a shock. This party reacts to some domestic political support shock with a protectionist policy measure, and the victim of that measure has little means of knowing (or finding out) what caused the injurer to act: an opportunistic contingency or true regret.361

What kinds of incertitude is the market access entitlement burdened with? Chart 4.5 (overleaf) summarizes our findings by revisiting Chart 3.3 and highlighting the relevant paths. We see that WTO Members must deal with “unforeseen contingencies”, since it is impossible to anticipate, define, agree on, and write down all possible market access contingencies. Transactors prefer to leave certain contingencies unmentioned and certain policy measures unconsidered, due to cost-induced inefficiencies connected with considering all relevant contingencies and having to prescribe the adequate policy responses (or due to constraints on rationality, for that matter). Signatories may or may not take up the effort to work out the probability of some contingency outcomes. Due to transaction costs, however, they will have to leave contingency outcome space uncovered. Despite this, any market access contingency is most certainly private knowledge to the party affected by a political support shock. Therefore the market access entitlement is beset with one of two kinds of incompleteness: Necessary (or type B), or inexorable-type B incompleteness, depending on whether parties are willing to research the probability density function of contingencies and write a contractual provision accordingly.

361 To see why this is so, consider the balance-of-payment (BoP) contingency measures in the WTO. GATT Articles XII and XVIII allow for the enactment of trade restrictions for BoP purposes. The regret contingency “BoP crisis” (actually an event) is private knowledge to the domestic administration; it is not symmetrically known to every WTO Member. How can any outsider know with certainty if there is a threat of “serious decline in monetary reserves” (as Art. XII.2(a) mandates)? No WTO Member, except for the government affected, is likely to know whether a BoP crisis really takes place, and if so what caused it, and what domestic social, political, and economic reverberations entail as a result. The affected country may appeal to the BoP-clause for legitimate reasons; it may well, however, invoke it for selfish, protectionist reasons. Given this asymmetrical revelation of information, the party wishing to deviate from the Agreement is likely to use its informational edge in order to engage in opportunistic behavior (but see panel and AB reports of the India – Quantitative Restrictions case, WT/DS 90). India, for example, has a long history of making ample and long-term use of the BoP-clause of Art. XVIII GATT. In 1997, the country justified quantitative restrictions on 2700 agricultural and tariff lines with recourse to Art. XVIII.B (cf. Mavroidis 2007, chapter 4.3).
4.3.2 Contingencies, incertitude and incompleteness affecting minimum standard entitlements and other multilateral entitlements

As with market access contingencies, those states of nature that affect previously agreed minimum standards and other basic multilateral entitlements are a subset of policymakers’ political support contingencies: Domestic political support shocks may prompt policymakers to retract from previously agreed minimum standards or other
basic entitlements, such as transparency obligations or timelines. Self-interested policymakers thus wish they had contracted for a different level of concessions, or more/less rigid regulation.

There are, however, two noticeable differences between market access and multilateral contingencies. One lies in the opportunistic/regret contingency dichotomy: Due to the multilateral nature of said entitlements, every backtracking behavior taken in response to a domestic political support crisis must be assumed to be opportunistic, unless there is unanimous consent by all contracting parties about its practicability (cf. de Bièvre 2004, pp. 4). The second difference between market access- and multilateral contingencies resides in the urgency and immediacy of response: Shock situations requiring emergency relief action are less likely to occur with respect to multilateral contingencies. It is difficult to come up with examples of surprise contingencies that would require immediate remedial action with respect to minimum standard obligations, such as the establishment of a patent office. Since multilateral obligations are usually written with a longer-term objective, temporary and short-term shocks will less often warrant immediate relief action that must not be delayed by the injurer.

How do multilateral contingencies affect the minimum standard and other multilateral entitlements? Following the same suit of questions applied to the case of market access contingencies, we find that multilateral contingencies are (i) not predictable, but (ii) their effect or outcome is well recognizable and therefore predictable: It is easy to observe whether minimum standards and other regulation norms have been infringed upon. (iii) The probability of contingency-induced outcomes may or may not be known in advance by signatories, but (iv) contingencies are certainly revealed privately to affected parties only. Hence, minimum standard and other basic multilateral entitlements – just like the market access entitlement – are beset by either necessary or inexorable-type B incompleteness – depending on whether parties can or cannot foresee the probability density function of contingencies. Chart 4.5 thus may equally be applied to incompleteness affecting all non-market access entitlements.

---

362 As stated in footnote 332 above, room for opportunism concerning minimum standard entitlements may be seen as a contradiction to the nature of a coordination game, which we hypothesize to be at the core of the new issue Agreements in the WTO. We believe that the rationale for minimum standards lies not in overcoming opportunism, but in achieving a general level of efficiency. This is different from contending that opportunism will never occur.
4 The WTO as an incomplete contract

4.4 Conclusion: The WTO – an incomplete contract based on market access externalities and minimum standards

Writing a complete contingent trade agreement is not in the realm of possibility for WTO signatories. Even if signatories were fully rational and willing, they could not anticipate all the bundles of political support contingencies that may have an impact on policymakers’ market access and minimum standard entitlements – not least, since unanticipated contingencies are privately revealed to affected parties: Private revelation of previously unforeseen contingencies makes incompleteness of the contract an inevitable consequence. The WTO is fraught with incompleteness of the types previously defined as necessary (type B) or inexorable-type B incompleteness.363 Informed affected parties (injurers) have the option to sham a domestic crisis and (allegedly as a reaction to the crisis) enact a trade-related policy measure that does more harm to the other parties than good to the injurer. Thus, the nature of a trade agreement based on market access externalities and minimum standards is such that it is incompleteable.

Stating that the nature of the WTO is that of a contract of necessary or inexorable-type B incompleteness does not mean that the WTO Agreement is free from the other types of incertitude discussed in Chapter 3.1.3: By trying to overcome and master the original contractual incertitude, WTO Members have introduced other types of incompleteness “through the contractual backdoor”. How did this happen? Subsection 3.2 showed that contracting parties, when designing the contractual governance structure, can make use of one or more of the following strategies: (i) minimize the number of gaps through comprehensive contracting; (ii) draft flexibility mechanisms; (iii) exercise precaution; and (iv) delegate responsibility to a third party.

As we will demonstrate, the WTO Agreement makes use of all of these strategies. In an effort to overcome “natural” incertitude in the form of necessary and/or inexorable-type B incompleteness, Members added additional contract language: Over and above substantive and basic auxiliary entitlements, the WTO framers introduced dependant auxiliary entitlements in the form of supplementary contract language on contingencies

---

363 Refer to our discussion of necessary and inexorable-type B incompleteness in subsection 3.1.3.
and policy measures, trade policy flexibility provisions, and enforcement mechanisms.\textsuperscript{364} Yet doing so came at the cost of \textit{additional incertitude}: Whenever explicit contract language is inserted, room for haphazard gaps, ambiguities, ambivalence, and rigidity is created.\textsuperscript{365} What’s more, the incorporation of additional auxiliary entitlements is likely to create room for unforeseen and even unforeseeable contingencies. Thus, fighting incompleteness may have proven to be akin to “robbing Peter to pay Paul”: One type of incompleteness being addressed by introducing another. The next chapter proceeds to discuss how WTO framers dealt with the treaty’s contractual incompleteness.

\textsuperscript{364} The merits of this endeavor and the actual quality of the Agreement’s governance structure are elaborated in the following section.

\textsuperscript{365} See Box 3.1 in Chapter 3 for an explanation of the dark side of contractual language. Article III GATT is a good example of a contractual clause that utilizes very broad and unspecific language and therewith introduces \textit{additional} incertitude. This article is aimed at curbing discriminatory domestic practices. Yet the instruments affected are neither mentioned nor specified. Ambiguity and opposing interpretations are the natural consequence (Mavroidis 2006 at footnote 17; WTO 2007, section II.C.1.b(iv)).
5 Analyzing the system of non-performance in the WTO

Substantive entitlements are the commitments that form the essence of contracts. Any contract that does not consider every possible contingency is by definition incomplete and must have in place a governance structure that protects its substantive entitlements from un-coordinated and, in particular, uncompensated *ex post* escape. At the same time, the institutional design should leave room for welfare-enhancing *ex post* flexibility. Safeguarding efficient post-contractual discretion is the role of a contract’s *system of non-performance*. To that end, signatories shape secondary and tertiary rules of contracting.366

The benchmark for an incomplete contract is given by the achievable first-best, the efficient “breach” contract (subsection 3.4.2). The EBC mimics the outcome of the hypothetical complete contingent contact. It strikes the optimal balance between flexibility and entitlement protection: Every contractual entitlement is protected in a way that prohibits opportunistic opt-out. This is equivalent to saying that injurers may engage in flexibility so as to seize welfare-enhancing *ex post* non-performance opportunities without harming the victims of contractual escape. Whenever one of these two conditions fails to hold, the contractual system of non-performance is out of balance. If signatories to a contract provide for provisions that are unsuccessful in safeguarding the initially traded level of commitment, a signatory expecting to assume the role of a victim is likely to commit to less cooperation upfront. Un-seized flexibility opportunities, on the other hand, discourage future injurers from engaging in contractual exchange in light of unforeseen or unanticipated contingencies.

This chapter will examine the current structure of trade policy flexibility and entitlement protection in the WTO. We will review how the market access entitlement, the minimum standard entitlements, and all other basic auxiliary entitlements are protected both *de iure* and *de facto* (subsections 5.1-5.3). Based on the insights gathered in Chapter 3 which discussed signatories’ strategies of overcoming contractual incompleteness, subsection 5.4 will examine whether the GATT- and WTO framers have chosen the best

---

366 Secondary entitlement rules lay out the scope and limits of *intra-contractual* *ex post* behavior; tertiary rules of enforcement outline how *extra-contractual* behavior (if discovered) shall be punished.
5 The WTO system of non-performance
governance structure to address and overcome the unavoidable incompleteness of the WTO contract. Assessing flaws and problems in the WTO system of non-performance, we shall conjecture about the consequences of the international trading system’s malfunctioning.

The verdict is not complimentary. The WTO as it stands today is rather far from being an EBC. It will be argued that the current functioning of contractual escape in the WTO does not address the Members’ needs adequately, and consequently crowds out cooperative zeal. The WTO system of non-performance is in dear need of reform.

5.1 Trade policy flexibility and protection of the market access entitlement

The GATT/WTO framers at their time had realized that the market access entitlement must be protected against undesired *ex post* backtracking. Technically, trade liberalization commitments – positive tariff commitments and GATS-concessions – can easily be undone by all sorts of tariff- and non-tariff measures, such as quotas, voluntary export restraints, export subsidies, regulatory red-tape measures, or the erection of technical impediments. *Ex post* trade barriers have the potential to completely vanquish the initial promise.

Yet the trade entitlement is not protected by an inalienability rule that would mandate unconditional, obligatory specific performance. Signatories were well aware that the WTO Agreements must leave injurers with some “breathing space” in case unforeseen market access contingencies occur that would justify temporary or permanent, partial or total *ex post* non-performance.\(^{367}\) To that end, positive market access concessions are supplemented by a protection belt of auxiliary, non-substantive, *contracting* rules. These contracting rules lay out how and to what extent the mutual market access entitlement is shielded from *ex post* discretion. These secondary rules of entitlement dictate who “owns” the contractual gap (should one exist), and under what circumstances and at what

---

\(^{367}\) This was indeed a wise decision of the WTO founding fathers. Many WTO scholars have convincingly shown: A market access agreement beset by unforeseen contingencies which permits trade policy flexibility Pareto-dominates one which rigidly demands mandatory specific performance (Dixit 1987; Downs and Rocke 1995; Herzing 2005; Rosendorff 2005; Rosendorff and Milner 2001, confer also our discussion in subsection 7.2.2 below).
price. This subsection will review the existing governance structure which protects the market access entitlement and authorizes trade policy flexibility.\textsuperscript{368}

\subsection*{5.1.1 De iure protection of the market access entitlement\textsuperscript{369}}

The WTO has various auxiliary entitlements that protect the trade entitlement so as to ensure that each party’s expectations are not frustrated in the performance phase of the trade contract. \textit{De iure} the contractually agreed level of market access is safeguarded by a defense line consisting of four kinds of contracting rules: (1) negative integration provisions, (2) non-violation complaints, (3) contingency measures, and (4) a default rule.

\subsubsection*{5.1.1.1 Negative integration provisions}

Instead of specifying detailed market access contingencies, signatories to incomplete contracts can engage in restricting or encouraging certain \textit{policy behavior} by injuring parties, if four preconditions hold: First, transactors must concentrate on regulating relevant policy instruments which are apt to cause the largest market access impact. To that end, signatories must anticipate the probability of the occurrence of those domestic political support contingencies that prompt injuring parties to enact certain policy measures. Second, the benefits of negotiating and writing down the acceptable policy instruments must outweigh the costs. Third, signatories must make sure that the defined (trade-)policy measures are \textit{effective} in separating opportunistic policies from those policies that are driven by real regret.\textsuperscript{370} Fourth, by writing down permissible and prohibited policy instruments, signatories must be careful not to restrict the sovereign policymaking discretion of policymakers. Domestic policies may be enacted for non-contractual reasons (in the trade realm: health, public morals, standards, etc.), yet display

\footnote{\textsuperscript{368} We re-emphasize that, conceptually, the question of trade policy flexibility is largely the flipside of the issue of legal entitlement protection (cf. Dunoff and Trachtman 1999, p. 32): The level of entitlement protection determines the action space of contracting parties. It sets out whether and how parties are allowed to react to changing circumstances that have not been considered explicitly at the time of the contract formation. Analogously, the choice of trade policy flexibility mechanisms lays down what behavior is permissible in case of a \textit{contractual gap}, i.e. in case of an unforeseen/unspecified contingency. So, while flexibility provisions nail down the legitimate (\textit{intra-contractual}) behavior of the \textit{active} party, the entitlement issue is concerned with the scope of protection granted to the \textit{passive} party (see our discussion in section 3.3 at footnote 200).}

\footnote{\textsuperscript{369} The discussion of trade entitlement protection in this chapter shall be reduced to trade in \textit{goods} covered by the GATT (and appending Agreements). Similar instruments for the trade in \textit{services} are provided for in the GATS.}

\footnote{\textsuperscript{370} If some policy measure can be used in response to both regret and opportunistic contingencies alike, its regulation is useless and needs additional definition so that it unambiguously separates permissible from opportunistic behavior.}

- 206 -
significant spillover potential. National decision-makers, however, cannot be expected to forgo their duty to regulate domestic affairs.

WTO signatories have gone down that route of contracting. They have integrated explicit contract language in the WTO Agreements that strives to regulate those policy instruments that are prone to cause market access externalities. These so-called *negative integration provisions* are not flexibility rules. Quite the contrary, for they demonstrate an effort towards contractual completeness. Aimed at preventing injurers from *ex post* escape in certain situations, negative integration rules prohibit those trade-restrictive measures, or policy instruments, that are decidedly enacted with an opportunistic intent.\(^\text{371}\) Examples of negative integration provisions are:

- Rules of non-discrimination (e.g. Arts. I and III GATT);
- Rules dealing with specific clusters of non-tariff barriers (e.g. regulated in TBT, SPS, GPA, or ROO);
- The prohibition of quantitative restrictions, voluntary export restrictions and ordinary marketing agreements (VERs and OMAs; GATT Art. XI);
- The prohibition of export subsidies (Art. XVI GATT and parts of the SCM Agreement);
- The prohibition of unfair trade practices (Art. VI GATT and ADA);
- Directives on customs valuations (Art. VII and the Agreement on Customs Valuation).

Aside from the apparent problems of ambiguity, ambivalence and rigidity that the inclusion of negative integration rules create (see Box 3.1 in Chapter 3), these WTO provisions cannot possibly regulate in advance all those policy instruments that are apt to partially renge on the contract: It lies in the nature of market access contingencies that unforeseen eventualities are bound to happen during the performance phase of the trade contract. In the same vein, WTO Members are able to engage in unanticipated trade-

\(^{371}\) In chapter 3.2.3.1 we called this gap-filling strategy *diligence*. Diligence basically means that transactors make efforts to write the best contract they can. Horn et al. (2006; 2005) call these efforts *discretion* (prohibiting policy instruments). According to the authors, *discretion* stands in contrast to *rigidity* which represents an effort towards anticipating environmental contingencies.
related policy behavior.\textsuperscript{372} Therefore, the market access entitlement is not protected in an absolute way. Signatories possess the discretion to temporarily renge on previously made trade liberalization commitments.

\subsection*{5.1.1.2 Non-violation complaints (NVC)}
Subsection 4.3.1.1 stated that market access contingencies can arise in the form of inadvertent spillovers of domestic policy instruments. A purely domestic policy measure may have trade-related consequences of sorts. NVC, now, protect victims from such unintentional, haphazard spillovers. NVC are described in Art. XXIII.1b GATT and Art. 26 DSU.\textsuperscript{373} The injurer is not under an international legal obligation to withdraw the measure in question. A mutually agreed solution is encouraged. If parties fail to achieve agreement over compensation, the WTO arbitrator under the mandate of Art. 22.6 DSU issues a non-binding damage measure (Petersmann 1991; Mavroidis 2007, section 4.5.6.2).

\subsection*{5.1.1.3 Contingency measures}
Turning from dedicated rules of entitlement protection to explicit provisions of trade policy flexibility, contingency measures were defined in subsection 3.2.2.1 as rules of \textit{ex post} flexibility geared towards seizing the potential of \textit{ex post} regret contingencies. Contingency measures are additional auxiliary entitlements driven by the desire to circumscribe the outcome of certain groups of (previously unspecified) contingencies or events, and to impose in exact terms the permissible action to be taken by signatories in response. Contingency measures lay out the broad contours of regret, without tackling the impossible task of specifying the underlying contingencies themselves.

\textsuperscript{372} Ex ante, signatories are likely to be ignorant of the underlying probability of outcomes provoked by potential market access contingencies. They cannot decide which instruments and states of nature to focus on. Even if they could foresee the probability of the occurrence of certain outcomes, laying down an exhaustive list of prohibited instruments would be prohibitively expensive: Horn et al. (2006) show that \textit{discretion} costs of tailoring instruments can be expected to be extremely high. A thorough assessment has to be conducted, as to whether a policy instrument can have trade-related effects, what its cross-country impact will be, and how harmful this is going to be for affected parties. This process is certainly research-intensive and tedious to negotiate. It is therefore completely rational to leave some, or many, policy-measures unnamed.

\textsuperscript{373} “The idea underlying [Article XXIII:1(b)] is that the improved competitive opportunities that can legitimately be expected from a tariff concession can be frustrated not only by measures proscribed by the General Agreement but also by measures consistent with that Agreement. In order to encourage contracting parties to make tariff concessions, they must therefore be given a right of redress when a reciprocal concession is impaired by another contracting party as a result of the application of any measure, whether or not it conflicts with the General Agreement” (GATT Panel report on \textit{EEC – Oilseeds} at § 144).
WTO signatories crafted the following contingency provisions of flexibility for trade in goods:\footnote{Similar contingency measures exist for the trade in services, regulated by GATS (see footnote 369 above).}

- \textit{Art. XIX GATT} allows WTO Members to take \textit{safeguard action} (by means of the erection of protectionist barriers) so as to react to a sudden, unforeseen surge in imports. The nature and origin of the principal contingencies notwithstanding, Art. XIX allows for adjustment in the event of serious economic pressure. The multilateral Agreement on Safeguards (SGA), which adds detail to Art. XIX GATT, was concluded during the Uruguay Round. At face value, the safeguards clause is a liability-type rule of opt-out: The injurer can decide whether and when to enact them so as to receive temporary relief in case of domestic distress.

In order to be able to perform escape via the safeguards clause, the potential injurer is faced with a substantial threshold of application:\footnote{To recall: The enactment threshold puts down the legal conditions imposed on the injuring signatory. It is usually determined in the form of a series of prerequisites that have to be fulfilled before the contracting parties may engage in \textit{ex post} discretion.} An enacting country must show that “\(i\) as a \textit{result} of unforeseen development; \(ii\) imports in \textit{increased quantities}; \(iii\) have \textit{caused} or threatened to cause; \(iv\) \textit{serious injury} to the domestic industry producing the \(v\) \textit{like product}”.\footnote{Howse and Mavroidis (2003 at p. 686). For specifics of the legal test, and additional restrictions for the enactment of Art. XIX by case law, see also Mavroidis (2007, chapter 4.7.4), or Roitinger (2004 at pp. 102). For example, according to the AB ruling of \textit{US – Wheat Gluten} (WT/DS 166/AB/R at § 55), the injurer, when claiming “serious injury”, must take into consideration \textit{all} relevant factors that determine injury (emphases in original).} Except for “critical circumstances” (Art. 6 SGA), Art. XIX.2 GATT, Art. 3 and Art. 8 SGA oblige the injuring party to give a public notice, engage in consultations with the potential victims (of at most 30 days duration), and to offer the victim adequate trade compensation for the partial withdrawal from its market access obligations. If a mutually agreed solution is not found, the victim is free to suspend substantially equivalent concessions and other obligations.

The application scope of Art. XIX is relatively restricted:\footnote{Safeguard measures can be invoked exclusively in times of economic distress and applied only once. Tariffs and quantitative restrictions are the only permissible trade instruments. The duration of safeguard measures is for a period of four years, although this can be extended up to eight years, subject to the findings of a mandated review panel (Art. 7.1 SGA).} Safeguard measures can be invoked exclusively in times of economic distress and applied only once. Tariffs and quantitative restrictions are the only permissible trade instruments. The duration of safeguard measures is for a period of four years, although this can be extended up to eight years, subject to the findings of a mandated review panel (Art. 7.1 SGA).
principle, safeguards cannot be directed at the country or set of countries that are the source of the injury (by virtue of Art. 2.2 SGA). Instead, they have to be applied on a non-discriminatory basis.\textsuperscript{378} The remedy rule accompanying the opt-out clause is probably most accurately described by the \textit{reliance} damage measure, which mandates a re-establishment of the victim’s \textit{status quo ante} the breach.\textsuperscript{379}

\begin{itemize}
  \item \textit{GATT} Arts. \textit{XX} and \textit{XXI} are contingency measures written with the intention of allowing for flexibility in the event of threats to \textit{public order} and \textit{national security}. Both provisions are also liability rules, but they alleviate the injurer from paying any remedies to the victim.\textsuperscript{380} Using these contingency measures as tools of market access flexibility, however, is constrained by a high level of preconditions, and a restricted application scope: First of all, injurers have to prove that their reason for opting out falls under the \textit{exclusive scope} of the letters (a)-(j) of Art. XX, or (a)-(c) of Art. XXI, respectively. Second, discrimination by intent is foreclosed by the requirement laid down in the chapeau of Art. XX. This is consequential, since it does not allow a policymaker to accord less favorable treatment to \textit{imports} than to \textit{domestic} producers when resorting to the “general exceptions” clause. Such unequal treatment, however, is part and parcel of trade policy flexibility (Roitinger 2004, p. 21). Although this does not mean that \textit{GATT} Art. XX cannot be abused (by enacting protectionist measures under the guise of non-discrimination), this caveat makes the contingency measure of Art. XX \textit{GATT} a difficult tool for \textit{ex post} discretion.

  \item \textit{GATT} Arts. \textit{XII}, \textit{XV}, and \textit{XVIII} on \textit{balance-of-payment crises, exchange arrangements}, and \textit{infant industries}, are also liability-type opt-outs, albeit accompanied by a high threshold of enactment, as well as a narrow scope of application. The purpose of these contingency measures is to protect Members from
\end{itemize}

\textsuperscript{377} The application scope specifies the contractual strings attached to the use or application of a trade policy flexibility mechanism.

\textsuperscript{378} In other words, safeguard measures are a non-selective, or non-discriminatory, flexibility tool. They must be “MFN’ed”, in \textit{WTO} parlance. Whenever safeguards are applied in the form of quantitative restrictions, however, Art. 5.2 SGA on tariff modulation may leave room for selectivity.

\textsuperscript{379} If no agreement on compensatory action is reached, Art. XIX.3(a) allows the affected Member (the victim) to “suspend … substantially equivalent concessions or other obligations.” This demand for commensurate damages has been interpreted by \textit{WTO} arbitrators to include only prospective, direct trade damages, a situation that roughly re-establishes the \textit{status quo ante} the breach (more on that in our discussion of \textit{extra-contractual} remedies below). A special provision of Art. XIX \textit{GATT}, however, is the “grace period” of Art. 8.3 SGA which mandates that “the right of suspension […] shall not be exercised for the first three years that [the] safeguard measure is in effect”.

\textsuperscript{380} \textit{Art. XXI} can be seen as an exception to the obligations under the \textit{GATT}, instead of a contingency measure (e.g. Mavroidis 2007, chapter 4.1.2). In this case, the provision would not be a liability rule of flexibility backed by a zero-rule of remedy, but a property rule granted to the injurer.
financial crises and to maintain financial stability (Arts. XII, XV and XVIII.B). GATT
Art. XVIII, sections A, C, and D protect nascent industries in developing countries
from international competition and let them establish a comparative advantage in the
production of skill-intensive goods. Just like Arts. XX and XXI of the GATT, these
contingency measures largely avail injurers of remedy payments. However, the use of
these flexibility tools is crippled by a rather high level of precondition.381

5.1.1.4 The default rule

In the realm of trade in goods, Article XXVIII GATT on Modification of Schedules is the
de iure default rule of the market access entitlement. Art. XXVIII can be invoked
independently of any contractually agreed contingency. Political expediency by the
injurer, or “requesting country”, suffices to initiate tariff renegotiations at any time.382
The renegotiations requirement shows that Art. XXVIII is a property-rule-type flexibility
mechanism accompanied by a negotiated remedy of at least commensurate damages.383
The injurer is asked to engage in tariff renegotiations with the “primarily concerned
Members”, i.e. with those WTO Members holding “initial negotiation rights”, those that
qualify as “principle supplying interests”, and those that have “substantial interests”.384

The enactment of a protectionist measure is burdened by a substantial threshold, as well
as by a limited scope of application: First, Art. XXVIII allows market-closing measures
in the form of tariffs only. No other trade protection instruments are granted.385 Second,
should the renegotiation request not coincidentally fall into a triennial renegotiation
period of three months length (laid out in Art. XXVIII.1), the injuring Member must

381 See Roitinger (2004, p. 20). When a country enacts a policy measure due to a BoP-problem, for example, it
must notify all Members, must confer with the WTO BoP-committee, and remains under the periodic review
and constant monitoring of that committee (for details, cf. Mavroidis 2007, chapters 4.3 and 4.4). Whenever a
country wants to engage in infant industry protection, it has to consult with Members concerned and offer
adequate compensation. If those consultations fail, the entire WTO membership presides over the issue.

382 Although it is nowhere stated expressis verbis that GATT Art. XXVIII on tariff renegotiations is the market
access default rule, we infer that this is the case: Whenever none of the contingency measures of flexibility
apply, a country wishing to alter its previously negotiated market access balance is left with the possibility of
tariff renegotiations as the sole flexibility solution.

383 Art. XXVIII.2 mandates that the WTO Members participating in the renegotiation of the concession “…shall
endeavour to maintain a general level of reciprocal and mutually advantageous concessions not less favorable
to trade than that provided for in this Agreement prior to such negotiations.” Although this adds a nice touch to the
article, the stipulation of commensurate damages is somewhat superfluous: As was demonstrated in subsection
2.2.2 (at footnote 92) above, no reasonably rational victim country is going to settle for anything less than
commensurate damages anyway, since specific performance is its best alternative to a mutually agreed solution.

384 See Interpretative note ad Article XXVIII under § 1, as well as the detailed discussion in Mavroidis (2007,
chapter 2.3.6.3).

385 Yet by virtue of the equivalence propositions (e.g. Krugman and Obstfeld 1994, chapter 9), the erection of an
additional tariff can achieve the same market-closing outcome as any other trade measure could.
secure the prior authorization of the entire membership in order to enact its right to renegotiate (Art. XXVIII.4). Third, given the elaborate and extensive procedure entailed in tariff renegotiations, this instrument is evidently not designed for temporary, but for permanent deviations. Therefore, Art. XXVIII is not well-suited to dealing with emergency situations or transient shocks. Finally, there is the issue of selectivity: Tariff renegotiations are an untargeted, non-discriminatory measure. Tariff concessions offered as compensation have to be offered on an MFN-basis. This essentially turns the bilateral nature of market access entitlement and market access contingencies into a multilateral obligation.

Who owns the gap of the market access entitlement according to the letter of the WTO contract? We see that the market access entitlement is not protected in an absolute sense. Instead, ex post trade policy flexibility is possible. The market access entitlement is guarded by a divided entitlement protection rule (see our discussion in subsection 2.2.4.2 above): Some gaps are owned by the victim and are protected by (well-described) negative integration rules. Other gaps are owned by the injurer, who can engage in trade policy flexibility by resorting to contingency measures. If a contingency occurs that does not fall under the ambit of negative integration and contingency measures but causes regret in injurers, the Member concerned is obliged to resort to Art. XXVIII GATT. By virtue of the PR of default, most unforeseen gaps are thus owned by victim Members, who are free to sell off parts of their entitlements to the injurer. Any other behavior must be considered a violation of WTO law. Whenever the injurer engages in behavior that is not deemed illegal, but nevertheless impairs the victim’s level of market access, the resort to NVC protects that affected party.

5.1.2 De facto protection of the market access entitlement

The de iure system of entitlement protection and non-performance is not the end of the story. In the contemporary WTO contract there are various informal, de facto escape tools available to WTO Members. Although they are in contravention of the letter of the law, or at least of the spirit of the Agreement, injurers make regular use to the following trade policy flexibility tools: (1) VERs and OMAs, (2) subsidies, (3) non-discriminatory

---

Messerlin (2000, p. 162) contends that renegotiation under Art. XXVIII is a disproportionate instrument for the aim of temporary protection.
domestic policies, (4) antidumping (AD) and countervailing duty (CvD) measures, and (5) violation of the Agreement.\footnote{5}{387}

As will be shown in the next four subsections, these trade policy instruments are – to varying degrees – suitable protectionist escapes from previously made trade liberalization commitments. Injuring Members appreciate and choose these instruments for their ease of use, low enactment costs, a far-reaching scope of application, and damage requirements that are strictly lower than those of official flexibility instruments.

5.1.2.1 Voluntary export restraints (and ordinary marketing agreements)\footnote{5}{388}

A VER is a mutually agreed reduction of exports between an importing and an exporting country. It lies in the nature of the quantitative restriction instrument that the exporter thereby collects the quota rents (e.g. Smith and Venables 1991). VERs, thus, are property rules of flexibility featuring negotiated remedies.\footnote{5}{389} The use of VERs as trade barriers was common in the 1970s and 80s (Baldwin 1989; Bhagwati 1988), but was prohibited in the course of the UR negotiations (by virtue of Art. 11.2 SGA) for three reasons: Firstly, although VERs were never formally tested by a GATT panel, they probably violate Art. XI GATT (because they are quantitative restrictions) and Art. XIX (because they are targeted against single export nations). Secondly, VERs undermine the spirit of the multilateral trading system: They impose externalities on third parties. Third, the origin of VERs is notoriously difficult to prove (but see the GATT panel on Japan – Semiconductors).

The prohibition of VERs by the WTO contract, however, does not mean that the incentives to conclude such arrangements have been abolished. Actually, VERs are probably still in full practice as protectionist tools: First, the proverb “no plaintiff, no judge” applies. How can an outsider find out whether exporters reduce their output due

\footnote{5}{387} The practice of “binding overhang” is sometimes included in the list of informal contingent devices of trade policy flexibility (see Roitinger 2004, pp. 16). Indeed, some WTO Members are using the margin between bound and applied tariff rates as a temporary tariff flexibility tool. The applied tariffs are increased to shelter domestic industries from imports but not to the extent that they move above the bound rate, so that no WTO commitment is breached (see e.g. Bagwell and Staiger 2005b; Horn, et al. 2006). However, we note here that binding overhangs, although technically a tool of \textit{ex post} discretion, should not be included in the list of market access flexibility instruments, since applied tariffs do not form part of the initially agreed mutual market access balance. Consequently, a deviation from something that is not factored into the initial balance of concessions does not constitute a partial backtracking from \textit{ex ante} made trade liberalization commitments.

\footnote{5}{388} For our purposes, OMAs and VERs are substantially equivalent trade protection instruments. We will only discuss VERs in detail.

\footnote{5}{389} But note that no extra-contractual remedies (enforcement instruments) are available, since these arrangements take place in the shadow of WTO law.
to, say, bottlenecks in production or due to a secret bilateral agreement? Rosendorff (1996), for example, points out that even today many VERs are the outcomes of negotiations which originated as AD investigations or actions. Bown (2002b, p. 53) suggests that AD actions themselves offer a loophole for signatories to engage in managed trade and VER-like price undertakings. Second, VERs are even legally possible through a loophole created by the application of Art. XIX GATT in connection with the footnote to Art. 11.2 and Art. 5.2 SGA.

VERs are easy to enact, and relatively cheap for the injuring country (the price for the victim’s consent is that of forgone tariff revenues). Importantly, VERs are far-reaching in their application scope: They are free from official restraints, such as timeframes, sunset reviews, or enactment preconditions; the underlying market access contingency can be technical, political, social, or economic. Also, in contrast to GATT Art. XXVIII, VERs are a selective measure, which can be targeted at a specific exporter of concern.

### 5.1.2.2 Subsidies

Export subsidies are sometimes added to the list of informal trade flexibility mechanisms (e.g. Kleen 1989). However, the economics of subsidies as a strategic tool of trade protection are rather shaky: It is difficult to construe cases where a production subsidy is actually advantageous to the policymakers of the subsidizing country (Roitinger 2004, pp. 128). Consider the following reasons: First, not every subsidy is apt to raise the export volume of a country at the expense of production in other countries. Second, export subsidies are prohibited by virtue of Art. 3 SCM. Countries have to concoct alternative production subsidies which are apt to circumvent the purview of Art. 3. Third, export subsidies can function as trade policy flexibility tools as long as a unique game set-up is given, as used in Brander and Spencer’s famous model of strategic trade (Brander and Spencer 1985; Brander 1987, 1995): Only in the presence of a Cournot duopoly or oligopoly, and if countries have a predatory intent, can the subsidization of exporters be used to gain world market shares at the expense of foreign rivals. Under less

---

390 Rosendorff (ibid.) shows that a VER is preferred to the erection of AD duties by self-interested governments of both the exporting and the importing Members. Hence, a VER is an equilibrium outcome.

391 See Mavroidis (2007, chapter 4.7.5): Art. 5.2 SGA allows for the discriminatory application of safeguard measures, whereas the footnote to Art. 11.2 SGA permits import quotas as safeguard measures to be administered by the exporter. This is little more than having VERs enter through the backdoor.

392 Predation occurs if three cumulative conditions hold: First, thanks to subsidies foreign exporters can afford to out-price domestic competitors. Second, exporters have the motivation and means to drive competitors out of the market and to obtain monopoly power. Third, once the domestic competitors have exited, the remaining monopolist raises the prices so as to maximize its welfare.
clinical circumstances, this outcome does not hold anymore. In general, subsidies tend to lead to more competitive industries and lower world prices, so “victim” countries should actually “write a note of thank you” to the subsidizing country (Bhagwati 2002).

To conclude, subsidies are liability-type rules of flexibility. They are not accompanied by an official remedy to the victim (an indirect compensation in the form of cheaper inputs for victim countries, however, can occur). Yet as rules of trade flexibility, subsidies are inapt: They are costly to enact, and indirect and uncertain in their outcome.

5.1.2.3 Non-discriminatory domestic policies

Opportunistic beggar-thy-neighbor policies can result from the use of domestic trade-related instruments. In the absence of exhaustive contractual regulation of all instruments, domestic policy measures may be used as ex post flexibility tools (Mavroidis 2007, chapter 3.1 at pp. 270). Domestic policies, be they related to the environment, health, competition or human rights, may be defined unilaterally by each WTO Member government. The effect of such measures can be market-closing, since they may depress world prices (improve TOT), or produce political support disadvantages to policymakers abroad. Hence, non-discriminatory trade-related policies are liability-type opt-outs without any remedy payments to victims. Remedies are only payable if the victim sues for non-violation, and if the dispute panel/AB concurs.

Art. III GATT states that Members are free to choose their policies as long as they are not applied in a manner that confers an advantage on domestic over foreign production. Yet it has to be noted that opportunistic trade decision-makers can circumvent the non-discrimination stipulation of Art. III by crafting targeted policy instruments in the guise of a non-discriminatory measure. Especially if the enacting country itself does not produce any “like” or “directly substitutable” products, a measure can be crafted that pinpoints exactly a certain export good from a certain country.

---

393 For example, if countries engage in a Bertrand competition (over prices), instead of a Cournot competition (over volumes), the Brander/Spencer outcome of monopolistic competition breaks down (see Eaton and Grossman 1986; for a general critique of theories of strategic trade see Grossman 1987; Krugman 1993).

394 This informal trade policy flexibility instrument is much like GATT Art. XX without the conditionality of titles (a)-(j), but also without the opportunity to renege upon explicit negative integration provisions.

395 For example, country A, itself not a cheese producer, enacts a sanitary measure that prohibits the purchase and sale of “white, soft cheese made from buffalo milk and pickled in brine”. This policy measure would pretty much exclusively target original Italian buffalo mozzarella.
5.1.2.4 Antidumping and countervailing duties

It is no secret that AD and CvD measures are quite frequently abused as protectionist tools of flexibility. Officially, both these tools only can be invoked to counter “unfair” trade practices by foreign exporters, subject to the occurrence of material injury caused by the allegedly unfair imports. AD duties address private practice, whereas CvD action addresses government practice (in the form of subsidization). Unofficially, these two measures are highly fungible mechanisms which are easily turned into protectionist instruments.\(^{396}\) Since they are very similar tools (Mavroidis 2007, chapter 4.6; Roitinger 2004 at p. 135), we will discuss AD and CvD together.

The level of precondition for the enactment of AD and CvD actions is reportedly very low: Domestic investigating authorities have to (i) prove the existence of dumping, or the use of a (prohibited or actionable) subsidy; (ii) demonstrate material injury; and (iii) establish a causal relationship between the two. Investigating national authorities have ample leeway to initiate and investigate dumping allegations, and in so doing are largely unchecked by the WTO or any other multilateral organization. They profit from a striking lack of basic contract language, common methodology, and calculation standards for dumping margins, injury, and antidumping tariffs.\(^{397}\) Thanks to the extremely strong deferential standard of review mandated by Art. 17.6 of the ADA, national AD authorities have additional latitude in their investigation.\(^{398}\)

Coupled with a lax level of precondition, AD and CvD actions are also staffed with a comfortable scope of application for injuring countries. The instruments can be used in

---

\(^{396}\) Trade scholars have called AD and CvD actions “ordinary protection with a good public relations program” (Finger and Zlate 2003), or “a poor man’s escape clause” (Hoekman and Leidy in Rosendorff and Milner 2001, p. 830), because they are easily enacted without requiring compensation of the victim(s). The protectionist abuse of these trade remedy measures is the dominant perception in the trade literature (examples include Finger, et al. 1982; Palmeter 1991a, 1991b; Barfield 2005; Bown 2001, 2002b; Finger, et al. 2001; Messerlin 2000; Neufeld 2001; Prusa and Skeath 2002; Schuknecht 1992; Sykes 1989; Tharakan 1995; Tharakan and Waelbroeck 1994; Trebilcock and Howse 2006).

\(^{397}\) This is not the place to discuss in detail the weak deferential enactment standard of AD and CvD measures (see Mavroidis 2007, chapter 4.5; or Roitinger 2004, sections 5.2 and 5.3; Barfield 2005; Lindsey 2000; Lindsey et al. 1999; Lindsey and Ikenson 2003, 2002). Special interest groups reportedly bring to bear a significant influence on national bureaucracies in the process of AD/CvD investigations (Finger, et al. 1982). Empirical research substantiates the role played by political influence in affecting the decisions taken by domestic investigation authorities (e.g. Schuknecht 1992; Tharakan and Waelbroeck 1994).

\(^{398}\) Article 17.6 ADA comes very close to a carte blanche for domestic AD authorities. It reads in pertinent parts (emphasis added): “If the establishment of the facts [of the investigation] was proper and the evaluation was unbiased and objective, even though the panel might have reached a different conclusion, the evaluation shall not be overturned”. Schropp (2005 at footnote 37) notes in this regard: “Given that the ADA completely lacks i) a clear objective of antidumping; ii) a concrete methodology; and iii) calculation standards for dumping margins, injury and dumping tariffs, on what basis can a panel ever assess whether the domestic valuation was unbiased or objective?” (emphasis in original).
versatile fashion: AD and CvD punishments can take the form of cash deposits or bonds (Art. 17 SCM/Art. 7.2 ADA), “voluntary undertakings” by the “culprit” (Art. 18 SCM/Art. 8 ADA), or of non-MFN’ed countervailing/ antidumping duties (Art. 19 SCM/Art. 9 ADA). Next, because “unfair” trade practices may be supplier- or country specific, AD and CvD measures can be pinpointed at exporters down to the firm-level; they are imposed only on those suppliers found to be dumping or receiving subsidies. Since AD/CvD actions are applied in response to “unfair” trade, there is no requirement to offer compensation to the affected trade partner. There is no possibility of retaliation by the country found guilty of dumping or subsidizing, either. Finally, neither countervailing nor antidumping duties have a specific time-frame by which they must be terminated. Both are, however, subject to sunset reviews at maximum intervals of 5 years (Art. 21 SCM/Art. 11 ADA).

5.1.2.5 Violation of the Agreement

A violation of the WTO Agreement is a final informal trade policy flexibility mechanism that allows for partial defection from agreed-upon market access concessions. Violations constitute illegal, extra-contractual behavior.

Engaging in contract violation is a liability-rule-type opt-out, free from prerequisites, limitations of application, and without compensation requirements. If violations are discovered and condemned, however, the official enforcement procedure of the WTO sets in (more on enforcement and extra-contractual remedies below). In addition to the official punishment, convicted WTO Members may suffer reputation losses for their uncooperative behavior.

To conclude this section on the de iure and de facto protection of the market access entitlement: Although the de iure protection of the market access entitlement is rather strong in protecting the market access rights of victims, the de facto situation of trade

399 The victim Member can, however, bring a nullification and impairment claim against the injurer (Art. XXIII.1.a GATT). In fact, many GATT and WTO disputes are challenges of illegitimate AD and CvD action: As the WTO Secretariat (WTO 2007, section II.D.3.b, Table 19) finds, roughly 16 per cent (or 79 out of 505 dyadic disputes) that have been initiated between 1.1.1995 and 28.2.2006 challenge “measures taken to offset ‘unfair’ trade practices”, viz. protectionist abuses of AD and CvD actions.

400 The reputation loss, or “name-and-shame” factor, resulting from a violation of the WTO Agreement has been stressed by various authors (such as Charnovitz 2002c; Dam 1970; Chayes and Chayes 1993b; Hudec 2002; Schwartz and Sykes 2002b; Büttler and Hauser 2002; Kovenock and Thursby 1992; Guzman 2002a; Guzman 2002b; Hauser 2000). There is mistrust of the reputation-loss hypothesis for conceptual reasons (Schropp 2005, section 3.3). Also, reputation gains and losses are notoriously difficult to formalize, let alone measure. We do not consider reputation losses in connection with WTO violation any further in this study.
policy flexibility paints a different picture. Since countries can readily resort to informal instruments of *ex post* discretion, the market access entitlement is protected less rigidly than the letter of the Agreement makes believe. The contractual gaps are comfortably owned by the *injuring* WTO Members.

5.2 *De iure* and *de facto* protection of non-trade entitlements

As was shown in subsections 4.2.2 and 4.2.3, multilateral non-trade entitlements, consisting of various minimum standard- and basic auxiliary entitlements, are of a quite different nature as compared to the bilateral market access entitlement. These latter entitlements also are prone to significant incertitude but, as we shall demonstrate, their protection is more straightforward and less complicated than that of the bilateral trade entitlement. By virtue of their multilateral nature, the protection of non-trade entitlements is more absolute, i.e. there is less room for regret contingencies that would justify formal *ex post* flexibility.

5.2.1 *De iure* protection of multilateral entitlements

Multilateral entitlements are written with an eye to reaping transaction cost efficiency. They streamline the channels of international trade. This is made possible by obliging every signatory party to adhere to a commonly applicable code of rules or standards. As discussed previously (footnote 362 and accompanying text), it lies in the nature of *erga omnes partes* entitlements that unilateral escape by an injuring Member affects *all other* contracting parties negatively. Injuring behavior brings down the general level of operations in the system, and harms the competitive opportunities of *all other* signatories. Hence, chances are that this backtracking behavior is driven by opportunistic intent or is welfare-depreciating at least.

But how to assess or calculate damages incurred by victims of *ex post* backtracking from multilateral obligations? It is notoriously difficult to put a “price tag” on the defection from multilateral entitlements. The harm caused lies in having negatively affected the entire trading system, and hence is very diffuse and hardly palpable. How can the community of Members assess the disutility caused by one Member infringing upon its
obligation to notify a policy instrument, to pay its membership-fees, or to have in place a functioning patent office?401

WTO framers have addressed this conundrum: At face value, non-trade WTO entitlements are protected stricter than the reciprocal market access entitlement. Fewer official trade policy flexibility tools are available. Unless unforeseen contingencies fall under the restrictive ambit of a general or security exception (e.g. Art. 73 TRIPS, or Arts. XX/XXI GATT), multilateral non-trade entitlements are protected quite strictly. No de iure contingency measures of ex post flexibility can be detected when it comes to multilateral entitlements.

It is difficult to identify a default rule for multilateral entitlements, since nowhere in WTO Agreements is there explicit mention of how to proceed in case something previously unforeseen happens. Art. X of the WTO Agreement could be seen as some sort of a default rule: It explains the specific procedure to be followed whenever a Member tables a proposal to amend the WTO Agreement or any Agreement mentioned in Annex 1. If consensus is not reached, a qualified two-thirds majority of the Ministerial Conference is required for an amendment proposal to be submitted by the Ministerial Conference to WTO Members for their approval.402 Art. X of the WTO Agreement could thus be seen as a property-rule-type DR of flexibility.

5.2.2 De facto protection of multilateral entitlements

Most multilateral obligations are phrased as positive, not negative, obligations. They lay out what WTO Members are to do, instead of trying to specify and circumscribe which measures and instruments they are prohibited from undertaking. Provisions written in positive language are much more easily protected from illegal and opportunistic ex post

---

401 To see the logical and practical difficulties connected with the defection from a multilateral entitlement, take as an example the Norway – Trondheim Toll Ring case from 1991 (a GATT dispute): Norway failed to respect its transparency obligations under the GPA. The authorities in the city of Trondheim assigned un-notified public works to a Norwegian company. This infringement of GPA transparency provisions may have led to damage. Yet who was harmed by the measure and how could such damages be quantified? Note that in principle any company operating in the relevant field and originating from a GPA signatory country could have won the contract bid. Hence, any supplier could have successfully litigated against Norway (cf. Mavroidis 1993).

402 Amendments to five imperative provisions set forth in Art. X.2 of the WTO Charter require unanimous acceptance by the WTO membership (negative consensus). So far, only one amendment has ever passed the WTO General Council (see WTO Doc. WT/L/641 of 8 December 2005).
defection.\textsuperscript{403} Whereas the trade entitlement is plagued by many informal opt-out, there are less \textit{de facto} flexibility loopholes when it comes to non-trade entitlements.

However, one significant caveat should be noted: As was the case with the trade entitlement, multilateral non-trade entitlements can also be reneged upon through the backdoor by \textit{violating} the Agreement: Members can simply breach the Agreement, risk getting sued, and sustain the punishment that they may have to incur. We shall proceed to show that the punishment for injuring countries is generally quite low.

5.3 Rules of enforcement\textsuperscript{404}

As discussed in detail in Chapter 2 (subsection 2.1.3), contract enforcement is a fundamental aspect of contracting in general. The quality of enforcement crucially determines the willingness of contracting parties to cooperate in the first place. Signatories design tertiary rules of contracting that lay down how the exchanged entitlements, that is, of the primary and secondary rules of contracting, are to be protected from \textit{extra}-contractual, defective, behavior. The DSU is the principal legal text governing enforcement issues within the WTO contract.\textsuperscript{405} In the DSU, dispute and litigation procedures are laid down, as are enforcement provisions for the violation of substantive and auxiliary rules.

Whenever signatories are in conflict about some contested measures, they are obliged to first engage in mutual settlement negotiations (Art. 4 and 5 DSU). If these consultations prove futile, the dispute settlement body (DSB) issues a report or ruling.\textsuperscript{406} After the lapse of the “reasonable period of time” (RPT) granted by Art. 21.3 DSU, and after the compliance panel has established that possible novel measures taken by the defendant (read: the injurer) are inadequate (Art. 21.5 DSU), Article 22 DSU comes into play. It

\textsuperscript{403} For example, the obligation to grant 15 years of patent protection for pharmaceuticals leaves little room for legal loopholes: Signatories don’t care \textit{why} and by \textit{which instruments} an injuring country defected from its obligations. What matters is \textit{that} it did so.

\textsuperscript{404} This section draws in parts on Schropp (2005, section 2.1).

\textsuperscript{405} By virtue of Art. 23 DSU, the WTO subjects all Members to the exclusive jurisdiction of the dispute settlement mechanism (DSM). Therewith, the DSM is “the only game in town”, precluding both unilateral actions and the use of other fora for the resolution of a WTO-related disputes. The DSU applies to all covered agreements (by virtue of Art. II.2 of the WTO Agreement and Art. 1 DSU).

\textsuperscript{406} The DSB is composed of all WTO Member countries, and passes recommendations by dispute panels or the AB by virtue of a negative consensus rule (cf. e.g. Pauwelyn 2000, p. 336). Therewith, panel or AB recommendations become DSB rulings quasi-automatically. The two terms “recommendations” and “rulings” can thus be used interchangeably.
lays down the enforcement procedures. \textit{De facto}, the defendant now has two options.\footnote{The qualification \textit{de facto} is appropriate here because WTO scholarship is in dispute about the legal nature of panel recommendations. It is subject to debate whether or not a condemned defendant is under an international-law \textit{obligation} to comply with the panel or AB recommendation, which usually advises the defendant to withdraw the illegal measure in place (Chapter 6 below deals with this question in detail).}

It can either re-enter into negotiations with the injured State in order to negotiate “a mutually acceptable compensation”, as Art. 22.2 DSU stipulates; alternatively, the violating party can decide to stay recalcitrant and endure retaliation. The remedy of retaliation provides the complainant (victim) with “the possibility of suspending the application of concessions or other obligations under the covered agreements on a discriminatory basis \textit{vis-à-vis} the other Member, subject to authorization by the DSB of such measures” (Art. 3.7 DSU). A winning complainant can engage in pre-sanctioned unilateral tariff increases (or other barriers to trade) against some of the non-compliant Member’s export sectors.\footnote{Before an injured complainant can bring into place its unilateral tariff hikes, however, it has to notify an “authorization request” to the DSB, indicating its intended retaliation schedule (Art. 22.2 and 22.3 DSU). This schedule specifies the complaining party’s plan of retaliation: The mix of target industries, the nature of retaliation, and the division of the retaliation award on target sectors. This retaliation schedule has to be in accord with basic principles and procedures of retaliation (Art. 22.3 DSU). If challenged by the defendant country (which is most often the case), an arbitration panel (Art. 22.6 and 22.7 DSU) will review the victim’s retaliation schedule and set the quantitative amount of “suspension of concessions or other obligations”, and the mix of sectors.}

Compensation and retaliation (“countermeasures” in WTO parlance) are temporary solutions only, and are – arguably – “merely instruments to ‘restore the balance of concessions’ with compliance as the ultimate objective” (Bronckers and van den Broek 2005, p. 102). Both remedy instruments must at least match the damages suffered by the injured party.\footnote{Throughout the WTO Agreement, the fallback alternative to offering compensation is the remedy of retaliation. Retaliation is mandated to be commensurate to the harm suffered by the injured party by virtue of Arts. XIX.3(a), XXVIII.3(a),(b) GATT, or 22.4 DSU. Given the retaliation prospect as the best alternative to a breakdown of negotiations, no complainant (victim) will accept a compensation offer that is substantially less than what it expects to gain from retaliation.} Compensation offers, however, in contrast to being exposed to retaliation, are voluntary (Art. 22.2 DSU), and consequently not an automatic obligation on the part of the injuring Member.

\textit{De facto}, the size of remedies for \textit{extra-contractual} behavior is less than the \textit{reliance} measure of damages. Art. 22.4 DSU demands that the “level of the suspension of concessions or other obligations authorized by the DSB shall be equivalent to the level of the nullification or impairment”. This condition of commensurate damages has consistently been interpreted by arbitration panels and the AB as \textit{prospective direct} trade damages apt to re-establish the \textit{status quo ante} the breach. To illustrate this, consider the
four quantification issues authoritatively established by WTO panels and AB case law (Mavroidis 2007 at section 5.4.6.6.2): (i) Remedies are calculated prospectively; (ii) indirect benefits are not recouped; (iii) only added value matters; (iv) legal fees are not reimbursed.

(i) **Prospective remedies.** Remedy awards have largely been future-oriented (see e.g. Bronckers and van den Broek 2005; Grané 2001; Pauwelyn 2000; Spamann 2006). Injury is calculated *not* from the time when an illegality was committed, but from the end of the RPT. The prospective nature of WTO remedies is not put down *expressis verbis* in any WTO provision, but so far only manifested in coherent WTO jurisprudence.

(ii), (iii) **Indirect benefits and added value.** DSB arbitrators have interpreted the “level of nullification and impairment” to be tantamount to *direct trade damages*, i.e. equivalent to the effective trade losses (Charnovitz 2002c at p. 418; Hudec 2002, p. 86; Lawrence 2003, p. 37; Mavroidis 2000 at pp. 774; WTO 2004 at § 243). The arbitrators in *EC – Bananas III (Ecuador) (Article 22.6 – EC)* denied the United States indemnity for lost profits resulting from the EC banana importing regime which *inter alia* harmed the Mexican banana industry, a major consumer of U.S. fertilizer (WT/DS 27 at §§6.12-14). A direct outcome of the discussion on indirect benefits is the decision by the arbitrators in the same case to compute only *value added* when calculating the level of nullification and impairment (at §§6.18).  

---

410 In total there were five panels that departed from the standard of prospective remedies, all of them dealing with AD/CvD duties or subsidies. Panels recommended revocation and reimbursement of illegally imposed duties (Mavroidis 2000, pp. 775; Lawrence 2003, chapter 3). Disputes in subsidy and CvD matters are *not* regulated by the DSU, but by special procedures in the SCM Agreement (Arts. 4 and 7 SCM). Crucially, Art. 4.10 SCM has a different standard for remedies. It allows victims to enact “appropriate”, rather than “substantially equivalent” countermeasures. Panels and the AB have interpreted the language of Art. 4.10 SCM to bear a more extensive meaning than Art. 22.4 DSU. Therefore, in the recent WTO past *Brazil – Aircraft, Canada – Aircraft*, and *US – FSC* (WT/DS 46, 70, 108; all subsidy cases) applied retroactive damages.

411 Bronckers/van den Broek (2005, p. 103) or Grané (2001, p. 768) claim that the *prospective* nature of WTO remedies is justified by virtue of Article 19.1 DSU. Mavroidis (2000, p. 789; 2007, p. 584), however, does not detect any constraint on *retroactive* remedies as a matter of positive DSU law, and sides with the *Australia – Automotive Leather II* panel report (§§6.26).

412 See also Anderson 2002; Schropp 2005; Spamann 2006; Trachman 2006; Breuss 2004; Schwartz and Sykes 2002b; Grané 2001. *Direct trade damages* are estimated as price increase (or decrease) due to the tariff measure, multiplied by import (or export) losses, multiplied by import (or export) substitution elasticity (WTO 2005, section III.A).

413 As to the concept of *value added*, consider the following example (taken from Mavroidis 2007 at pp. 585): Assume that a good costs 10$ on the world market. 4$ of the total value is from imported goods. The exporting country of the said good will lose 10$ per unit in revenue when an illegal trade barrier has been erected against its exports. However, as a result of the trade measure it is facing, it will probably stop importing the input costing 4$. As a result, the actual nullification and impairment will not be 10$, but 6$ per unit.
The WTO system of non-performance

WTO case law has ruled out any consideration of compensation for efficiency losses and other second-order effects resulting from illegal interruption in trade. Hence, these losses always come at the expense and to the detriment of the complaining party.414

(iv) Legal fees. The arbitrators in their report on US –1916 Act (EC)(Article 22.6 –US) made it clear that legal fees paid cannot form part of the calculation of nullification and impairment (WT/DS 136 at §5.76).

In conclusion, Art. 22.4 DSU mandates commensurate damages for victims who saw their benefits under the Agreement nullified or impaired. By granting prospective remedies amounting to direct trade damages, WTO arbitrators have interpreted the term “commensurate damages” to imply the re-establishment of the status quo ante the breach. Actual WTO damage awards under DSU Article 22.6 are roughly apt to restore the trade level that would exist, had the injurer brought its contravening measure into conformity. Therefore, the current system of extra-contractual remedies is probably best characterized as a remedy short of reliance damage measure.415

5.4 Does the current system of trade policy flexibility and entitlement protection make sense?

We are now equipped to put the WTO system of non-performance into perspective and submit it to a test. How well does the current WTO framework dealing with the contractual incompleteness of the Agreement? Does the current regime of entitlement protection in the WTO make sense? There is evidently a need for improvement. In fact, the WTO trade policy flexibility regime as it stands today pretty much defies what contract theory has to say about flexibility instruments and efficient entitlement protection in incomplete contracts. Essentially, violation-cum-retaliation is the ultimate default rule for all entitlements traded in the WTO contract. Violation-cum-retaliation is

---

414 Efficiency losses are opportunity costs or losses in domestic value-added (cf. Mavroidis 2000 at p. 800; Lawrence 2003, p. 36) caused by the partial breach of the Agreement over and above direct trade effects. Those efficiency losses include the present (discounted) values of profits forgone, lost economies of scale and scope, costs of finding new markets/partners, switching-costs in production, production downsizing costs etc. Second-order effects are costs created by the application of the countermeasure of retaliation: Suspension of concessions depreciates the initially agreed-upon mutual balance of market access and leads to two-way trade on a lower, hence suboptimal, level (Charnovitz 2002c, p. 418; Schropp 2005 at footnote 10).

415 We say “roughly” and “short of” the reliance damage measure, because the prospective nature of the arbitration award renders impossible the exact re-establishment of the status quo of the market access balance as it existed before the breach. The remedy applied by WTO arbitrators fails to compensate the victim for direct trade damages incurred between the enactment of the measure in question and the lapse of the RPT.
a *de facto* liability rule accompanied by remedies which are systemically under-compensatory. Since contract-conforming and *extra*-contractual behavior are sanctioned in the same way, enforcement of violations is deficient as well. Thus, the WTO misses the benchmark of the EBC by far, as we will demonstrate below.\(^{416}\) We proceed by discussing the flaws in the protection of the market access entitlement first (5.4.1), and will continue with the assessment of non-trade entitlements (5.4.2). Subsection 5.4.3 discusses the dynamic effects of the defective system of non-performance for the international trading community.

### 5.4.1 The flawed protection of the market access entitlement

What is wrong with the WTO’s protection of the reciprocal market access entitlement? We find that both the *de iure* and the *de facto* entitlement protection regime show remarkable weaknesses. Coupled with ineffective enforcement of *extra*-contractual behavior, the system of non-performance protecting the market access entitlement is in trouble and could benefit from a healthy reform.

#### 5.4.1.1 Examining flaws in the *de iure* system of entitlement protection

Take note of six concerns in connection with the *de iure* system of flexibility and entitlement protection of the market access entitlement:

1. **Article XXVIII GATT is a questionable rule of default.** The renegotiation provision of Art. XXVIII GATT is a weak and inept default rule of flexibility. First, as was mentioned previously (footnote 386 and main body text), the scope of Art. XXVIII is limited to permanent, not temporary, tariff modifications. The procedural build-up connected to tariff renegotiations renders emergency and temporary protection impossible.\(^{417}\) Second, the renegotiation provision of Art. XXVIII GATT is foiled by a

---

\(^{416}\) Many of the current problems of trade policy flexibility have been addressed by WTO scholarship as isolated issues; some have been subject to litigation in high-profile WTO disputes. In this section, we try to give a *comprehensive* and systemic critique of trade policy flexibility and enforcement in the WTO.

\(^{417}\) If the renegotiation request does not happen to fall into an official triennial renegotiation period of three months (laid out in Art. XXVIII.1), the injuring Member has to secure the prior authorization of *all* Members in order to enact its right to renegotiate (Art. XXVIII.4). As per *Interpretative Note ad Art. XXVIII GATT* (§§4.1, 4.4), the injurer must submit a written request to the Council of Trade in Goods, the relevant organ to decide. The requesting Member must supply comprehensive statistical and other information justifying its appeal and listing the effects of the envisaged measure. The Council will then give notice of its *consensus* decision within 30 days. *Interpretative Note ad Art. XXVIII GATT* at §4.5 states that later on in the process the same Council determines – i.e. *all* WTO Members decide unanimously – whether the compensation offered by the injurer is sufficient. This sort of conditionality is certainly apt to slow down the process of reacting promptly to unforeseen contingencies and unanticipated shocks.
The WTO system of non-performance

weak remedy accompanying the entitlement protection rule. Contract theory tells us that a property rule of renegotiation must be complemented by remedies in excess of the expectation damage rule in order to be operable. Anything less reduces a property rule to a liability rule, because the injurer will practically always choose to pay the remedy instead of engaging in renegotiations. This is exactly the case with Art. XXVIII: Although at first sight it looks like a property rule of flexibility backed by negotiated compensation, the tariff renegotiation clause is effectively a liability-rule-type opt-out accompanied by WTO arbitrators’ interpretation of commensurate damages. Paragraph 3 of Art. XXVIII states that an injuring Member may proceed to unilaterally withdraw concessions in cases where negotiations over MFN compensation break down. Adversely affected trading partners may then bilaterally retaliate by withdrawing substantially equivalent concessions or other obligations through recourse to binding arbitration (as per Art. 22.6 DSU).

2. What is the substantial difference between GATT Arts. XIX and XXVIII? The finding that tariff modification is actually a liability-type escape provision then begs the question of whether there is a substantial difference between GATT Arts. XIX and XXVIII. It is somewhat elusive as to why the trade entitlement is protected by two flexibility mechanisms, both of which feature a liability rule and the same remedy (“substantially equivalent” damages, cf. footnote 409 above). Both articles allow injurers to opt out at their discretion for the price of offering compensation or enduring suspension of concessions or other obligations. Only the set of enactment preconditions, as well as the scope of application differ. These differences notwithstanding, it is incomprehensible from a contract-theoretical point of view why a contingency measure and a default rule should be so similar in nature and design.

3. Insufficient scope of de iure escape mechanisms. Another concern is the insufficient scope of existent de iure escape clauses. The WTO is a political contract. For this reason market access externalities are only indirectly economic and largely political in nature. It

418 As was demonstrated above (see text accompanying footnotes 92 and 225 in Chapter 3), negotiated remedies usually safeguard this outcome, since the victim country’s best alternative to a mutually agreed solution is to insist on contract performance.

419 Pertinent parts of Art. XXVIII.3 read (emphasis added): “If agreement between the Members primarily concerned cannot be reached […], the Member which proposes to modify or withdraw the concession shall, nevertheless, be free to do so and if such action is taken, any Member with which such concession was initially negotiated, any Member […having] a principal supplying interest and any Member […having] a substantial interest shall then be free not later than six months after such action is taken, to withdraw […] substantially equivalent concessions initially negotiated with the applicant contracting party.” This paragraph does no more
is somewhat odd that the drafters of the WTO/GATT have not provided for any political escape mechanism, and in addition have given such a strong conditionality to Art. XIX GATT (Roitinger 2004).

The EC – Hormones cases demonstrated that the European Communities, for political or health reasons, wished to withdraw from a previously made market access commitment. This endeavor was not backed by any formal WTO escape clause: Given GATT Art. XIX’s narrow application scope of reacting to economic shocks, and heeding the apparent inadequacy of Art. XXVIII renegotiations as a safety valve, the EC felt obliged to keep on violating the GATT Agreement. Lacking any formal means to spontaneously and temporarily withdraw from existing concessions is a serious issue, because it effectively blurs the line between good-faith and bad-faith (opportunistic) behavior.

4. The application scope of non-violation complaints is limited. WTO signatories are heedful of domestic policy measures that may display external trade effects. A victims’ right to balanced market access is protected from inadvertent backtracking by way of the non-violation complaint in GATT Art. XXIII.1.b. Technically, NVC can be applied whenever a victim government feels that its previously made trade liberalization concessions have been nullified and impaired by a legal (or rather: not illegal) policy measure of another signatory. De facto, however, WTO dispute panels have been extremely reluctant to admit NVC in situations other than regulatory subsidies. WTO

---

420 Sykes (1991 at p. 289) concurs that high requirements undercut the political utility of the escape clause.

421 “Measures affecting livestock and meat (Hormones)” (DS 26 and 48), and “Continued suspension of obligations in the EC – Hormones dispute” (DS 320 and 321).

422 This is, of course, only one interpretation. Other observers may come to different conclusions, arguing that the EC acted genuinely malevolently, or was “putting to a test” the infant DS system of the WTO.

423 This reading of Art. XXIII.1.b is confirmed in EC – Asbestos (at §§ 188). The case law in EC – Asbestos “did not limit the realm of possible applications of a non-violation complaint. […]NVC are ] based on the open-ended language employed in Art. XXIII.1.b GATT. Following this jurisprudence, it seems plausible to argue that non-violation complaints can be raised against practically each and every government measure that might have an impact on the value of negotiated concessions” (Mavroidis 2007, p. 571).

424 Bagwell and Staiger (2002b), and Bagwell (2007) take the view that non-violation complaints are an attractive yet starkly underutilized means of ensuring that reciprocal commitments will be observed.
practice has thereby burdened victims with a relatively high enactment threshold for invoking an NVC.425

5. **WTO framers picked the wrong battlefield: Negative integration clauses are over-engineered whereas default rules are neglected.** By putting too much effort into the drafting of negative integration norms, the framers of the WTO not only created the questionable impression of contractual completeness, but also neglected the much more salient issue of drafting solid default rules. The WTO founding fathers shoved the square peg of explicit treaty language into the round hole of inevitable contractual incompleteness. Consider the following arguments:

First, concentrating on negative integration rules is a problematic endeavor. The trade liberalization negotiations of the Uruguay Round witnessed a boom in negative integration provisions protecting the market access entitlement (e.g. in TBT, SPS, GPA, ROO, etc.). Some authors interpreted this impulse as a general drive towards a rule-of-law over the last decade.426 To us, however, this evolution does not mark a paradigmatic shift in the world trading community, let alone in the contractual logic (as sources claim). Rather, we interpret this evolution as an attempt to ameliorate the general contracting environment. WTO signatories tried to mend legal loopholes that had become apparent through the GATT years by adding explicit (and quite often complicated) language that anticipates possible contingencies and prohibits certain types of policy behavior.427

We are critical of these efforts in the contracting strategy of diligence: The incertitude connected with the market access entitlement in the WTO is insurmountable (subsection 4.3.1.2). Trying to overcome incompleteness by writing down ever more explicit

425 Standing GATT/WTO case law (e.g. in Kodak – Fuji, Spain – Un-roasted Coffee, or EC – Asbestos) has ruled that for a WTO Member to successfully launch a non-violation complaint, the victim must demonstrate that four conditions are cumulatively met: (i) As a result of some action taken by a WTO Member, the consistency of which the WTO Agreement is not in dispute; (ii) provided that such action occurred subsequently to the conclusion of a tariff concession, and (iii) provided that such action could not have been reasonably anticipated by the complaining, (iv) the action at hand impaired benefits accruing to the victim, i.e. reduced the value of the concession negotiated between the complainant and the defendant (Mavroidis 2007, p. 571; Petersmann 1991).

426 Charnovitz (2001; 2002c; 2002a), among others, alleges a gradual move from a diplomatic bargaining forum under the GATT (1947-94) towards an international rule-of-law system with proper norms and common values in the WTO. The drive towards a stronger rule-of-law has also been termed “rule-orientation” (as opposed to “power-orientation”, Bagwell and Staiger 2002b, pp. 5 and 36; Jackson 1997a at pp. 109; Roitinger 2004, p. 143).

427 This drive towards completeness seems to be an organic development in relational contracts. As Cohen (1999, p. 82) puts it: “[T]he actual historical development of contracts is probably best described as starting as incomplete as possible, then becoming more complete and formal as governance mechanisms other than the written contract proved to be inadequate.” It seems debatable, however, whether this evolution is smart.
The WTO system of non-performance obligations must necessarily remain a patchwork endeavor. Next to creating additional room for ambivalence and ambiguity (cf. text accompanying footnote 372), more contract language brings with it the illusion of completeness. Parties think they have “nailed down” all eventualities. On the one hand, this creates unnecessary rigidity, deterring injurers to react to regret contingencies. On the other hand, only when they are actually involved in lengthy and costly disputes do signatories realize that they did not nail down all eventualities at all.

The second aspect of the argument is the striking imbalance between “complete” and “incomplete” contracting in the WTO: As shown, market access contingencies are of such a nature that they render a completion of the contract impossible. This calls for the design of flexibility rules, more precisely, of an efficient rule of default. Chapter 3.2.5.2 above demonstrated that DR are important provisions in incomplete contracts, because they enter into effect whenever a previously undescribed contingency occurs. In order to fit many different contexts, DR should be easy to enact, have the broadest possible application scope, and be accompanied by the appropriate rule of remedy. Art. XXVIII GATT as the DR for the market access entitlement, however, is evidently at fault. Its enactment is slack and cumbersome; it is protected by an inadequate intra-contractual remedy rule; its application scope is porous. Instead of shaping a more efficient rule of default, WTO framers put the emphasis on crafting ever more elaborate negative integration rules.

6. Retaliation is a questionable mechanism of remediation. Most de iure flexibility mechanisms couple the injurer’s ex post escape with a compulsory remedy, which is the offer of tariff compensation. If compensation negotiations break down, the victim is authorized to engage in suspension of concessions or other obligations. As we will lay out in our discussion of DSU enforcement infra, the countermeasure of retaliation as a mechanism of remediation is questionable and certainly inferior to compensation offers.

To sum up our qualms with the de iure system of market access entitlement protection: Any incomplete, but efficient contract must protect its substantive entitlements against ex post opportunism, but nevertheless allow for Pareto-superior post-contractual flexibility. On the one hand, the WTO features entitlement protection instruments that

428 The application scope of GATT Art. XXVIII is quite obviously porous. It is deficient from the point of view of injurers, as we showed above. It is also insufficient from a victim’s point of view: If it were an efficient rule of entitlement protection, victims would not need to resort to non-violation claims. But since the available
are imperfect and porous: Negative integration norms neither cover all possible market access contingencies, nor prohibit all sorts of opportunistic behavior – as the presence of NVC attests. The provision of NVC, in and of itself a valuable tool of entitlement protection, is crippled by WTO practice. On the other hand, the de iure configuration of trade policy flexibility is also less than satisfactory: A country that wants to opt out for non-opportunistic reasons cannot easily do so in the face of rigid flexibility mechanisms that are difficult to enact and narrow in their application scope. The default rule that eventually protects the market access entitlement is thwarted against potential injurers.

5.4.1.2  Examining flaws in the de facto system of entitlement protection

So much for the systemic flaws of the de iure system of non-performance. Things get more irritating, once the de facto situation of trade policy flexibility enters our assessment: The de facto system of flexibility is apt to annihilate many a market access provision written down in the WTO contract. For injurers, informal flexibility tools (especially domestic policy measure, AD and CvD action, and violation of the Agreement) supersede de iure trade policy flexibility tools, simply because they are (1) easier to enact, (2) possess a broader scope of application, (3) are cheaper in terms of political currency, and (4) are inexpensive in terms of remedies payable to the victim.

1. Informal flexibility instruments have lower enactment thresholds. As was indicated in subsection 5.1.2, informal instruments of ex post flexibility have lower enactment thresholds precisely because they are enacted in the shadow of the law: AD and CvD, for example, were written to counter unfair trading, not to be abused as tools of protectionism. Violation of the Agreement and trade-related domestic policy measures can be enacted by injurers anytime and without any precondition.

2. Informal flexibility mechanisms possess a broader scope of application. Informal escape tools are either illegal or at least counter to the spirit of the Agreement. Hence, they are less impeded by contractual language geared towards restricting their application scope. Market-closing domestic policy instruments, for example, can be enacted for all sorts of reasons, not just because of unforeseen economic shocks (as prescribed by Art. XIX GATT). An important factor is the issue of selectivity: Many informal escape mechanisms, most notably VERs, AD, CvD and violation of the Agreement, can be pinpointed against those exporting countries, industries or firms that

contractual rule of default patently fails to kick in whenever an unforeseen contingency occurs, the design of non-violation claims seemed apposite to the framers of the WTO.
are of special concern to an importing WTO Member. Instead of having to treat all countries in the same way, injurers can better address their protectionist needs.

3. Informal trade policy flexibility tools are politically more convenient to policymakers. Numerous authors have examined the political economy of flexibility mechanisms. They find various reasons why informal escape mechanisms (in particular AD, CvD, and violation-cum-retaliation) are politically more opportune to protectionist policymakers than formal flexibility tools. There is a rather intuitive explanation for the reluctance of countries to utilize de iure escapes: First, policymakers are averse to overtly legitimizing a protectionist measure applied against fairly traded goods – this would be tantamount to an official admission of guilt and incompetence. Informal opt-outs such as AD/CvD action or VERs circumvent problems of this kind. Second, the direct nature of the de iure measures does not allow for “blame-shifting” or “scapegoating” of allegedly unfair foreign trade practices. Third, formal flexibility tools display a lack of selectivity. Finally, the requirement that compensation be offered immediately – and not after a lengthy litigation – undermines the engagement in formal escape.

4. De facto flexibility tools are accompanied by lower remedies. An additional and indeed compelling feature of informal escape mechanisms for injurers is that damage remedies payable to the victim pursuant to an unofficial escape are unequivocally lower than those looming in the aftermath of a formal act of non-performance. Depending on which de facto flexibility tool is chosen, remedies range from zero to damages that are strictly less than those “substantially equivalent to the level of nullification or impairment” – which WTO jurisdiction has interpreted to be roughly equivalent to the reliance damages. This is so because damages are payable only if the victim sues and the injurer is subsequently found guilty of the violation. Due to the probability that the victim may not go to court at all, to the possibility that the injurer may actually win the case, and to the probability that the victim country may refrain from enacting its awarded retaliation rights, the de facto remedy expected by the injurer is strictly smaller than the

---

429 See Schropp (2005, section 3) for details and literature references.
reliance damages payable under the formal opt-out mechanisms.\textsuperscript{432} We omit considerations of potential reputation losses on part of the injurer (see footnote 400 above).

\textbf{5.4.1.3 Examining flaws in the system of enforcement}

This discussion of remedies regarding informal, and largely extra-contractual, flexibility measures immediately leads to an examination of problems connected with the WTO system of enforcement (the tertiary rules of contracting). The contract’s extra-contractual remedies, which signatories have shaped in order to safeguard the abidance to intra-contractual rules of flexibility, have three serious drawbacks:

1. \textit{Intra- and extra-contractual behavior get sanctioned in the same way}. Throughout the GATT and the DSU, the same remedy of withdrawal of “substantially equivalent concession” appears as a countermeasure to legal and illegal nullification or impairment of previously agreed market access concessions (see Art. 22.4 DSU, GATT Arts. XVIII.7.b, XIX.3.a, or XXVIII.3.a/4.d). It is hard to think of any reason why the framers of the GATT/WTO could have felt that intra- and extra-contractual behavior should be sanctioned identically. On the contrary, contract theory postulates that illegal \textit{ex post} behavior should be punished by high extra-contractual remedies. Yet by sanctioning extra-contractual defection as lightly as contract-conform default, the WTO reduces the distinction between illegal and legal behavior to a legalistic formality.

The bottom line is that deliberately violating the Agreement \textit{de facto} is indeed penalized less than if a Member resorts to the \textit{de jure}, intra-contractual flexibility mechanisms (see point 4 above). Thus, the WTO contract effectively establishes violation-cum-retaliation as the \textit{de facto} default rule. The lax punishment of Art. 22.4 DSU turns violation-cum-retaliation into the fallback option for any injurer, since this fallback strategy is at the same time one of the most attractive escape tools (Mavroidis 2000). This is consequential: For the potential injurer, the fallback option sets the benchmark for all other (formal and informal) escape remedies the WTO knows today. Moreover, it determines the power relationship in all settlement negotiations between injurers and

\textsuperscript{431} The non-discriminatory feature of formal safeguard measures has been cited as one reason why it is infrequently used: Governments may prefer a more targeted instrument which can be directed at the country or set of countries that are the source of the injury (e.g. Barton, et al. 2006).

\textsuperscript{432} An exception here is the instrument of VER, when, due to a mutual side-agreement, the victim does not file a complaint with the DSB. Instead, the remedy is bilaterally negotiated. However, given that the injuring
victims: When bartering over voluntary compensation, no injurer will be willing to settle above its reservation utility, i.e. the expected cost of enduring retaliation.

2. **Extra-contractual remedies are systematically under-compensatory.** Article 22.4 DSU mandates that extra-contractual remedies are to be “substantially equivalent” to the damage done. As just demonstrated, the WTO’s interpretation of commensurate damages is effectively standard-setting for all flexibility mechanisms of the WTO. However, arbitration panels have shown a somewhat inadequate understanding of commensurate damages. So far, WTO arbitrators have awarded retaliation roughly equivalent to reliance remedies – at best. The currently applied countermeasure of retaliation amounting to *direct (prospective) trade damages* is insufficient. These trade damages may roughly re-establish the *status quo ante* the breach, but do not even satisfy the benchmark of *expectation damages* needed to meet the proportionality principle usually demanded by *intra-contractual remedies*.

Many authors have criticized the way in which equivalent damages have been calculated pursuant to Article 22.4 DSU. They have argued that official retaliation awards failed the benchmark of rebalancing the mutual market access balance (Anderson 2002; Breuss 2004; Mavroidis 2000; Spamann 2006; Trachtman 2006; as well as various contributions in Horn and Mavroidis 2004; 2005; for an in-depth discussion see Keck 2004). Two main criticisms have been brought forth:

First, a systematic under-compensation is said to result from the absence of *retroactive* damage awards (cf. Bronckers and van den Broek 2005; Pauwelyn 2000). Prospective damages fail to compensate the victim for losses incurred between the enactment of the violating measure and the end of the RPT.433

A second point of concern is dispute panels’ interpretation of “substantially equivalent damages” as the *reliance* measure of remedy. Apart from the fact that WTO panels have not been fully transparent in the methodology they applied for calculating retaliation

---

433 In addition, prospective damages coupled with weak procedural rules invite opportunistic “foot-dragging” tactics by offending Members. The most powerful procrastination strategy is to swap one non-compliant measure with another one (Lawrence 2003; Schropp 2005). Also, countries can enact any illegal measure for the time period until a report is adopted which determines its illegality (Bronckers and van den Broek 2005; Pauwelyn 2000; Trachtman 2006). The United States in *US – Steel* case (WT/DS 248, 249, 251-254, 258, 259) tellingly revealed that this indeed is a workable strategy (Hufbauer and Goodrich 2003a, 2003b).
awards, the reliance damage measure misses the minimum target of the expectation damage measure. Reliance remedies simply do not make up for the actual welfare loss which the injurer caused by the offending measure, because these damages are strictly less than the replacement value which includes efficiency losses that the victim requires to be properly compensated.

3. Retaliation is a suboptimal countermeasure. Although bilaterally negotiated compensation is the preferred countermeasure in the WTO (by virtue of Art. 3.7 DSU), “compensation is a rare event” (Pauwelyn 2000 at p. 337). Retaliation is the more frequently applied remedy. The basic idea of retaliation is that sectors not directly involved in the dispute are punished, and consequently exert pressure on the non-compliant government to bring its measure into conformity. Besides the deterrent effect on the offending party, protection of its own import-competing sectors constitutes at least a partial indemnification for the complaining party (the victim). The significant advantage of retaliation over compensation is its self-enforcement quality;

---

434 See WTO (2007, section II.D.3.c.(iii)). For example, in order to estimate the trade effect of a disputed measure in EC – Bananas (22.6) (WT/DS 27), the value of EC imports under the disputed banana import regime needed to be compared to an estimated value under a WTO-consistent regime. Arbitrators did not report how they determined the counterfactual scenarios, and why the final award differed from some of the counterfactuals proposed by the litigants. But see the modeling approach in US – Continued Dumping and Subsidy Offset Act of 2000 (22.6) (WT/DS 217) for an illustration of a more systematic, stringent, and theory-based damage calculation approach (WTO 2005, section III.A).

435 As laid out in Chapter 3.3.2.1, in a situation of bilateral ex ante commitment and in the presence of unforeseen contingencies, the expectation damage measure Pareto-dominates all other remedies accompanying a liability rule of remedy. Expectation damages can only effectuate a “breach” decision which mimics non-performance of the hypothetical complete contingent contract; they induce the injurer to engage in Pareto-efficient ex post non-performance. The reliance damage, on the other hand, is systematically under-compensatory. It induces injurers to utilize the opportunity to default from contractual obligations more often than is efficient. Thus, in certain situations where the complete contingent contract would mandate performance, the reliance measure incentivizes the injurer to default nevertheless.

436 Expectation damages comprise of direct trade damages and all efficiency costs from the moment the measure in question was enacted. As to the nature of efficiency losses in the trade context see footnote 414 above.

437 The Sutherland Report (WTO 2004 at §243) notes: “Valuation [i.e. the monetized calculative basis for trade damages] would have to consider not only effective losses, but also potential gains that are nullified and impaired.”

438 There is only one official dispute in which compensation was agreed. Following arbitration in US – Section 110(5) of the US Copyright Act (WT/DS 160), the U.S. paid financial damages to the European music industry until the offending law was repealed. Roitinger (2004, pp. 37) shows that compensation offers under the safeguards clause of Art. XIX GATT have also been quite rare: Between 1948 and 2003 compensation has been offered by temporarily escaping countries in only 16 safeguard cases. Moreover, ever since 1978 there has not been a single notified compensation offer under XIX.3 GATT. Schropp (2005) assesses why compensation is such an unattractive tool for injuring countries: Among other factors (see footnote 429 and main text), it seems that tariff compensation is procedurally disadvantaged compared to the instrument of retaliation, because it is not supported by an official DSU mechanism.

439 Retaliation leads affected exporters to lobby with their government to keep foreign markets open and to act as a counterweight to the influence of import-competing industries.
retaliation can be executed by the complainant without prior consent by, or accommodation of, the defendant.

Yet retaliation as a mechanism of remediation poses grave disadvantages for WTO signatories: 441

First, retaliation is economic nonsense. Higher levels of protection introduce additional economic inefficiencies on both sides. This brings down the general level of the previously negotiated (presumably politically optimal) market access commitments. 442

Second, some authors contend that the instrument of suspension of concessions violates human rights by barring uninvolved individuals from economic activity (Bronckers and van den Broek 2005; Charnovitz 2001; Petersmann 2003, 2002). Retaliation harms “innocent bystanders”, such as consumers and competitive industries. Third, retaliation is likely to lead to trade diversion, and thus has economic spillovers on uninvolved third countries. Fourth, the complaining Member itself may have no interest in implementing retaliatory measures, when the costs of raising tariffs on much needed imports are considered too high, both economically and politically. Even large WTO Members may face strong resistance by domestic consumers and importers of intermediate products, who suffer from higher prices or disturbed relations with regular suppliers (Anderson 2002). Fifth, many WTO scholars allege that the instrument of tariff retaliation is severely biased against small countries, and hence inherently unfair (Bronckers and van den Broek 2005; Charnovitz 2001, 2002c; Diego-Fernandez 2004; Hudec 2002; Pauwelyn 2000; Anderson 2002). Small countries, in view of their limited market size, may well be unable to exert sufficient pressure on larger Members to alter their behavior (e.g. Hudec 2002). Retaliation threats of small countries thus fail to deter economically powerful Members from committing a violation (Mavroidis 2000; Pauwelyn 2000). Large countries may either remain non-compliant, or offer the complaining small

440 This is so owing to either political economy considerations or to a positive terms-of-trade effect, if the retaliating country is large enough to affect world prices.

441 Cf. also Schropp (2005, section 2.2), and WTO (2007, section II.D.3.c).

442 Many scholars have noted that retaliation restricts (rather than promotes) economically beneficial trade. It harms consumers and competitive industries worldwide, while protecting declining sectors that vie for protectionism (Bronckers and van den Broek 2005 at pp. 107; Charnovitz 2001 at pp. 811; 2002c at p. 419; Mavroidis 2000, pp. 800; Pauwelyn 2000; Brainard and Verdier 1994).
country settlements at unfavorable conditions.\textsuperscript{443} The futility of retaliation gives rise to the suspicion of “missing cases” due to a significant “chilling effect” (Bown and Hoekman 2005): Small-country complainants may not bother to bring up disputes, because the costs of dispute settlement are incurred without any hope of obtaining reparation (Bronckers and van den Broek 2005).\textsuperscript{444} Sixth, and summarizing all previous drawbacks, the countermeasure of retaliation completely frustrates the spirit and purpose of the WTO as a whole. Raising protectionist barriers in response to market access defections runs counter to the liberalizing spirit of the WTO, and its objective to secure predictable business opportunities.\textsuperscript{445}

To conclude our discussion of \textit{de iure} and \textit{de facto} flexibility mechanisms and the accompanying extra-contractual remedies with regard to the market access entitlement: Intra- and extra-contractual behavior are sanctioned in the same way. This opens the floodgates to unchecked use of informal trade policy flexibility instruments in the WTO. As a consequence, the contemporary trade flexibility regime practically obliterates any \textit{de iure} rule. To a considerable extent, it defies what contract theory has to say about efficient entitlement protection.

While the \textit{de iure} system of trade protection privileges victims and exacerbates \textit{ex post} discretion of potential injurers, the \textit{de facto} system of non-performance seriously disadvantages victims of \textit{ex post} discretion: No matter what the various WTO Agreements may say, with regards to the market access entitlement the current flexibility regime is tantamount to a pure liability rule with varying \textit{substandard} remedies. This sets strong incentives for injurers to simply disregard the rules of the game.

It is not too difficult to see that in a given situation a reasonably rational injurer will always go for the very escape instrument which promises “most mileage”, i.e. the fewest

\textsuperscript{443} Latin American trade diplomats are quoted with the remark that “trade sanctions are a huge club in the hands of industrial giants and a splinter in the hands of developing countries” (quote from Charnovitz 2001 at note 211; cf. also Palmeter 2000, p.472). Indeed, developing countries have never suspended concessions so far. In at least nine pertinent instances small countries won trade disputes, but no action followed – neither compliance by large Members, nor a mutually agreed solution, nor retaliation. These cases are: DS 27, 122, 217, 241, 267, whereby co-complaints are counted as separate instances (see Horn and Mavroidis 2006a, 2006b).

\textsuperscript{444} Bown (2004) shows that the retaliatory capacity of complainants is the crucial determinant affecting injuring governments’ policy decision to comply with panel rulings, or to stay recalcitrant in the face of the adopted panel/AB report. Retaliation capacity is thereby understood as the complainant’s market power vis-à-vis the defendant: The more dependant the defendant’s exports are on the market of the complaining party, the more credible and deterring are the latter party’s self-enforcement threats.

\textsuperscript{445} On that account, see the Sutherland Report (WTO 2004 at §240), or Hudec (2002, p. 88).
enactment costs, the lowest compensation, and the largest scope of application. With the informal flexibility tools VERs/OMAs, subsidies, domestic trade-related policies, AD or CvD actions at its discretion, the injuring country can renegade on the trade entitlement practically for free. Alternatively, the injurer can simply infringe upon the contract at the price of losing a trade dispute and having to pay prospective reliance damages, which – in expected terms – are strictly less than the harm done to the victim(s). Violation-cum-retaliation is the ultimate de facto default rule for the market access entitlement, whereby the court-ordered retaliation sets the standard for all intra- and extra-contractual remedies.

5.4.2 The flawed protection of multilateral non-trade entitlements

Turning to the protection of multilateral non-trade obligations traded in the WTO (minimum standard and basic auxiliary entitlements), our biggest criticism of the de iure situation concerns the absence of a clear and unambiguous default rule: Quid in casu a situation occurs that is not anticipated in the letter of the contract? The WTO Agreement is largely mute on that issue. Earlier, in subsection 5.2.1, we mentioned WTO Charter Article X on treaty amendments. However, as Mavroidis (2007, p. 547) notes, the WTO – in contrast to the Vienna Convention on the Law of Treaties – does not distinguish between amendments and modifications. Amendments are once-and-for-all changes of the treaty language, whereas modifications bear a more temporary, intermittent and discretionary connotation.

In the absence of any explicit treaty language concerning a DR of flexibility, we are forced to conjecture as to how the WTO framers intended to deal with unforeseen non-trade contingencies. We can see three possibilities on the nature of default rules of multilateral WTO entitlements:

- Alternative 1 is that Members chose to make ex post changes in multilateral obligations difficult, yet not impossible. Therefore, Art. X of the WTO Agreement (a property rule of renegotiation) is actually the only way for signatories to react to unforeseen contingencies.447

---

446 For example: Between 1994 and 2007 there were 20 times more antidumping initiations (2,851) than there were safeguard initiations (142), and nearly 26 times more (1,804 compared to 70) AD than safeguard measures applied (source: WTO 2007, section II.C.4).

447 There are some problems connected with this interpretation. For example: Does Art. X of the Charter protect all multilateral entitlements in the same fashion? This would mean that peremptory norms of international law are
Alternative 2 would assume that, when drafting the Agreement, WTO Members desired that there should be no possibility to temporarily escape from previously agreed multilateral concessions. To that effect, signatories deliberately omitted any language on emergency actions or temporary modifications. Post-contractual non-performance would automatically be interdicted by virtue of a general rule of inalienability. This, however, would have been extremely poor drafting on the part of the founding fathers: The WTO as an international treaty is not concluded in a vacuum, but is part of the grander structure of public international law. That means that wherever the WTO contract is mute on a certain issue, external norms of international public law automatically take effect.448

Alternative 3 is that WTO framers conscientiously did not “contract around” customary rules of international law, so that the VCLT takes effect as the adequate rule of default rule of multilateral entitlements whenever an unforeseen regret contingency occurs (Pauwelyn 2006).

All three alternatives are unsatisfactory, because they lump together all multilateral entitlements, and do not give consideration to the idiosyncrasies of each entitlement type: Is it sensible to protect procedural rules or transparency provisions in the same way as a ius cogens norm or minimum standard entitlements? There is no logical or economic reason for that.

Considerations of de jure default rules notwithstanding, the de facto DR of multilateral entitlements is violation-cum-retaliation, just as is the case for the market access entitlement. Violating the Agreement and then enduring retaliation is a liability rule granted to the injurer, backed by less-than-reliance damages by virtue of Art. 22.4 DSU. Consequently, WTO Members have an incentive to over-“breach” WTO Agreements: Injured parties may or may not challenge that measure, and may or may not succeed. Arbitrators will face the immensely difficult task of putting a price tag on the challenged measure (see footnote 401 above and accompanying text), and the successful victim may

---

448 Dunoff and Trachtman (1999, p. 35) opine that the Vienna Convention on the Law of Treaties (VCLT) and the ILC Draft on State Responsibility may be seen as collections of external DR of international contracting conduct between sovereign States (see also Pauwelyn 2003; Mavroidis 2000; Grané 2001). On the relevance of public international law in the WTO legal order, see the comprehensive studies by Pauwelyn (2003), and Trachtman (1999).
or may not actually engage in retaliation. This *de facto* result should, by any reckoning, be the exact opposite of what framers must have had in mind when they were contemplating the nature of entitlement protection of the various multilateral entitlements.

### 5.4.3 Conclusion and consequences

To conclude this section on the assessment of the current WTO system of entitlement protection, trade policy flexibility and enforcement, let us offer a short summary of our findings and discuss the dynamic implications that a flawed system of non-performance in international trade entails.

#### 5.4.3.1 Summary of the main flaws of the current WTO system of non-performance

It is well-established that contractual escape mechanisms are an indispensable feature of incomplete trade contracts. Yet the contemporary system of trade policy flexibility and entitlement protection in the WTO is remarkably flawed.

The efficient “breach” contract – the first-best achievable contract – mandates equivalence between rules of *ex post* flexibility and entitlement protection: A system of non-performance must prohibit opportunistic *ex post* market-closing behavior, yet reap the opportunities looming by unforeseen regret contingencies. It must also compensate victims of post-contractual discretion so that their *ex ante* commitments do not get frustrated by *ex post* escape. Finally, the system must be able to protect signatories from illegal *extra*-contractual behavior.

The current WTO system of non-performance misses that mark. It is plagued by striking imbalances: Its governance structure fails to balance between the market access entitlement with non-trade entitlements, its approaches to incomplete contracting and complete contracting, the *de iure* and *de facto* organization of trade policy flexibility, and the way injurers and victims are treated. Our critical analysis of the current WTO system of non-performance revealed a series of grave shortcomings:

---

449 The imbalance between trade- and non-trade entitlements is striking: As shown above, entitlement protection and flexibility rules abound when it comes to the market access entitlement. Eliciting the protection of multilateral non-trade entitlements on the other side is a bit like reading tea-leaves; one has to go hunting for clues about how the WTO framers might have wanted to protect multilateral entitlements.

450 See the discussion of point 5 in subsection 5.4.1.1.
First, the WTO’s system of entitlement protection fails to draw the line between welfare-enhancing regret contingencies and opportunistic behavior: On the one hand, injurers acting in good faith and striving to react to regret contingencies cannot easily do so legally. The scope for good-faith escape from previously made concessions is too narrow to suit signatories’ trade policy flexibility needs, especially concerning market access obligations. Injurers are much better off resorting to illegal means of opt-out.\footnote{Scholars argue that \textit{de facto} “breaches” of WTO obligations often occur because of the rigidity with which formal escape mechanisms are written and have been interpreted so far. The current WTO safeguards regime allegedly does not sufficiently address Members’ needs for policy flexibility. Barred from the necessary breathing space, WTO Members instead turn their attention to informal means of flexibility (see e.g. Horn and Mavroidis 2003; Finger 1993; Bagwell and Staiger 2002b; Sykes 2003; Mavroidis 2006).}

On the other hand, bad-faith injurers are highly advantaged by the \textit{de facto} system of escape which offers many convenient ways of unencumbered defection.

Second, even if the system could neatly separate acts of good faith from bad-faith \textit{ex post} non-performance, the remedies for legal escape payable to victims would still be chronically under-compensatory. \textit{Prospective reliance damages} do not put the victim in as good a position as if the injurer had performed. The effects of \textit{any} market-closing \textit{ex post} behavior will be partly shouldered by the victim(s) of that measure – a classical externality that the WTO was founded to overcome in the first place. This is not only somewhat unfair, but has repercussions on the cooperative behavior of all those signatories that must expect to become victims at any point of time during the existence of the contract (on that note, see \textit{infra}).

Finally, the tertiary rules of contracting – the system’s \textit{extra}-contractual remedies – are too weak to deter illegal behavior. Insufficient punishment converts legal and illegal flexibility mechanisms into \textit{substitutes}. This propels violation-cum-retaliation as the \textit{de facto} rule of default: The one trade policy flexibility mechanism that is always available and applicable is to violate the Agreement and to endure the bearable consequences in the form of sub-standard countermeasures. Furthermore, \textit{extra}-contractual remedies in connection with violation-cum-retaliation establish a benchmark for \textit{all} contractual remedies available in the WTO: Reasonably rational injurers will not negotiate or settle for any compensation \textit{in extenso} of prospective reliance damages.
5 The WTO system of non-performance

5.4.3.2 **Over-“breach” and under-commitment**

The current system of entitlement protection and flexibility in the WTO sets the wrong incentives for injurers and under-compensates victims of escape. What are the systemic consequences of inefficient entitlement protection?

First, parties cannot distinguish whether an injuring party acts in good faith, defaults by mistake or negligence, or behaves opportunistically. This foments suspicion in victims that every *de facto* flexibility behavior is driven by opportunistic guile. An aura of mistrust must be assumed to pervade the system.

Second, as was shown in Chapter 3 (subsection 3.3.2.1), whenever a contract – explicitly or implicitly – allows injurers to defect from their previous commitments for less than the real incurred damage (which is the victim’s expectation damage), injurers have an incentive to *over-“breach”* the agreement. Compared to what the hypothetical complete contingent contract would mandate, injurers are inclined to escape inefficiently often. This is opportunistic bad-faith behavior, since it entails welfare-depreciating redistribution to the detriment of the victim and to the benefit of the injurer.

*Over-“breach”* in the WTO not only creates frustration and discontent among victim parties but also destabilizes the WTO in the long run: The dynamic consequences of overzealous non-performance may be threefold:

1. Whenever a contract cannot prevent opportunist activity from happening, contracting parties are reluctant to enter into it in the first place. All those contracting parties who ever expect to become victims in the course of the contract,\(^\text{452}\) will liberalize their *ex ante* trade to a lesser extent than would be potentially achievable under the EBC, the politically realistic first-best contract. Rational anticipating actors will drive down the level of participation and cooperation (initially and in future trade liberalization rounds), because they have to assume that they get “less miles for the buck” in case a trade partner “defaults on them”.

2. Frustration may even lead them to exit the system partially, *de facto*, or fully.\(^\text{453}\)

\(^{452}\) Since the WTO is a long-term, open-ended, repeated-interaction contract where Members trade a myriad of goods day in and out, practically every country must anticipate to assume the role of the victim at some point in time.

\(^{453}\) Note that Member exit is not a dichotomous variable. Rather, there are various degrees of exit. We can witness partial exit in the form of engagement in preferential trading agreements and bilateralism, withdrawal from plurilateral agreements and protocols, resort to extra-WTO policies, or nonparticipation in trade talks.
3. Alternatively or concomitantly, victims may refrain from utilizing the WTO dispute settlement mechanisms in the first place. Instead of litigating, they may engage in extra-contractual means of retribution and/or aggressive self-help behavior: They may opt for bilateral resolution outside the WTO forum,\(^454\) seek retaliation outside the trade realm (e.g. political coercion), engage in unilateral retaliation (e.g. *Section 301 of the U.S. Trade Act of 1974*), enact retaliatory AD action,\(^455\) enter into retaliatory litigation, or design strategic retaliation tactics (e.g. “carousel” retaliation).

Any of these three dynamic consequences can be expected to seriously harm the system and sour the atmosphere and general WTO spirit of trade liberalization.

\(^{454}\) Examples of bilateral arrangements outside the DSB-ambit include VER-type accords, or an offer of favorable GSP-treatment for the victim by the injuring country. These extra-WTO settlements can be win-win solutions for the participating countries to the detriment of non-participating third parties.

\(^{455}\) Note that retaliatory dumping has proliferated vastly throughout the last decade (for empirics on retaliatory AD, cf. Feinberg and Olsen 2004; Prusa and Skeath 2002).
6 The compliance-vs.-rebalancing debate: A missed opportunity

Chapter 5 identified significant shortcomings in the current system of trade policy flexibility, entitlement protection, and enforcement in the WTO. As was shown, the governance structure of the WTO is not well equipped to deal with contractual incertitude. The system of trade policy flexibility, as it stands today, chronically disadvantages victims of contractual breach. The possible consequences for the international trading system were identified. Before we proceed to lay out a comprehensive agenda for reform in Part 3 of this study, we want to pause for a moment to reflect on why WTO scholarship has failed to address these profound systemic deficits displayed by the current WTO system of non-performance.

It was the so-called “compliance-vs.-rebalancing debate”, which brought WTO scholarship closest to tackling the problems inherent in the WTO system of non-performance. As will be shown in this chapter, the compliance/rebalancing controversy missed the mark of identifying the real problems of the WTO governance structure of non-performance. Reviewing the origin and essence of the compliance/rebalancing controversy which has been simmering inconclusively for the last 10 year, we will find that scholars on both sides overlooked the need to refocus their controversy: Its disputants put an unduly rigid perspective on WTO enforcement, and at the same time neglected systemic issues of contracting. Instead of solely concentrating on the object and purpose of WTO enforcement, the scope must be much broader: The compliance/rebalancing rift must be redefined as a controversy over the nature of the WTO contract and its system of trade policy flexibility and remedies in general. Reframing the research questions which should have been at the core of the debate from the very beginning, we find that neither of the schools of thought has succeeded in giving a convincing and accurate picture of the WTO contract so far.
6.1 The origin of the compliance/rebalancing debate: The objective of WTO enforcement

At its inception on January 1, 1995, the WTO Dispute Settlement Understanding was termed the centerpiece of WTO architecture, “the jewel in the crown of the WTO”. It was hailed by observers and practitioners as the major innovation towards the establishment of rule of law in the global trading order (see exemplarily Davey 2005a). As time went by and countries collected DSU experience, disillusion about the system of enforcement slowly settled in. Over the last few years, the WTO dispute settlement has been facing severe criticism voiced by trade practitioners and WTO scholars as well as civil society advocates and developmental NGOs. Next to issues of adjudication, the center of criticism is focused on the system’s allegedly slack enforcement instruments.456

The countermeasure of retaliation and the size of arbitration awards, in particular, have received scathing criticism from free-traders and trade skeptics alike (see the discussion in section 5.4.1.3).457 The remedy of retaliation has been attacked from two sides: The prevailing reproach is that retaliation does not have enough “teeth” to counter persistent non-compliance with rulings of the DSB. It is argued that retaliation fails to deter countries from upholding measures in violation of WTO law (e.g. Bronckers and van den Broek 2005; Charnovitz 2001; Mavroidis 2000; Pauwelyn 2000), and that retaliation as a matter of legal interpretation is largely under-compensatory to the victims of a violating measure (e.g. Pauwelyn 2000; Spamann 2006; Trachtman 2006; Breuss 2004; Bronckers and van den Broek 2005; Schropp 2005). Other authors, in contrast, bemoan that, at times, DSB panels showed too much “teeth”. Especially in disputes concerning subsidies, some authors contend that instances of over-compensatory retaliation were recorded (cf. Greenwald 2003; Lawrence 2003, chapter 3; Mavroidis 2000).

WTO Members, cognizant of the weaknesses inherent in the DSU system of tariff retaliation, have started complaining openly about the faults of WTO’s dispute settlement regime. Eighty-nine proposals to reform the DSU have been put forward in the DSU negotiations so far (as of July 2007) – by individual Members or country

---

456 Issues of adjudication concern the manner in which the DSB operates. The main reproaches are usually that (i) the DS-system has become too legalized; (ii) the DSB is exceeding its mandate and engaging in “judicial activism”; and (iii) WTO dispute settlement procedures lack transparency (see WTO 2007, section II.D.3.c(ii) for analyses and literature references).
groups from all geographical regions, initially covering 24 out of the 27 Articles of the DSU.\textsuperscript{458} Some of the tabled reform suggestions are concerned with institutional or procedural reforms.\textsuperscript{459} A good many, however, bring forth substantive reform proposals along various trajectories – and indeed often heading in opposite directions, some requesting stronger punishment for defection, some suggesting more lenient sanctions.\textsuperscript{460}

Lest the observer gets completely confused in the face of the multitude of reform suggestions, it seems requisite to first engage in a discussion of what is effectively wrong with the current rules and procedures of WTO enforcement – and why so. A handful of authors correctly hold that, as a matter of logic, it is crucial to have a clear conception of the object and purpose of WTO dispute settlement before addressing the subordinate

\textsuperscript{457} Jackson (2004, p.123), for example, judges: “In many ways, the DSU provisions on remedies, especially the temporary measures of compensation and suspension, are deeply flawed, and even dysfunctional. WTO members are questioning whether they really contribute to the effectiveness of the DS system.”

\textsuperscript{458} A number of WTO Members, including several developing countries such as the African Group, Mexico, and Chile, have spoken in favor of opening renegotiations on DS and have tabled proposals for DSU reform. See generally the WTO document series TN/DS/xxx.

\textsuperscript{459} Institutional reform issues include proposals such as the introduction of a system of permanent panelists. Procedural issues are concerned with issues such as enhancing transparency, allowing amicus curiae submissions, introducing remand authority for the AB, improving the regime on the RPT, refund of litigation costs for least developed countries, etc.

\textsuperscript{460} See WTO (2007, section II.D.3.c) for the most salient DSU reform proposals. Zimmermann (2006) provides for an excellent overview of the history and issues of DSU reform. Some substantive reform proposals tabled by WTO Members concern the multilateral level. Such community-based innovations include making tariff compensation mandatory and automatic (proposals come from the Meltzer Commission to the US Congress [International Financial Institutions Advisory Commission Report, March 2000], Lindsey, et al. 1999; Charnovitz 2001, 2002a, 2002c; Pauwelyn 2000; Roitinger 2004, pp. 188, and others). Other multilateral reform proposals of the WTO enforcement regime include the suspension of membership rights, for instance the right to attend meetings, to participate in decisions, to use the DSM or other WTO provisions or to receive technical assistance (Charnovitz 2001; Lawrence 2003, chapter 5). India and nine other developing countries expressed the need for a collective retaliation scheme along the lines of the “principle of collective responsibility” championed in the UN Charter (TN/DS/W/19, see also Maggi 1999a; Pauwelyn 2000). Mexico tabled a proposal which directly addresses the issue of asymmetry between WTO players, namely tradable remedies. Bagwell et al. (2005), and Limao and Saggi (2006) offer theoretical support. Multilateral soft-law reform avenues (including improved surveillance of Members’ implementation of DSB reports, to rally international mobilization vis-à-vis non-compliant states, technical and financial assistance, reporting, and transparency of information) have been proposed by advocates of the “managerial school” (see Charnovitz 2002b; Chayes and Chayes 1993a; Guzman 2002a; Mitchell 1993). (Cf. also our discussion in footnote 73.) Another strand of reform proposals has argued against multilateral and in favor of bilateral mechanisms: Some authors simply argue for “stiffer”, i.e. more coercive retaliation penalties so as to induce WTO Members into compliance (see for instance Busch and Reinhardt 2002; Bown 2002a). Other proposals (e.g. Charnovitz 2002b; Hudec 2002; Subramanian and Watal 2000) suggest the increased and easier use of cross-retaliation. The 2003 joint reform proposal of the U.S. and Chile (TN/DS/W/52) requests less multilateralism and panel interference into what the two governments view as an essentially bilateral settlement process. Stimulating the use of the countermeasure of bilateral tariff compensation is another reform avenue promoted (Lawrence 2003; Pauwelyn 2000; Schropp 2005). Replacing bilateral tariff compensation with monetary fees has also been proposed (Davey 2005b; WTO 2004; Bronckers and van den Broek 2005).
discussion of enforcement tools and instruments.\footnote{The precedence of why? (enforcement objectives) over how? (enforcement instruments) is after all a logical prerogative. Unless we know what the DSU is meant to do, it is difficult to discuss whether it performs well. Unfortunately, this first step is quite frequently missing in the literature: As Hauser and Roitinger show (2004), most trade practitioners and scholars tend to question enforcement instruments – without addressing what object and purpose the enforcement regime is to achieve in the first place. Instead, these groups have merely argued about conclusions drawn from divergent sets of underlying – implicit – assumptions.} Otherwise it would be simply impossible to give meaning to the ongoing criticisms of WTO enforcement. There is, however, wide disagreement among WTO scholars as to whether DS is to compensate the victim, or whether DS procedures should be aimed at inducing compliance by deterring future deviations from happening. WTO scholarship is thus neatly divided into two camps: A “compliance” school of thought and a “rebalancing” school of thought (Charnovitz 2001, pp. 802; Lawrence 2003, chapter 2).

The compliance/rebalancing debate was triggered by a *AJIL*-editorial by Judith Hippler Bello (1996). Her article was a reaction to American opponents to the WTO Agreement, who feared that the treaty would threaten U.S. sovereignty by allowing unelected bureaucrats to undermine democratic legislation. Bello reassured her compatriots that Members in breach of a WTO Agreement have a distinct choice between adhering to a dispute panel’s/the AB’s recommendations (mandating that the Member in question brings its practices or law into consistenc y with the texts of the Annexes to the WTO Agreement) on the one hand, and deliberately “opting out” of the respective agreement on the other. The precondition for opting out, the author claims, is the provision of compensation, or – in case mutual agreement is lacking – the toleration of a suspension of concessions. Concerning the bindingness of panel recommendations, Bello contends that “[t]he only truly binding WTO obligation is to maintain the balance of concessions negotiated among members” (ibid., p. 418). This view was vividly refuted by John Jackson (1997b), who contended that a WTO Member is in no way free to “pay or perform” after having been condemned by the international trade court.\footnote{The term “pay or perform” is linked to Judge Oliver Wendell Holmes’ famous dictum, namely that private contracts are promises to perform according to the letter of the agreement, or to opt out of the deal and compensate by paying damages (Holmes 1920 at p. 175).} Rather, each Member is under a strict international legal obligation to comply with a dispute panel’s recommendations.

It soon became obvious that Bello’s and Jackson’s contentions on the legal status and effect of dispute settlement reports were each representative of two entirely distinct schools of thought, perspectives or preconceptions of WTO enforcement. The two mindsets, in fact, seemed to neatly divide WTO scholarship into two rivaling camps –
that of compliance advocates and that of rebalancing proponents. The academic dispute of compliance-vs.-rebalancing is waged in only a handful of papers, in which authors of both schools of thought attempt to frame the controversy, to condense the logic inherent in their respective contention, and to disqualify the other camp. Yet the framers of this controversy apparently see their contentions as representative of an entire school of thought. In support of their respective view, authors draw widely upon relevant WTO literature, extract the underlying (largely implicit) assumptions and methodologies therein, and bring forth their inferences accordingly. The argumentative logic behind these two rivaling mindsets has to be analyzed: The scope and dimensions of the compliance/rebalancing cleavage will become apparent.

6.1.1 Rebalancing as key objective of enforcement

Basically, the rebalancing camp basically views the objective of dispute settlement as supplying an insured safety-valve for injurers in a non-stationary world. By equilibrating the mutual balance of concessions, DSU enforcement mechanisms ensure the twin-goal of compensating the victim and of providing the injurer with an efficient opt-out possibility.

6.1.1.1 Theoretical foundations of the rebalancing approach

Scholars delivering the theoretic foundations of the rebalancing camp tend to originate from the disciplines of trade economics, law & economics, and the international political economy branch of IR. This approach to WTO enforcement is highly influenced by economic logic and the paradigm of rational choice. The rebalancing approach to WTO enforcement is rooted in the conviction that the world trading system is fundamentally driven by reciprocal promises of trade liberalization which give rise to a “balance of
concessions”.\footnote{The rebalancing view of the WTO is best described as “a system of reciprocal rights and obligations to be maintained in balance”, as Dam (in: Charnovitz 2001, p. 802) notes.} For a large faction of WTO-scholars, enforcement in the WTO-context is tantamount to achieving a rebalancing of commitment levels in case the original equilibrium of concessions is out of synch.\footnote{For a legal explanation of rebalancing, see Dam (1970, pp. 357); Charnovitz (2001 at p. 801; 2002c, pp. 414); and Jackson (2004, p. 109; 1997a at p. 60). For an interpretation of the economic principles of rebalancing in dispute settlement, cf. Lawrence (2003, pp. 19); Bagwell/Staiger (2002b, pp. 58, 104); Ethier (2001a, p. 3); Rosendorff (2005, pp. 390); or Bown (2002a, p. 288).}

According to rebalancing proponents, the essence of the WTO – just like the preceding GATT – is a mutual, correspondent exchange of market access (export) opportunities that WTO Members grant to each other: Signatories to the WTO agree to constrain their subsequent behavior because they believe this to be a mutually beneficial arrangement. Market access concessions are the currency of the trade deal. Therefore, the WTO is best conceptualized as a web of bilateral market access contracts (Lawrence 2003, chapter 2) or a “packages of bilateral equilibria” (Pauwelyn 2000 at p. 340). Rational bargaining by Member governments now mandates that in the initial trade negotiations market access concessions be granted on a reciprocal basis, creating the balance of concessions.\footnote{According to Finger and Winters (2002, pp. 50) reciprocity has been the motivating principle of the GATT/WTO system (cf. also Bagwell and Staiger 2002b, chapter 4; 1999).}

This initial “balance of benefits and burdens contemplated in the covered agreements vis-à-vis other Members” (Vazquez and Jackson 2002, p. 563) as the core of the trade accord is to be defended against any kind of ex-post backtracking by any signatory. Backtracking, or partial reneging of initially granted concessions, can take any form whatsoever: It can be voluntary or accidental, open or concealed, formal or informal, temporary or permanent. Form notwithstanding, its substance, namely a partial denial of initially granted market access (analogously: the erection of some protectionist barrier) stays the same, and needs to be made up for: If ever the initial balance of concessions happens to be in disequilibrium, it nullifies or impairs advantages previously promised to the victim party. In such circumstances, contracting parties undertake efforts to restore the equipoise of initially conceded economic advantages and to bring the balance back into equilibrium (hence: re-balancing). The general techniques for achieving rebalancing vary: negotiation and out-of-court settlement, withdrawal of the offending measure by
the injurer, alternative offers of compensation in sectors other than the one in question, or suspension of concessions enacted by the victim.469

For proponents of this economicist view, the inception of the WTO DSM substantially improved the old GATT rules for settling disputes, but did not alter the fundamental nature of the negotiated bargain among sovereign Member States.470 As in GATT times, the proceedings of the new WTO dispute settlement are neatly embedded in the wider systemic logic permeating trade agreements. Trade disputes – be they under GATT 1947 or the WTO – are indicative of the fact that the original market access balance is out of synch. Enforcement measures and instruments under the WTO DSU now have to ensure that proper rebalancing occurs. Tariff retaliation features prominently due to its undeniable advantage of being self-enforceable. However, the victim should not be allowed to react with overzealous retaliation (Schwartz and Sykes 2002b), nor should the injuring Member be permitted to hitch a free ride by escaping its compensation duty or endurance of retaliation. The magnitude of compensation and retaliation is logically limited to being strictly commensurate to the damage caused by the “imbalancing” act.471

This “proportionality principle” (Ethier 2001a) renders punitive damages logically incompatible. In their damage calculation (Arts. 22.6 and 22.7 DSU), panels and the AB are strictly bound by rebalancing requirements.

Drawing analogies to economic theories of private commercial contracts, some L&E scholars have gone a step further: Bello (1996), Sykes (2000), Schwartz and Sykes (2002b), Hauser and Roitinger (2003), Dunoff and Trachtman (1999), and Trachtman (2006) claim that the purpose of WTO DS goes well beyond securing a passive rebalancing, in case some external shock disturbs the initial balance of market access concessions. Rather than serving as a mere insurance for potential victims, dispute procedures also benefit injuring parties: The DSU, the authors contend, facilitates (and

469 Hence, as with any ordinary balance, equilibrium can be restored either by removing weight on one side (i.e. the injuring party withdraws the initial measure, or offers compensation in return for nullification and impairment), or by adding weight on the other side (i.e. the violated party reciprocates by enacting retaliation measures). See Pauwelyn (2000 at pp. 320).

470 Under the rebalancing view, the inception of the DSU into the framework of the WTO treaty constituted an antidote to the procedural deficiencies of Arts. XXII and XXIII of the GATT – but did not change its deeper logic (Schwartz and Sykes 2002b; Hauser and Roitinger 2003). To liability advocates, the DSU mainly codified the evolution of GATT trade dispute practice from the “1979 Understanding Regarding Notification, Consultation, Dispute Settlement and Surveillance” to the “Montreal Rules” from 1989 (Mavroidis 2000, p. 777).

471 Although in contrast to Art. XXVIII GATT permanent compensation is not permitted de iure, rebalancing proponents argue that it is de facto a possibility: Contracting parties just come to a mutually accepted agreement and drop the dispute.
protects) contractual non-performance in situations where standard provisions for renegotiation of tariff bindings, or contingency measures for deviation under specified conditions, are inadequate or insufficient in ensuring efficient adjustment to changing circumstances (Schwartz and Sykes 2002b, section III; Sykes 2000, p. 348).

According to these authors, the WTO is an incomplete contract which cannot explicitly take into consideration the complexities of the relationship between Members and which does not anticipate all future contingencies. Performance strictly according to the (rather limited) letter of the treaty is not always jointly beneficial when circumstances change. In a non-stationary world contractual performance is desirable only if the cost to the non-performing party is exceeded by the gain from performance to the other party. Where this condition does not hold, non-performance is jointly preferable and the conditions for efficient “breach” are met.

The strict rebalancing requirement now fulfills a double role and solves Ethier’s “reciprocal-conflict problem” (cf. footnote 23): On the one hand commensurate punishment compensates the victim of a backtracking measure (and therefore formally insures it). On the other hand, the requirement of substantial equivalence of the countermeasure to the damage done allows the injuring party to make use of the DSU as a ready safety valve, without being punished by overzealous sanctions on behalf of the victim. This dual role of the rebalancing principle in essence ensures that rational parties use the safety valve flexibility that the DSU offers only in situations where adjustment to outdated modalities non-performance is effectively Pareto efficient (Bagwell and Staiger 2002b, chapter 6; Ethier 2001a; Rosendorff 2005; Rosendorff and Milner 2001).

472 The proportionality principle of rebalancing would explain why WTO sanctions (just like sanctions under the preceding GATT) under Art. 22.4 DSU are so “toothless” in nature. For Schwartz and Sykes (2002a section IV.B at p. 26) the main improvement of the DSU (over the old GATT system) was the institutionalization of the efficient “breach” principle: “[T]he innovation of the DSU was intended not so much to deter violation of most substantive rules […] What the system really adds is the opportunity for the losing disputant to ‘buy out’ of the violation at a price set by an arbitrator who has examined carefully the question of what sanctions are substantially equivalent to the harm done by the violation […] The new system does a better job of protecting violators from the actual or threatened imposition of excessive sanctions. In turn, it ought to perform better than the old system at ensuring that opportunities for efficient breach are not undermined.”

473 For rebalancing proponents a growing use of the DSM and the countermeasure of retaliation is a sign of a system at work – not at fault. The DSU mechanisms are the quasi-legal opt-out or trade flexibility mechanism that countries need in order to react appropriately to the ever-changing realities of world trade. For them, the clamor about rising non-compliance or a spiraling into “trade wars” is in fact a non-issue, since tariff retaliations are proof of a functioning system of flexibility (Hauser and Roitinger 2003).
6.1.1.2 De iure support for the rebalancing view

So much for the theoretical underpinning of Bello’s original contention that “compliance with the WTO, as interpreted through dispute settlement panels, remains elective […] A government can change its mind and raise a particular tariff, provided it offsets such nullification and impairment of the delicate GATT balance through compensatory tariff reductions” (1996, p. 417). Sykes (2000) and Schwartz and Sykes (2002b; 2002a) have argued that the treaty language of the WTO in its present form – particularly the rules and procedures of the DSU – is in strict consistency with the Holmesian pay-or-perform system: A potential injurer has the right to unilaterally opt out of, or escape, any treaty obligation – inasmuch as the victim of the measure gets compensated for the level of nullification and impairment it sustained through the measure in question.

According to the authors, the WTO treaty provides for a general de iure fallback rule of liability by virtue of the DSU mechanisms and procedures, mainly DSU Arts. 3.7, 19.1, 22.1, and 22.4 (cf. Sykes 2000, pp. 349). This implies that every contingency laid down in the Agreement can be reneged upon, and that every contractual gap can be seized by the injurer – given the victim’s level of expectancy is preserved. The existence of non-violation claims allegedly weighs heavily in their favor (e.g. Bagwell 2007). The victim – agnostic as to how the initial level of market access is recovered – may always unilaterally re-establish the balance of concessions by engaging in tariff retaliation.

To reiterate: Under the presumption of a general liability rule of entitlement protection, all opt-out behavior is permissible and covered in the terms of the Agreement. Ultimately, this would imply that no extra-contractual behavior exists. The DSU as a legitimate tool for non-performance “internalizes” all sorts of reneging behavior in the terms of the Agreement. Contractual escape coupled with indemnity (damages payments) is part of the rules.

6.1.2 Ensuring compliance as a key objective of WTO enforcement

The compliance school of thought takes an entirely different perspective on the object and purpose of WTO enforcement and on the role of dispute settlement procedures. This
faction of trade scholarship contends that WTO enforcement is to induce compliance with panel/AB rulings, and to deter future violations for the sake of legal predictability and the stability of the global trading system. Compliance proponents submit that not rebalancing, but inducing ongoing and strict compliance with the explicit terms of the Agreement, is the key objective of WTO dispute settlement, and that the legal effect of an adopted panel report is the international legal obligation to adhere to the panel/AB rulings.

6.1.2.1 Theoretical foundations of the compliance approach

Whereas the rebalancing view of the WTO borrows extensively from the economic theory of private contracts, the compliance school of thought is deeply rooted in the discipline of public international law. For international lawyers it is the treaty text that sets the parameters of Member conduct. Compliance scholars rely mainly on textual evidence and on other standard legal treaty interpretation techniques: Subsequent practice; interpretation of the treaty by panels, AB, and trade practitioners; drafters’ intent, i.e. preparatory notes; context of the object and purpose of the treaty as found in other passages of the Agreement.

According to compliance theorists, the rules of the game in the world trading system changed decisively with the inception of the WTO in 1995. Whereas the GATT 1947 was a negotiation forum ruled by power politics and diplomacy, the modern-day WTO represents a rule-of-law-based international economic regime guided by strong underlying norms and values.475 “This gravitation of the whole system toward rules suggests that rebalancing has less and less of a role to play in the WTO context” (Jackson 2004, p. 121; cf. also Charnovitz 2001; Charnovitz 2002c; Pauwelyn 2000). Compliance advocates contend that the basic values and objectives cherished by the “new” WTO are security and predictability of the world trading system (as Art. XVI of the WTO Agreement and Art. 3.2 DSU instruct Members). The aim of the WTO is hence to protect the expectation and competitive relationships of all economic agents, including traders and commercial entities, consumers and uninvolved third party

474 Termed differently, for proponents of the rebalancing paradigm the DSU (rather: violation-cum-retaliation) is the liability default mechanism of the WTO that kicks in whenever unanticipated circumstances occur. Although the DSU procedures are conceptually similar to other treaty provisions (e.g. GATT Arts. XIX or XXVIII), the latter are limited in their scope of application by the hefty conditionality laid out in the text of the articles (Schwartz and Sykes 2002b, section II). That this view conveniently omits reference to Art. 22.1 DSU, which forestalls this notion, shall be discussed in subsection 6.3.2 infra.

475 See our related discussion in subsection 5.4.1.1 at point 5.
governments. This, so compliance backers claim, presupposes a credible, objective and apolitical DS that redresses power asymmetries and levels the playing field between large powerful States and small weak ones. Aspects of rebalancing or efficient breach are not “central or even operative in the normal DS processes” (Jackson 2004, p. 118). In addition, rebalancing undermines the longer-term goals of legal security and predictability, which can be ensured only by constant application of the rules and by binding third-party arbitration. “This feature is important to every type of juridical system, whether national or international” (ibid.).

In addition to a WTO-intrinsic argumentation of the compliance effect of panel reports, compliance advocates resort to the tenets of general public international law. The WTO, property advocates claim, is not a self-contained regime that exists in isolation, but is fundamentally rooted in the grander scheme of international public law, as the very first AB report made clear.476 A DSU panel recommendation (which is quasi-automatically adopted by the DSB thanks to the reverse consensus rule) is a court-ordered sentence – just like in any domestic or international context. Members condemned for having violated a WTO Agreement are therewith infringing upon their obligations under international law. Under public international law these Members have the duty to comply with dispute settlement rulings (Grané 2001; Jackson 2004, 1997a, 1997b; Mavroidis 2000). There is no option of deviating from this court-ordered international legal obligation – everything else would render the international law meaningless.477 Opting out by disobeying a conviction brought forth by an international adjudicating body is an abhorrent thought to international lawyers.

This view of the unconditional compliance-inducing nature of WTO enforcement is analogous to what we know in contract theory as a property rule of entitlement protection: The compliance view of the WTO concedes that the treaty is indeed an incomplete contract in a non-stationary environment (Sykes 2000, p. 347), but fundamentally disagrees with rebalancing proponents as to the way efficient ex post adjustment is to take place. Whereas under the precepts of a rebalancing system an injurer can adapt to changing circumstances unilaterally by enacting a general liability

476 US – Gasoline, informally known as the Superfund case (WT/DS 2 and 4, at §17).
477 To compliance advocates, disregarding a court’s ruling is the same as disregarding the rule itself. If a party to the treaty is permitted to disobey the rules, the entire system of law is futile. DSB rulings are international legal obligations erga omnes partes liable to the community of Members, and not just owed bilaterally (Pauwelyn 2001). Panel recommendations are public goods and not just a matter of concern to the disputing States – not only due to the precedence they set, but more so for the legal stability of rules they ascertain (Jackson 2004, pp. 120).
rule, compliance advocates beg to differ. Under a PR of enforcement, contracting parties have a right to specific performance of the terms of the Agreement. A party wishing to change or modify the agreed-upon terms can always approach another party and attempt to negotiate a release from performance. No backtracking measure can be enacted against the consent of the victim, or else international law is violated.

Concretely this means that the only permissible way of escaping unilaterally is prescribed exhaustively in the treaty’s contingency measures, as well as in Art. 3.4 DSU (“mutually agreed solutions have to be consistent with the obligations of the covered agreements”).478 Dispute settlement procedures are – legally and logically – separate from these clearly circumscribed escape options:479 “But [despite the possibility to escape under certain well-described circumstances,] the ultimate idea that full compliance is an international law obligation can still be crucial to the notion of a rule-oriented system that is objective and creditable and provides for a basis of security and predictability for all members of the organization, as well as non-governmental beneficiaries of the system” (Jackson 2004, p.122).

Treaty enforcement plays an integral role in the compliance system. Since every party is contractually obliged to specific performance, any extra-contractual behavior must be sanctioned and future deviations deterred (Mavroidis 2000, p. 811). Punitive damages in the form of penal trade “sanctions” are not alien to this perspective – although the letter of Art. 22.4 DSU (but not 4.10 SCM) presently forestalls this option (see for example Charnovitz 2001). Nevertheless, extra-contractual remedies are to be interpreted as penalties (cf. Sykes 2000, p. 351) intended to coerce recalcitrant violators into obedience, and to deter all WTO Members from future defections. A link between court-ordered remedies on the one hand, and the damage caused by an injuring party on the other, is a necessity only due to contractual provisions to the contrary – not due to systemic logic. In order to improve on the efficiency of the system, WTO remedies should be decoupled from the amount of damages suffered. Punitive damages, monetary fines and other collective penalties are perfectly acceptable in principle, since they

---

478 Although certain contractual provisions in the WTO allow for a limited liability-rule application (Arts. XIX, XX, XXI, XIV GATT, e.g.), the applicability of these rules is strictly confined to the contingencies laid down therein.

479 Charnovitz (2001, p. 818) notes that Art. 22 DSU is not to be used as a safeguard. Jackson (2004, p. 121) states that the two concepts of safeguards and dispute settlement were distinctly separated in the GATT drafting process, and that four decades of history let dispute settlement evolve rather consistently and persistently towards a more juridical, rigorous, and creditable system.
disincentivize WTO-inconsistent behavior and induce compliance. All incentives to adapt to changing circumstances would thereby be channeled through the mechanism of mutual consent and renegotiation.

6.1.2.2 De iure support for the compliance view

In order to support their contention that the WTO features a *de iure* compliance obligation, WTO authors of the compliance camp resort to “textually oriented” (Jackson 2004, p. 111) legal interpretation techniques. To buffer their argument, WTO lawyers point to various separate passages in the WTO DSU, Preamble and text of the Marrakech Agreement, preparatory notes, and various panel and AB reports. Charnovitz (2001; 2002c) sets out to show how the perception of the WTO has evolved over the years – from a rebalancing to a compliance objective of WTO enforcement. In support of his argument the author extensively cites official WTO documents, GATT and WTO negotiating history, statements from trade practitioners and the assertions of learned scholars. Compliance advocates see no textual and international legal leeway for anything but a general rule of unquestioning specific performance, i.e. strict compliance with the letter of the law. For them, the treaty can be augmented or modified only through mutual consent. If a contractual gap appears in the course of the conduct of the WTO treaty, or if one party is significantly unhappy with either the concessions promised, or with pertinent passages of the treaty text, it is the respective Member’s duty to engage in formal, constructive renegotiations.

In summary, in the current controversy over purpose and objective of WTO enforcement scholarship is still divided into two opposing perspectives. Both schools of thought acknowledge the incompleteness of the WTO contract. The rebalancing camp, on the one hand, views enforcement as a general liability rule with the aim of providing an insured safety-valve in a non-stationary world. DSU enforcement is to ensure the twin goal of compensating the victim and providing the injurer with an efficient opt-out possibility. The compliance faction of WTO scholarship, on the other hand, contends that WTO enforcement is to induce compliance with panel/AB rulings, and to deter future violations for the sake of legal predictability and stability, as Art. 3.2 DSU

480 Grané (2001, p. 763) notes: “Even if we accepted that both compensation and withdrawal of concessions were originally conceived as counterbalancing actions or re-balancing of trade concessions, whose objective was to reestablish the lost balance of rights and obligations, they are nowadays perceived in practice as sanctions or punitive measures for failure to comply with the adopted recommendations of the DSB. At best, they are considered means to induce compliance” (emphasis added).
mandates. The WTO is thus conceived as a general property rule, backed by the strong enforcement belt of unambiguous language in the dispute settlement understanding.

6.2 Why the compliance/rebalancing debate is misguided: Research questions reframed

The compliance/rebalancing controversy seems to have reached an impasse. The conflict between the two perspectives has been described as a “discours des sourds” by Hauser and Roitinger (2004, p. 641) for good reason: Although the controversy should have stimulated closer occupation with the other camp’s underlying assumptions, authors on both sides of the divide have just deepened their trenches. They continued promoting their respective perceptions and methodologies, without considering alternative DSU objectives brought forth by the other school of thought.481 The fault lines between economistic logic (“rebalancers”) and legalistic thinking (“compliancers”) seem unbridgeable: Whereas rebalancing proponents point to the realpolitik of Members’ de facto opt-out behavior in support of their thesis (as the disputes EC – Hormones, EC – Bananas, US – FSC, or US – Byrd482 seem to confirm), compliance advocates find it hard to fathom that anybody seriously challenges the literal bindingness of international legal obligations.

Occasionally, the controversy has been tagged as “insignificant” or “minor” in recent mainstream WTO scholarship.483 This belittlement is not a proper solution to what is effectively an important issue.

Why did the compliance/rebalancing controversy fail to gather momentum for reaching a solution and is unresolved as ever? We believe that this discussion constitutes a missed opportunity to address in a meaningful and scientific manner the pressing issue of trade policy flexibility in the WTO. The compliance/rebalancing debate should not be a discussion about the nature of enforcement (as both compliancers and rebalancers made believe). Rather, it should be properly conceived of as dealing with the much more encompassing issue of “breach”, breach, and remedies in general. Instead of grasping

481 The latest rebalancing contributions include Bagwell (2007), Rosendorff (2005), Bagwell and Staiger (2005a), and Herzing (2005). Recent compliance publications include (Bronckers and van den Broek 2005; Medrado 2004; Petersmann 2003).

482 DS 26, 48; 16, 27; 108; 217, respectively.

483 A WTO Appellate Body Member in personal communication with the author referred to the compliance/rebalancing debate as a “storm in a teacup”.

- 255 -
this broader scope of their debate, they maneuvered themselves into the cul-de-sac of discussing the nature of WTO enforcement.

As a matter of scientific rigor, and as a requisite of logic, it should have been realized that a discussion of enforcement is vacuous without a prior occupation with the nature of the contract. In the absence of a good grasp of the essence of the underlying contract and a proper deliberation of what constitutes legal behavior in the first place, any examination of rules of enforcement must necessarily remain patchwork. A proper argument by the compliance/rebalancing scholars should have been based on the following suite of questions:

1. Why did countries conclude a trade agreement in the first place, and what are the exchanged entitlements of the contract? (read: What are the primary contracting rules?);
2. How are entitlements protected, and what constitutes legal non-performance? (read: What are the secondary rules of entitlement?);
3. How are entitlements protected from extra-contractual, illegal behavior? (read: What are the tertiary rules of contracting?).

WTO scholarship has not engaged in these questions. Instead, it put the cart before the horse by arguing narrowly over the goal of enforcement and adequate dispute settlement mechanisms. When theorizing about the objectives of WTO enforcement, scholars have quarreled about how to deal with defection from the WTO treaty without having defined before what kind of behavior constitutes a defection in the first place. The debate was thus built on the wrong fundament in the form of an erroneous research question. This is consequential: Asking the wrong questions rarely leads to the right answers.

If we reinterpret the compliance/rebalancing controversy in light of the above suit of questions (which constitute the basic stages of contracting that every contract, including the WTO, adhere to), we come to an interesting conclusion: Paradoxically, as a matter of practice, compliance and rebalancing WTO scholars did broach the broader issues of trade policy flexibility and remedies – only that they did so in an implicit, unarticulated manner. Our presentation of the two schools of thought in section 6.1 above suggests that the compliance/rebalancing debate is at its core less a controversy about defection and sanctions, or about the legal bindingness of panel rulings. Rather, the cornerstone of

---

484 To this day, WTO scholarship has engaged in a dispute akin to whether the referee had made wrong decisions – without before having reconciled just which ballgame is the subject of discussion.
contention in the controversy is effectively the fundamental question of *contractual choice* (or *rule-making*): The compliance and rebalancing approaches are essentially different interpretations of the nature of the contract, and of the permissibility and desirability of partial withdrawal from previously made contractual obligations. To illustrate, consider Chart 6.1:

**Chart 6.1 Rephrasing the compliance/rebalancing debate in WTO scholarship**

<table>
<thead>
<tr>
<th>Contracting stage</th>
<th>School of thought</th>
<th>Rebalancing approach</th>
<th>Compliance approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Initially exchanged entitlements</td>
<td>Market access commitments</td>
<td>n.a.</td>
</tr>
<tr>
<td>II</td>
<td><em>Intra</em>-contractual entitlement protection</td>
<td>Liability rule</td>
<td>Property rule</td>
</tr>
<tr>
<td>III</td>
<td>Enforcement: <em>extra</em>-contractual entitlement protection</td>
<td>n.a.</td>
<td>Coercive punishments</td>
</tr>
</tbody>
</table>

Source: author
Notes: This Chart is a reinterpretation of the compliance/rebalancing debate. Instead of seeing the discussion in light of enforcement (or tertiary rules of contracting, III, shaded bottom row), the debate is reframed within the basic three steps of contracting: The two perspectives are compared in light of their interpretation of the nature of contractual entitlements (or primary rules, I), the rules of *intra*-contractual entitlement protection (secondary rules, II), and of enforcement. *n.a.* stands for “not available”.

Chart 6.1 revisits the compliance/rebalancing controversy in light of the three stages of contract design: the primary, secondary and tertiary rules of contracting (plotted on the vertical axis). Reframing the two schools of thought in that manner shows the actual points of contention between the two rivaling schools of thought (see section 6.3 for detailed analysis):

- The **rebalancing** approach to the WTO reduces the contract’s entitlements to reciprocal market access commitments. These market access commitments are *de facto* protected by a pure liability rule of entitlement protection. Every signatory can opt out of the Agreement anytime under any circumstance given the country pays its
compensation, or at least does not obstruct retaliatory self-help measures on the part of the victim. The rebalancing approach is agnostic over extra-contractual rules of enforcement, simply because under a pure liability-rule system an injuring country is contractually allowed to withdraw any concession at any point of time; this somehow renders considerations of illegal behavior futile.

- The **compliance** approach to the WTO does not distinguish between entitlements. *Every* contractual right and obligation is protected by a property rule of default. Any state of nature that is not explicitly listed in the text as an exception or contingency measure must be addressed through renegotiations. Everything else amounts to a violation of public international law. This property rule of flexibility is protected by the DSU, whose explicit task it is to coerce the violator into compliance with the rules of the game, and with the rulings of dispute panels and the AB.

To conclude: The present compliance/rebalancing controversy built a house with a second floor and neither fundament nor ground floor. The debate is misguided, because it started – and got stuck – in subsequent issues of dispute settlement and enforcement mechanisms. If the compliance/rebalancing debate is reframed and embedded in the larger context of contract design, the two opposite views can be re-calibrated. Comparing the different notions that compliance and rebalancing advocates have on the nature of the contract may help us understand better the confusion that this *discours des sourds* created: As a consequence of having started with follow-up questions, the debate took up a wrong spin, and spiraled out along two different trajectories. In fact, rebalancing and compliance scholars actually describe two different contracts: The rebalancing approach depicts a pure market access exchange contract, and reflects upon *intra*-contractual safety valve mechanisms; compliance theorists portray the WTO as a full-blown, multi-purpose, multi-entitlement treaty and contemplate the international legal bindingness of panel rulings and the enforcement of WTO rule violations in general. It is hardly surprising that the two sets of theories do not match well.

### 6.3 A critique of the compliance and rebalancing view

Repositioning the two views within the larger context of the nature of the WTO contract has additional advantages over and above explaining the origin of the *discours des sourds*. It also helps us assess how well the two views actually describe the WTO contract, and where it is inconsistently argued. We will assess issues that both schools of
thought fail to address. This is followed by a review of why and in what respects the two perspectives are unsuccessful in explaining the true nature of the WTO contract.

### 6.3.1 Omitted: The inalienability approach to the WTO

Had scholars engaged in a discussion over the nature of contractual obligations instead of getting entangled in the downstream issue of enforcement, they would have realized that there are at least three – not two – rivaling mindsets in WTO scholarship: Chart 6.2 illustrates that in addition to a compliance approach (which implicitly propagates a general property rule of default), and a rebalancing view (which alleges a liability-rule system of default), we can extract an *inalienability* perspective on the nature of WTO obligations which posits immutability with WTO rules.

**Chart 6.2 Extending the compliance/rebalancing debate**

<table>
<thead>
<tr>
<th>Contracting stage</th>
<th>School of thought</th>
<th>Compliance approach</th>
<th>Rebalancing approach</th>
<th>Inalienability approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Initially exchanged entitlements</td>
<td>Inalienability approach</td>
<td>Market access commitments</td>
<td>Trade liberalization commitments</td>
</tr>
<tr>
<td>II</td>
<td>Intra-contractual entitlement protection</td>
<td>Compliance approach</td>
<td>Liability rule</td>
<td>Inalienability rule</td>
</tr>
<tr>
<td>III</td>
<td>Enforcement: extra-contractual entitlement protection</td>
<td>Rebalancing approach</td>
<td>Coercive punishments</td>
<td>Coercive punishments</td>
</tr>
</tbody>
</table>

Source: author

Notes: This Chart is an extended version of Chart 6.1 above. There it at least one more conception of the nature of the WTO, namely the inalienability approach (grey column).

The inalienability perspective has a very different, and indeed rivaling, notion of the *nature* of WTO law as contracted: It views the WTO treaty as a *trade constitution* and as such as immutable. For proponents of the inalienability school of thought the WTO is a direct extension of national constitutions, aimed at securing the unconditional abidance of basic economic property rights against protectionist backtracking or any kind of governmental opportunism. Countries, or rather peoples, engage in welfare-enhancing
trade liberalization. Governmental opportunism under the inalienability perspective is seen as a form of unlawful expropriation.\textsuperscript{485} Theoretical support for this claim comes from the “constitutionalism” variant of the commitment approach to trade agreements (see subsection 4.1.2.2).

According to inalienability proponents, WTO rules were deliberately conceptualized by Member States in a rigid manner, so as to tie the hands of current and future trade policymakers, who may otherwise be prone to fall prey to domestic protectionist pressures. Constitutional rigidity upholds the predictability and stability of the system (as expressed in Art. 3.2 DSU), and advantages non-state actors who are normally under-represented in the domestic trade policymaking process (consumers, enterprises). Just as you cannot step back from constitutional obligations in unspecified circumstances, so the argument goes, it is equally pernicious to change the content of the WTO agreements in any way.\textsuperscript{486} For another group of scholars, inalienability is just a logical consequence of the most widely acknowledged \textit{ius cogens} norm of \textit{pacta sunt servanda}. A rule of inalienability then is just the codification of this legal dogma.\textsuperscript{487}

An inalienability view of the WTO is one of a “renegotiation-proof” contract that cannot be enhanced, modified, or adapted in any manner except in ways originally laid down in the text. The issue of whether the WTO treaty is a complete or an incomplete contract notwithstanding, proponents of the inalienability school argue that any gain connected to the exercise of trade policy flexibility is outweighed by general (immediate, long-term, or systemic) welfare losses entailed in a system of \textit{ex post} non-performance.\textsuperscript{488}

\textsuperscript{485} Protectionist measures by domestic policymakers do not only give rise to economically inefficient production decisions, and therewith disadvantage domestic consumers and consuming industries. Moreover, retaliatory actions by victim countries may be targeted at uninvolved domestic sectors which have not been at fault in any way. A retaliatory denial of market access then deprives the targeted industries of their economic right to do business as previously consented to – it may reach a form of expropriation. In addition, the WTO rules on suspension of concessions or other obligations are sort of a “collective liability”, in which uninvolved parties get punished for acts not committed by them. Collective liability is a controversial legal concept. (See Bronckers and van den Broek 2005; Charnovitz 2001 at note 159; Petersmann 2002, 2003).

\textsuperscript{486} Also, since in some countries WTO jurisdiction has a \textit{direct effect} (or “self-executing” effect) on domestic legislation, reneging on WTO rules would be tantamount to infringing domestic legislation – a precedent with unfathomable consequences (cf. Jackson 1997b, pp. 61).

\textsuperscript{487} Dunoff and Trachtman note (1999, p. 32): “Some international lawyers will reject the concept of efficient breach on a normative basis. They might argue that accepting the efficient breach hypothesis would threaten precisely the feature that renders treaties the ‘major instrument of international cooperation in international relations’ – the belief that treaties will be obeyed, even when contrary to the state’s immediate, short-term interest. Encouraging, through law, ‘efficient’ breaches of these treaties would undermine the fundamental rule of \textit{pacta sunt servanda}, and likely render more difficult the possibility of sustained cooperation in an international community through treaty regimes” (cf. also Pauwelyn 2006).

\textsuperscript{488} Subsection 3.3.1 above gave economic reasons for a general default rule of inalienability in incomplete contracts.
Enforcement under inalienability is vital, even more so than under a property-rule protection. Any defection from the letter of the law is to be punished – with verve, if necessary. Inalienability theorists, accordingly, have not shied away from demanding high penalties for misdemeanors in the WTO-context.

We do not follow the thread of the inalienability approach to the WTO any further here. For one, the objective of this chapter is to broker between the compliance and rebalancing view. Furthermore, as we have argued in subsection 4.1.2.4.A above, we mistrust the commitment approach (on which the inalienability perspective is based) for conceptual reasons.

### 6.3.2 Why the rebalancing approach is misguided

The rebalancing approach to the WTO contract is subject to two serious and consequential flaws:

First, it runs the risk of confusing the actual *as-is state* of the WTO contract with an ideal *should-be state*. Rebalancing proponents should in fact be arguing that it is a systemic imperative to (re-)organize the WTO as a system of liability default. What they make believe is that the letter of the WTO treaty actually supported the rebalancing view. This is textually wrong. The rebalancing view thus forfeits logical traction, loses credibility with a critical audience, and diverts attention from its real strength, which is sound contract-theoretical analysis and systemic rigor.

As a matter of positive law, a liability-type pay-or-perform permission is absent in the wording of the DSU. Sykes (2000), and Schwartz and Sykes (2002a; 2002b) interpret various passages in the DSU (Arts. 3.7, 19.1, 21.1, 22.1, 22.8, 26.1.b) in a way that allegedly supports their liability-rule hypothesis. Each of the passages therein, they claim, can be read in a way consistent with a Holmesian liability system of entitlement protection. Their argumentation is creative, but ultimately far-fetched. There is hardly any doubt in the discipline of international trade law that the letter of the text states that WTO recommendations are binding, and that compliance with the obligations of the treaty is mandated for. An attentive reading of Arts. 21.1, 22.1 and 22.8 alone should dispel any doubts that specific performance is the compulsory fallback behavior.\(^{489}\)

---

\(^{489}\) Pertinent passages in the DSU wording read (emphases added): “Prompt *compliance* with recommendations or rulings of the DSB *is essential* in order to secure effective resolution of disputes” (Art. 21.1); “The suspension of concessions or other obligations shall be *temporary* and shall only be applied until such time as the measure
contrast to Art. XIX and XXVIII GATT which permit a liability-type escape (cf. Schwartz and Sykes 2002a, 2002b; Sykes 1991), the DSU does not contain a pay-or-perform-type escape. Escape policies may be negotiated between potential injurer and victim(s), but their enactment against the victim’s will is prohibited. Numerous panel and AB reports have interpreted the DSU to have exactly that compliance-inducing meaning.

Despite the fact that the mainstream interpretation of the DSU is one of strict legal bindingness of WTO rules and rulings, the DSU-wording that negotiators consented to might not reflect what they had intended to contract for. Another possibility is that the framers of the WTO may have negotiated according to their intentions, but have concluded a “bad” or “faulty” Agreement that fails to reflect what they would have opted for had they been more diligent, far-sighted, thoughtful, or under less time pressure. In both cases the DSU in its contemporary form would have to be called systemically flawed and in need of revision. However, this is a purely theoretical, and not a textual consideration. It is an issue which presupposes a theory of the system, its players and their preferences, negotiating obstacles, market imperfection, etc. Analytical and systemic reasoning is a strength of the rebalancing approach, since it draws analogies to L&E contract theory and to the economics literature. But this asset of possessing strong theoretical underpinnings is overshadowed by the camp’s untenable and ultimately far-fetched textual exegesis.

found to be inconsistent with a covered agreement has been removed, or the Member that must implement recommendations or rulings provides a solution to the nullification or impairment of benefits, or a mutually satisfactory solution is reached” (Art. 22.8); “Compensation and the suspension of concessions or other obligations are temporary measures available in the event that the recommendations and rulings are not implemented within a reasonable period of time. However, neither compensation nor the suspension of concessions or other obligations is preferred to full implementation of a recommendation to bring a measure into conformity with the covered agreements” (Art. 22.1). As for Art. 22.1 DSU, note that, counter to what Sykes (2000, p. 350) suggests, the word “preferred” relates to the terms “temporary measures” in the preceding sentence, and must not be read in isolation. Therefore Sykes’ (ibid.) contention that “the text [in Art. 22.1, sentence 2 DSU does not] expressly say that [compensation and retaliation] are illegal or that a Member who elects compensation or retaliation is in violation of the rules” must be seen as wrong.

490 Mavroidis (2000, p. 800) notes: “Article 22(1) DSU makes it plain that both compensation and suspension of concessions are not preferred options to full implementation. In other words, a WTO member should not be presumed to be in compliance with its international obligations when it continues an illegal act and at the same time either it agrees to pay compensation or concessions in its favour are suspended. In such a case a WTO member continues an illegality and has not fulfilled its international obligations.”

491 See Pauwelyn (2000) and Jackson (2004) for panel and AB reports in support of the international legal bindingness of WTO rules and rulings. Even assuming that Schwartz and Sykes were right in their reading of the DSU, one could easily argue that customary international law, or tacit consent among WTO Members, has developed over the course of the last ten years, reflecting dissent from the hypothetical treaty text, and hence legitimizing subsequent practice (Charnovitz 2001; Jackson 2004).
A second major flaw of the liability-rule perspective is its myopic view of the international world trading system. This may be provoked by an overzealous reliance on the results of formal contract theory. Rebalancing proponents adopt findings originating from standard contract theory in a reductionist and a-contextual manner. They tacitly insinuate assumptions which are neither implied by standard contract theory, nor maintained by contract theorists. The result of this over-confidence in formal contract theory is a misinterpretation and misapplication of academic findings.

Formal economic models of contractual relationships deal with extremely simplistic game settings, usually featuring two (but not multiple) contracting parties, one traded entitlement (but not multiple), and a complete set of actions (stationary environment), strategies and environmental contingencies.492 Reality is largely abstracted from, and the interaction is reduced to a theoretic bare-bone. As we discussed in subsection 3.1.6, whenever one is to adopt and integrate findings of formal contract theory literature, one must be very careful to adequately adapt them to the context at hand (here, the WTO). There is an inherent risk of losing explanatory scope, or, worse, generating wrong results by comparing the WTO to simple principal-agent or sales contracts.493 The WTO is much more complex and diverse a contract than anything formal contract theory usually deals with. Thus, results brought forth by contract theory have to be “customized” and reinterpreted to fit the WTO context. To that end, the contracting context has to be properly understood and sketched in the first place. Proponents of the rebalancing camp did not heed these principles.

A series of important consequences ensue that make the rebalancing view an untenable theory of the WTO contract:

1. Narrow scope of applicability: Exclusive focus on reciprocity-based rights and obligations. The rebalancing school of thought has its origins in economics. As was reviewed in section 6.1.1 above, for most economists the central objective of the WTO –

---

492 As laid out in chapters 2 and 3, contract theorists in industrial organization usually examine sales transactions (future exchange of goods for money). Alternatively, contract theorists deal with principal-agent relationships such as that between an employee (who performs only a single task or a well-defined set of simple tasks) and an employer. The contract then only consists of issues related to this task or set of tasks, and its remuneration. Other – more comprehensive – issues that we find in real employment contracts (multiple responsibilities, objectives, holidays, bonus packages, perquisites, unemployment and health aspects, training on the job, etc.) are neglected for the sake of formal tractability.

493 Jackson (2004, pp. 111, 119) and Dunoff and Trachtman (1999, p. 19) warn of overstraining analogies from domestic private law jurisprudence and of incorporating them into the international law context, since parties of interest, institutional settings and contextual circumstances are usually fundamentally different.
 Indeed its very raison d’être – is a mutual exchange of market access. To rebalancing proponents the substance of a trade agreement is the credible reciprocal exchange of market access concessions. All WTO rights and obligations are derived solely from this single substantive reciprocal commitment. Although we concur that the market access entitlement is the most important mutual obligation in the WTO (see remarks in subsection 4.2.4), there seems to be confusion between initial entitlements and rationale for the Agreement: Rebalancing advocates equate the two. Yet the WTO Agreement is evidently comprised of more rights and obligations than solely the trade entitlement.

2. How to rebalance non-trade entitlements? Non-reciprocal entitlements and their protection are of no apparent concern to rebalancing theorists. A natural corollary is that “re-balancing” as a response to partial defection of an injuring party breaks down for unreciprocated entitlements, such as minimum standard obligations or procedural rules of entitlement. It is hard to see how rebalancing can be achieved in situations where positive, non-reciprocated commitments are infringed upon (cf. e.g. Pauwelyn 2000, p. 342; Jackson 2004, pp. 121). As soon as a treaty obligation is violated that is not based on bilateral market access commitments, the rebalancing approach has no logical remedial reaction to fall back on.

3. Enforcement is not just another safeguard: No intellectual difference between contractual escape and dispute settlement. Chart 6.1 indicated another serious shortcoming of the rebalancing view: It grants no role to contractual enforcement. Liability-rule proponents seem to be guided by the following consideration: Under the conception of a contract, which is based solely on the exchange of market access, and is thus protected by a simple liability rule of default, virtually all instances of non-performance are dealt with in an intra-contractual way. The need to sanction unlawful behavior (violations) by means of countermeasures hence does not apply, since there

---

494 Trade concessions can take various forms: The market access entitlement comprises compulsory tariff bindings (Art. II GATT), positive GATS concessions in the four service modes, codes of conduct of how to deal with non-tariff barriers (SCM, TBT and SPS are examples), non-discrimination stipulations (e.g. Art. I and III GATT), the prohibition of quantitative restrictions (Art. XI), and others (cf. subsection 4.2.1).

495 Jackson (ibid.) notes: “[H]ow does one quantify a breach of a DSU norm? Rebalancing in any objective and meaningful way seems to be a fallacy in the light of the shifting perspective of the GATT/WTO system away from more quantifiable norms, such as tariff norms, toward broader rules that should arguably be shaped so as to provide benefits to all sides, not just a reciprocal ‘swap’”. See also our discussion around footnote 401.

496 Paradoxically, any “traditional” response to a non-market-access-related violation of WTO obligations (in the form of a countermeasure) exposes the liability-rule approach to a non-trivial puzzle: The victim’s countermeasures bring the initial balance of market access concessions into imbalance and should – within the strict liability logic – allow the original injurer to retaliate so as to bring the market access disequilibrium into balance again. This logic easily leads to a downward spiral of trade wars.
only exist lawful pay-or-perform actions. This reductionist thinking leads to the following consequences:

- The scope of application of contractual liability logically transcends the realm of the DSU: Using the violation-cum-retaliation strategy is just one of many (formal and informal) opt-out tools for ensuring contractual flexibility – albeit the most important one. For rebalancing proponents the DSU procedures constitute the contractual super-rule in case of any contractual gap.\footnote{According to the rebalancing camp, the DSU constitutes an implicit contractual rule that permits the breach of all other rules – at any time and under any circumstances. It is some sort of a super-rule which surpasses all other legal norms. This conception may have prompted Bello to contend that “WTO rules are not binding in traditional sense” (1996, p. 417).} The fact that contractual opt-out mechanisms transcend the narrow realm of DSB rules/procedures has been largely ignored by rebalancing advocates. (It has, however, been recognized by authors like Sykes 1991; Lawrence 2003; Roitinger 2004; see Schropp 2005).\footnote{An interesting side remark is that rebalancing backers cannot give meaning to the existence of non-violation claims under Art. 26 DSU: Questions of violations are out of the ambit of the simple rebalancing logic, therefore a distinction between violation claims and NVC is decrepit. Indeed: Under a simple liability-rule performance of WTO obligations is equivalent to ensuring market access. It is thereby irrelevant if non-performance occurs due to a violation or pursuant to a NVC.}

- Under the conception of a general liability-rule protection of entitlements, dispute settlement is therefore hardly about disputes, but about fostering contractual flexibility (or opt-out) and its consequences. A rebalancing regime is concerned with assessing and computing damages sustained by the victim of a measure in question. Hence the main role of the DSB panels and the AB is to be an arbitrator, information disseminator, conciliator, and monitor – not an adjudicator (Ethier 2001a; Rosendorff 2005; Rosendorff and Milner 2001; Thompson and Snidal 2005; see Keck and Schropp 2007 for an overview of the roles of the DSB in trade economics).

- The rebalancing approach ignores issues of extra-contractual behavior. The liability system of entitlement protection mandates the implementation of commensurate damages in reaction to non-performance. Yet what happens in the case of extra-contractual behavior? The simple system of indemnity breaks down in various circumstances: Firstly, as noted above, it defaults in situations where some non-reciprocal obligation is infringed upon. Secondly, the rebalancing perspective on the WTO is overstrained if an injuring Member acts in bad faith and does not abide by the rebalancing rules of the game. Imagine one Member using its “right” to opt-out to enact a protectionist measure. Yet assume that it neither offers voluntary damage
payments (tariff concessions) nor accepts the victim’s self-enforcement (tariff retaliation), and instead engages in counter-retaliation. Without any rules of enforcement in place, the system is unable to sanction, let alone deter, such bad-faith behavior. Faced with non-compliant conduct (in the rebalancing logic), the choice under a self-enforcing regime is then between doing nothing and seeing the system lapse into a trade war spiral of retaliation and counter-retaliation. Thirdly, the rebalancing argumentation collapses in a situation where a weak victim lacks the economic power to engage meaningfully in a retaliatory suspension of concessions.

- The rebalancing school of thought has a hard time dealing with contractual ambiguity or imprecise language. Situations of ambivalence or ambiguity – possibly giving rise to good-faith clashes – are notably absent in this economic thinking. They are absent not because they do not exist, but because they are assumed away from the outset (Masten 1999). In standard contract theoretical models this does not matter all that much, because the types of contract economic scholarship usually deals with are of an unambiguous nature. In more complex, long-term and repeated-interaction situations, such as the WTO, however, contractual ambiguity assumes a much more prominent role. In the same vein, standard contract theory models in traditional industrial organization lend little credence to courts. Courts are mechanical enforcers of law, conciliators, or calculators of damage (see our discussion in subsection 2.1.3 at footnote 76 above). The rebalancing approach to the WTO blindly adopts these results – stripping the DSB of any means to broker and adjudicate in good-faith clashes.

In sum, the rebalancing school of thought – incautiously or erroneously – equates contractual flexibility with enforcement rules and DSU procedures. When rebalancing advocates speak of dispute settlement and enforcement, they actually mean neither. They rather talk of arbitration/conciliation, and paying the contract-conform price of contractual escape, respectively. The truth of the matter is: Rebalancers do not discuss contractual enforcement problems at all. Yet in real life, a rebalancing system cannot operate properly without enforcement strategies and adequate mechanisms at hand. Unfortunately, proponents of the rebalancing paradigm have even “foreclosed” this avenue of true DSB enforcement by asserting that the violation-cum-retaliation was just

---

499 When facing a recalcitrant escapist Member, a weak victim’s hope of achieving the declared objective of rebalancing will be frustrated. The rebalancing mindset depends on a functioning self-enforcement regime. If a victim country does not have the economic (or political) capabilities to inflict retaliatory damage on the injurer, the concept of “rebalancing” falters without provision of fallback enforcement (see our discussion at footnote 443 and accompanying text).
another pay-or-perform opt-out mechanism. The same goes largely for dealing with contractual ambiguity: Since the DSB is interpreted as an opt-out facilitation device, value-creating gap-filling and adjudication by panels/AB so as to clarify and interpret ambiguous language is beyond the explanatory scope of the rebalancing approach to the WTO.

4. Confusing terminology. Rebalancing proponents utilize a baffling set of terminology. The terms “breach”, “violation”, “enforcement”, and “remedies” are apparently applied in a cooperative, intra-contractual fashion. This use of vocabulary, though occasionally championed in L&E literature on private and commercial contracts, is inappropriate for the context of public international law, where terms of this sort all have a strict extra-contractual connotation, and are reserved for punishable deviations from previously agreed rules and regulations. However, under the alleged general rule of liability, virtually all injuring behavior is covered within the confines of the contract. The wording used by liability advocates should thus be adapted to avoid disarray: The nomenclature “breach” or “violation” should be substituted by more neutral terms such as “non-performance”, “release”, or “excuse”. Likewise, “remedies”, “sanction”, or “countermeasures” should be properly termed “damage measure”, “indemnity”, or “price of non-performance”. “Enforcement”, finally, should read “implementation”.

5. Limited explanatory scope: Rebalancing is a theory about concession-escape and market access-related gaps. This point somewhat summarizes the previous shortcomings of the rebalancing approach to the WTO contract. The rebalancing stance is not a full-scale theory of international trade contracts. It is not a general theory about “breach”, breach and remedy in trade agreements. Rather, it is a theory about market access-related performance gaps and efficient gap-filling: A liability rule of entitlement protection works well in contexts featuring unforeseen market access contingencies, mutual exchange of commitments, and functioning (self-) enforcement. Only under these circumstances can contractual flexibility of the liability type enhance the general welfare of all participating contractors.

The rebalancing approach, however, falters whenever non-market access rights and obligations are at stake. Non-reciprocated legal WTO entitlements are outside the explanatory ambit of the rebalancing approach. In the same vein, it is mute on issues of extra-contractual behavior and ambiguous treaty language. Rebalancing proponents have answered to these shortcomings by disregarding non-market access-related eventualities,
just as they disregard instances of contractual ambiguity, or *extra*-contractual bad faith.\textsuperscript{500}

To conclude our discussion on the shortcomings of the rebalancing stance as an explanation for the WTO contract: A rather weak textual analysis of the positive law of the DSU raises serious doubts about the applicability of Art. 22 DSU as a generic liability tool of flexible contract adjustment. The credibility of the underlying theoretical considerations is gambled away by the rather bold contention that the WTO text *actually supports* this interpretation. In addition, the overzealous reliance on standard contract theory outcomes must raise doubts about the explanatory scope of the rebalancing mindset. As a result of these two major flaws, the rebalancing perception of the WTO contract reduces the multilateral trading system to a pared-down version of a mutual tariff agreement: When proponents of the rebalancing camp state that WTO *rules* are not binding in the traditional sense, they actually intend to say that *previously negotiated market access concessions* are not strictly binding, and are protected by an unconditional liability rule of flexibility.\textsuperscript{501} This is a somewhat different ballgame, and a consequential qualification in the scope of the applicability of the rebalancing perspective.

### 6.3.3 Why the compliance approach is misguided

The compliance school of thought, too, has significant weaknesses. The property-rule mindset is characterized (i) by too heavy textual reliance, and (ii) by the absence of a systemic vision and holistic perspective of the WTO contract.

First, the compliance school of thought relies too heavily on standard techniques of textual interpretation and legalistic analysis. In trying to reinforce their contentions that the WTO is about inducing compliance, proponents of the property perspective expend much effort on proving textual evidence about what today’s WTO *is*, instead of engaging

\textsuperscript{500} Rebalancing proponents might answer to these allegations by stating that they are exclusively concerned with market access contingencies. Thus, arguing about entitlements other than market-access related ones would be slightly futile, since these do not feature gaps. The argument might continue that non-market access entitlements are either contractually fixed in the WTO Agreements already, or secured by fallback rules of international law, viz. the VCLT. The quality of this argument notwithstanding, rebalancing proponents absolutely failed to make explicit and comprehensible these qualifications and the resulting limitation of explanatory scope. They are mainly responsible for the confusion created by the compliance/rebalancing rule debate.

\textsuperscript{501} To be sure: When proponents of the rivaling *compliance* perspective (property rule) speak about violation of WTO rules, they mean *all* the rules: They take into consideration all sorts of contractual deviation – be they related to market access, procedural, or obligations owed to the organization as a whole. Taking panel recommendations as a mere suggestion to comply, rather than an international legal obligation, essentially means nothing less to them than exerting impunity for all violations of WTO rules.
in a more analytical exercise of what the WTO was – and is – conceptualized as. To prop up their argument, its advocates engage heavily in textualism, endlessly citing treaty texts (GATT, DSU, WTO Charter, Preamble passages), preparatory notes, panel and AB reports, and statements by trade policymakers (see especially Charnovitz 2001, pp. 803; Jackson 2004, pp. 111).

The compliance perspective fails to acknowledge that the core of the matter is fundamentally theoretical in nature for at least three important reasons: For one thing, even Jackson (2004, pp. 112; 1997b, p. 62) concedes that textual analysis does not manage to conclusively “nail down” the matter of dispute between compliance and rebalancing proponents. The various WTO Agreements exhibit a good deal of textual imprecision. Textual ambiguity is arguably best dealt with by resorting to the systemic logic of the contract, and not just by going back to traditional legalistic treaty interpretation techniques, namely text, context, drafters’ intent, and practice under the Agreement. Second, since textualism in the WTO is obviously a double-edged sword, teleological interpretation techniques (“contextualism”) must be brought to bear. But contextualism in essence means going back to the object and purpose of a legal norm and a legal regime in general (Cohen 1999). Finally, in light of the severe discontent by WTO Member States, practitioners, and trade scholars alike, textual interpretation methods are of little help anyway: There is apparently something wrong with the WTO dispute settlement regime and WTO enforcement in general. Interpreting presumably faulty language must be seen as a futile task. Going back to the roots and asking what the WTO is conceptualized as seems much more fruitful an endeavor, and a necessary first step towards a workable institutional reform of the world trading system. The compliance approach largely ignores the route of contextualism.

---

502 The WTO treaty language is at times ambivalent and cloudy. Interpreting ambiguous language is necessarily colored by one’s concept on the WTO. Lawrence (2003, p. 13) submits: “By convention, major WTO decisions require a consensus, and since members often have differing views and interests, it is not at all surprising that the rules are characterized by what is sometimes called constructive ambiguity. The result is that agreements are subject to very different interpretations that reveal more about the perspectives of the interpreters than about the meaning of the text. This is particularly the case when those approaching the WTO have strong normative preconceptions that color their view of what the system should be.”

503 We use the term “contextualism” in a broader sense than applied by traditional lawyers. Generally, for L&E scholars (cf. e.g. Cohen 1999) contextualism means “searching for the deeper contractual intent” – an inherently theoretical and conceptual endeavor. Traditional lawyers, on the other hand, tend to look for the textual framework that a rule is embedded in. In the situation at hand, Jackson’s definition of contextualism is: “Analyzing the legal problem of obligation to comply therefore entails looking at the total framework of the DSU, to try to establish what the context is and the meaning of some of the specific clauses when viewed in that total context” (Jackson 2004, pp. 111). We have submitted supra that looking at the text of the DSU does not, however, resolve the issue of the nature of legal entitlements in the WTO.
There is a second flaw with the compliance perspective, the impact of which is amplified by its dogmatic adherence to the interpretative technique of textualism. The way property proponents argue must be criticized for isolating the issue of dispute settlement from the logic of the entire Agreement: The compliance view focuses exclusively on treaty enforcement, and grants little or no space to systemic issues of rule-making which logically transcend the narrow realm of the DSU. Nowhere in the debate do compliance theorists attempt to explain why sovereign States concluded the WTO contract in the first place, and what the nature of the exchanged entitlements is. Neither do they engage in a meaningful discussion on whether or not there is a deeper systemic logic in the WTO contract that would warrant additional trade policy flexibility. Nor do they challenge the alleged property rule of entitlement protection. If a property rule of renegotiation truly were the only means of contract adjustment, advocates should be able to come up with an explanation as to why WTO Members have opted for that rule of trade policy flexibility. So far, any systemic justification for the existence of compliance in the WTO is in fact wanting.\footnote{Arguably, a coherent theory of the system is what legal methodologies sometimes lack, as Dunoff and Trachtman (1999, p. 3) contend when uttering that “[l]nternational legal scholarship too often combines careful doctrinal description – here is what the law is – with unfounded prescription – here is what the law should be.}

These two failures bear some important consequences, which cast doubt on the significance of the compliance approach to the WTO contract:

1. **No distinction between trade entitlement and other entitlements.** As illustrated in Chart 6.1 the compliance view of the WTO contends that there is a strict specific performance duty for all WTO Members at all times. Unilateral opt-out for contractually unspecified reasons is generally impossible, least of all by using violation-cum-retaliation as an opt-out tool. It is consequent that property proponents only speak about – extra-contractual – “treaty violation”, “enforcement”, and “punishment”, since intra-contractual non-performance is not an option (save for explicit contingency measures mentioned in the Agreements).

However, the school of thought’s firmness on Members’ “obligation to perform” leaves a bulk of questions unanswered, including this one: “Performance of what obligation – and to what end?” Blind compliance with anything the WTO prescribes cannot be the purpose of enforcement. Contracts are concluded to achieve a specific objective, not for the purpose of complying with just anything. There must be a deeper contractual logic in
the form of substantive entitlements. But compliance advocates draw no systemic distinction between market access entitlements and other legal WTO entitlements. In their argumentation all entitlements are lumped together and put under the specific-performance umbrella. Compliance advocates will be hard-pressed to explain, why – conceptually – contractual escape is permissible under certain conditions and contingencies (e.g. GATT Arts. XIX, XX, XXI), but not under other – previously unspecified but potentially analogous – circumstances. The logic of contingencies in the mentioned WTO Articles exceeds the explanatory scope that compliance scholars can offer.

2. Compliance advocates accept the concept of flexibility by way of a “reasonable period of time”, yet deny flexibility via non-compliance. For obvious domestic political reasons (cf. Schwartz and Sykes 2002a, p. 15) the DSU grants the convicted injurers a reasonable period of time (RPT) to conform their policies, before sanctions or compensation may set in (Art. 22.6 DSU). It would seem that by tolerating the RPT, WTO framers explicitly acknowledged that at least for some time a violation coupled with the withdrawal of concessions is potentially superior to immediate compliance. As Schwartz and Sykes (ibid.) now contend, there is no systemic or theoretical reason to think that the same (or similar) factors, which make immediate compliance impractical, may at times make compliance feasible for an extended period. Just like granting an RPT is apparently mutually beneficial (otherwise it would not have been included into the contract), the joint interests of parties may be better served by letting parties opt out of the contract in return for offering compensation or enduring retaliation. Compliance proponents are mute on this issue.

3. An intra-contractual liability fallback rule? Compliance proponents fail to explain how a (hypothetical) institutionalization of a liability-type fallback rule into the WTO treaty would bear on their compliance dogma. A contract-conform liability rule of default would obviously form an integral part of the contract. That contract would then be protected by a property rule… This is certainly not what compliance advocates had in mind, when stating that compliance with the treaty is the central objective of WTO adjudication. However, apart from dogmatic adherence to a renegotiation rule, compliance advocates have little substantive argumentation to offer.

This scholarship often lacks any persuasively articulated connection between description and prescription, undermining the prescription.”
4. **Compliance proponents fail to explain the “toothlessness” of *de iure* sanctions.**
The compliance school does not offer a rationale for why the WTO framers did not go the extra mile to raise the bar for countermeasures to painfully punitive levels. Chapter 2 declared it a matter of systemic (Paretian) logic that negotiated contracts must be compatible with the incentive structure of all parties. If parties agree on a property rule of entitlement protection, they do so because a strict obligation to perform is in their mutual interest.

Compliance advocates, now, fail to explain, why DSU enforcement against illegal behavior is nevertheless so “toothless” in the WTO: Whereas contract theory would mandate that a property rule must be accompanied by prohibitive sanctions for illegal behavior, countermeasures in the WTO are limited to “substantial equivalence” to the level of harm done to the victim. Why did the framers of the WTO not choose stronger – incentive compatible – punishments when migrating from the GATT to the WTO regime of dispute settlement (cf. Schwartz and Sykes 2002b, section IV)?

The answers for property-rule proponents may lie in the weak benchmark default rules of public international law, as stipulated in the VCLT. But then the question remains as to why WTO framers did not “contract around” the fallback rules of international public law so as to install a more effective system of penal sanctions.

Alternatively, it could be argued that some limitation in the magnitude of sanctions is essential even under a rule of specific performance. However, this characterization again fails to explain why *equivalence* was chosen as the criterion for a ceiling on punishments if the only goal was to prevent trade battles from escalating. In addition, why did the system of the GATT prior to the creation of the WTO hold together very

---

505 Property-rule proponents like to argue that the world trading system changed from a pure rebalancing-oriented negotiation forum to a rule-of-law based organization, where compliance with the contracted rules is paramount. Note however that the punishment for violations has not changed from the “old” GATT to the “new” WTO system – it is still “equivalence” to the damage done (Pauwelyn 2000). “Remedies are there to ensure that contracts will under all circumstances be respected. Thus, remedies must be effective, and […] this is not the case in the WTO context” (Mavroidis 2000, p. 811). Compliance advocates fail to explicate convincingly just why the relatively weak language of Arts. 3.7 and 22.4 DSU has been transposed *verbatim* from old GATT documents, yielding evidently ineffective outcomes from a property-rule perspective. Schwartz and Sykes (2002a, p. 15) aptly state: “If WTO members really wanted to make compliance with dispute resolution findings mandatory, they would have imposed some greater penalty for noncompliance to induce it.”

506 As Pauwelyn (2006, pp. 20) asserts, in public international law (self-)enforcement has traditionally been limited to a simple tit-for-tat (see also Grané 2001).

507 By putting a cap on *extra*-contractual remedies, punitive overreach by WTO panels and the AB could be prevented, or an eruption of escalating trade wars – and therewith the risk of a system breakdown – avoided (Jackson 2004, p. 111).
successfully *despite* the absence of a meaningful constraint on the magnitude of sanctions?508

5. **Compliance proponents fail to explain the liability fallback of Art. XXVIII GATT.** Compliance advocates (in contrast to exponents of the inalienability perspective) do agree that the drafters of the WTO, cognizant of contractual incertitude, acknowledged that the treaty had to be modified from time to time. However, contrary to the rebalancing camp, they contend that all changes are to occur by way of mutual renegotiations – not by way of DSU proceedings. For a modification of the tariff concessions in goods, compliance proponents point to the procedures of Art. XXVIII of the GATT as the proper default rule of flexibility (and the analogous Art. XXI GATS for services, Jackson 2004, p. 121). Yet it must be stated that under this modification rule market access concessions are ultimately protected by a liability fallback rule – not by a specific performance rule (as shown above at footnote 419 and accompanying text).

In essence, the WTO presents compliance scholarship with a puzzle: The (alleged) compliance fallback of Art. XXVIII GATT in itself features a liability fallback, and the damage amount payable to the victim is mandated to be “substantially equivalent” *both* under Art. XXVIII.3 and under the DSU proceedings of 22.4 DSU. Hence, just as rebalancing proponents suggest, it is hard to see any qualitative, systemic difference between using Art. XXVIII’s opt-out opportunity and that of the DSU. Compliance advocates could have felt challenged by this contention and should have discussed this systemic puzzle. They did not. Compliance supporters cannot keep on pointing to the WTO framers’ unambiguous intentions in the DSU (and especially its Art. 22) without also explaining the (textually equally unambiguous) passages in Art. XXVIII GATT. The apparent conflict ought to be resumed by reverting to some sort of theory. Yet property advocates have so far given no answer to this issue.

In summary, the compliance view as to the nature of the WTO contract is generally too dogmatic in its textual interpretation of the WTO and too narrowly focused on issues of enforcement. That is probably why it loses sight of salient systemic questions of contracting, such as Members’ rationale for contracting, the nature of the traded

---

508 The consensus rule of dispute settlement under the old GATT 1947 allowed disputants to block the dispute resolution process from proceeding. The result of this positive consensus rule was that unilateral retaliation (such as resort to Section 301 of the US Trade Act of 1974) could be applied at will. Nothing prevented nations
entitlements, and the inherent logic in the system of trade policy flexibility. The compliance view is generally too much concerned with explaining what the DSU *is, says* and *does*, instead of looking for the bigger picture of the contract’s nature.

6.4 Lessons learned from the compliance/rebalancing controversy

Section 6.2 was devoted to overlooked issues in the decade-old compliance/rebalancing debate. On a positive note, the controversy can be seen as an important step towards structuring the eclectic and incoherent flood of proposals aimed at reforming the WTO enforcement regime: Scholars effectively realized that any fruitful discussion of reform suggestions must be vacuous in the absence of a clear perception of what dispute settlement and enforcement is to achieve in the first place.

On a negative note, we have contended that the debate ostensibly stopped at the question of the purpose and objective of WTO rule *enforcement*. WTO scholars did not broach the deeper issue of the nature of the WTO contract and the basics of rule-making. They missed a unique opportunity to engage in the superordinate discussion of entitlements exchanged in the WTO, and its system of trade policy flexibility and remedies. It should have been realized that the *nature of law contracted* is the prime input to discuss, and that the protection of WTO entitlements must be tackled before protection against contractual misdemeanor can be evaluated. In reality, the compliance/rebalancing rift easily extends the narrow scope of enforcement issues and dispute settlement. What is actually at issue is a much broader debate on trade policy flexibility and remedies in general.

We are witnessing a *discours des sourds* between the two rival camps. It seems that the two schools have lost touch with each other; no solution is in sight. We pointed out why the debate spiraled along two different trajectories. By re-calibrating the two opposite views to the issue of *rule-making*, we made explicit the largely implicit perceptions on the nature of the WTO contract according to the two perspectives. Rebalancing and compliance scholars describe two vastly different contracts: Rebalancing proponents portray a pure market access exchange contract, and theorize about *intra*-contractual safety valve mechanisms. Compliance advocates, negligent of the underlying

---

from using overzealous and punitive retaliation. Yet trade wars rarely emerged and multilateral liberalization made steady progress (cf. WTO 2007, subsection II.D.3.a; Sykes 2000, p. 352).
entitlements traded, introduce the WTO as a full-blown trade constitution, and contemplate the international legal bindingness of panel rulings and the enforcement of WTO rule violations in general.

Our analysis shows that neither of the two theories is flawless. The argumentation of each approach is inhomogeneous, and sometimes even erratic; neither compliance nor rebalancing paradigms lead to a consistent conception of the WTO contract.

Knowledge of the weaknesses of each approach provides some helpful lessons that we for our the analysis in Part 3 of this study: First, the WTO is a multi-entitlement contract; singling out one entitlement and its protection leads to flaws and misunderstandings. The rebalancing view of the WTO restricted itself to the study of the market access entitlement, and consequently inferred a wrong conclusion about the nature and objective of WTO enforcement. Second, in order to understand trade policy flexibility in the WTO, it is important to assess *intra*-contractual flexibility (secondary entitlements) before discussing tools of enforcement. Third, the *only* issue that compliance and rebalancing camps are *directly* at odds with is the question of adequate *protection of the market access entitlement*. Compliance scholars claim that efficient non-performance can be induced by means of a renegotiations clause, e.g. pursuant to Art. XXVIII GATT, while rebalancing proponents favor a liability-type opt-out along the lines of Art. XIX or, in the absence of any explicit contingency measure, of violation-cum-retaliation. Fourth, there are qualitative distinctions between the protection of the bilateral market access entitlement and non-reciprocated, multilateral entitlements. Fifth, in contrast to what rebalancing theorists may claim, as a matter of contractual logic enforcement mechanisms and escape mechanisms must not be substitutes. Treating illegal behavior just the same as cooperative action is not only morally dubious, but also vastly ineffective: Using violation of the Agreement as an *ex post* flexibility tool is irrational, since it strips signatories of any chance to sanction ill-meaning opportunistic behavior.

The dispute between compliance and rebalancing has not been resolved in a decade of debate. Can it be solved? Maybe there is much less to be solved than even the framers of this debate fathomed. Both perspectives on the WTO provide us with valuable conclusions about the nature of the contract: Whereas the compliance view delivers important realizations about the need for enforcement and entitlement protection in

---

509 It is imperative to note that the compliance and the rebalancing paradigm are not in opposition in other issues, just because they ignore them: The rebalancing school does not consider other entitlements, or enforcement of the contract. Because of this, it cannot be in disagreement with the compliance perspective on these issues.
general, the rebalancing view integrated the need for market access flexibility into the systemic whole. Since trade policy flexibility and enforcement is the concern of this study, we feel quite confident that we will be able to generate results that in many ways will lend themselves to reconciliation in the next section (see infra at section 7.6).
Part 3

Flexibility and enforcement in the WTO:
Towards an agenda for reform

Part 1 of this study gave an introduction into incomplete contract theory. It stressed the crucial role that contractual flexibility provisions, and default rules in particular, play in incomplete contracts. Part 2 presented a comprehensive contract-theoretical analysis of the WTO contract. There, we identified players, utilities and motives for contracting, and discussed the nature of contractual incompleteness prevalent in the treaty. This was followed by a portrait of basic entitlements exchanged in the WTO today and how they are protected from intra- and extra-contractual ex post non-performance. Pursuant to this systematic examination, we assessed in which respect the system of trade policy flexibility and enforcement is flawed in the current-day WTO. The consequences of these shortcomings for the international trading system in general were summarized.\footnote{510 The critical examination of the compliance/rebalancing debate, albeit a parenthesis of sorts, confirmed the aptness of our objective to examine the whole breadth of traded entitlements, not just parts of the contract.}

This final part of the study is geared towards an outline of reform. Chapter 7 will conduct the thought experiment of a hypothetical bargain analysis (cf. Scott 1990, p. 598). It theorizes about how reasonably rational, self-interested trade negotiators can be expected to organize and design a system of contractual non-performance in a multilateral trade agreement such as the WTO. Building on the previous analysis, we will search for the most efficient institutional rules for breach, “breach” and remedies trade negotiators will craft for themselves – cognizant of the idiosyncratic contracting context of the WTO. We want to elucidate which intra-contractual – legal – rules of trade policy flexibility best protect the various entitlements traded in the WTO, which enforcement provisions best safeguard compliance with these rules, and what cost injurers should incur for committing illegal behavior.

In short, Chapter 7 will hypothesize about how the WTO would look if it were organized along the precepts of the efficient “breach” contract (EBC). It will lay out a positive benchmark of trade policy flexibility and enforcement. Chapter 8 will conclude with detailed reform suggestions for a better and more viable system of flexibility and enforcement in the WTO.
7 Theorizing about the “WTO*” as an efficient “breach” contract

Seen from a contract-theoretical point of view, the current-day system of entitlement protection, trade policy flexibility, and enforcement in the WTO gives reason for concern: Currently, whenever an unforeseen (and hence unspecified) state of nature occurs, injuring WTO Members can choose between several substitutive – formal and informal – trade policy flexibility mechanisms, all of which are systematically under-compensatory. This results in (opportunistic) over-“breach” ex post, and under-commitment ex ante. The harm to the international trading system as a whole can be assumed to be significant. Economically weak players especially mistrust the contractual regime. This renders the global trading system dynamically unstable and under-performing.

Finding viable improvements and proposing sustainable reform avenues is imperative. Yet in order to make a meaningful contribution, it seems important that any proposal for reform is directed towards what is achievable, not just what seems desirable. Many reform proposals currently circulating in WTO scholarship are under a “prescriptive fallacy” (Dunoff and Trachtman 1999, p. 3): Authors base their suggestions on what they take to be right, fair, legal, distributively just, globally efficient or morally apposite. These are often intelligent, convincing, laudable and well-intentioned endeavors. However, they display one crucial error in their reasoning – they are not incentive compatible with what decision-makers desire. You cannot expect self-interested actors to agree to treaty details that are not conducive to their utility improvement, let alone reforms which might eventually harm them. Indeed, if those individuals who ultimately decide on reforms of the contractual design lose out from a given reform proposal, it stands little to no chance of ever being enacted. Thus, many prescriptions for WTO reform reside in a normative nirvana without hope of serious implementation.

To avoid this pitfall, let us take a step back and engage in a different consideration: “What would the right – read: incentive compatible – institutional design of flexibility and enforcement look like? What system of non-performance and enforcement can we expect reasonably rational trade negotiators to draft in the first place, given the trade-offs and constraints surrounding the initial negotiations?” Following the framework for examining incomplete contracts developed in subsection 3.4.1, this chapter will conduct
a hypothetical bargain analysis of the WTO system of non-performance. It constructs a hypothetical contract that notably uses the same contextual ingredients and inputs as the real WTO treaty – the same number of signatories, same entitlements, same types of incertitude, same timeline, same transaction costs, etc. Yet the hypothetical treaty will suggest a somewhat different governance structure. The outcome is a model contract of what the best achievable contract between self-interested policymakers could look like.

In absence of the unattainable first-best of the Pareto-efficient complete contingent contract, the contracting ideal is the efficient “breach” contract.\(^{511}\) The resulting EBC can be seen as an archetype of a multilateral trade agreement, and may therefore serve as a feasible benchmark institution against which to measure the reality of flexibility and enforcement in the contemporary WTO arrangement. This will help us bring forth a politically realistic, systemically viable and comprehensive reform agenda of the WTO system of non-performance in the final chapter.

In order to avoid confusion about what the WTO is today and what it is conceptualized as, we shall introduce the notation “WTO*” whenever talking about the idealized (hypothetical) version of the World Trade Organization.\(^{512}\)

Finding the optimal – incentive compatible – design of trade policy flexibility and enforcement in an incomplete multilateral trade agreement entails the following suite of questions:

1. What is the optimal protection of the market access entitlement, and of the various non-reciprocated multilateral entitlements? Will the Members allow for temporary deviation from all previously agreed commitments, or should some of the exchanged entitlements instead be protected by rules of inalienability? Is trade policy flexibility in the WTO best organized as a ready-to-use liability-type escape, or in the form of ex post renegotiations between injurer and victim?

2. How many different flexibility measures will safeguard the most efficient protection of an entitlement?

\(^{511}\) As was shown in section 3.4.2 above, the benchmark governance structure for an incomplete contract is given by the achievable first-best, the efficient “breach” contract. The EBC mimics the outcome of the hypothetical complete contingent contract by striking the optimal balance between flexibility and entitlement protection: Victims’ entitlements are protected in a way that prohibits opportunistic opt-out.

\(^{512}\) In subsection 6.3.2 we argued that the rebalancing view of the WTO must be reproached for blurring the lines between what the WTO currently is and what systemically would make sense. We wish to avoid this shortcoming by making very explicit that the hypothetical version of the WTO does not pretend to actually be the WTO in its contemporary form. The notation WTO* is thereby shorthand for “a multilateral trade agreement between reasonably rational, self-interested trade policymakers, such as the WTO”.
3. Should there be any strings attached to enacting a trade flexibility mechanism? In other words, what should be the optimal level of conditionality for making use of trade flexibility?

4. If liability-type flexibility is the negotiation equilibrium, what is the remedy benchmark, i.e. what is the *damage rule* of choice for *intra*-contractual non-performance? Should the victim be put in as good a position as if the injurer had performed? Should the *status quo ante* the breach be reestablished? Should the *status quo ante* the contract be restored? Or should the victim instead receive a fair share of the efficiency gains incurred by the non-performance of the injurer?

5. What is the *policy instrument* of choice for *intra*-contractual non-performance: tariff compensation or retaliation?

6. Finally, what kind of an *enforcement* regime should be in place to protect the contractual system of entitlement protection? Which *extra*-contractual remedies can safeguard adherence to the contractual rules – given the limitation of self-enforcement in international law?

Our analysis builds on the theoretical findings collected in Chapter 3 on incomplete contracting, and on Chapters 4 and 5, which characterized the nature of the WTO contract. Section 7.1 will give a complete overview of the trade game, and its inherent trade-offs and the constraints that potential WTO signatories are faced with at the beginning of contract negotiations. This is followed by a discussion of the efficient entitlement protection design of the reciprocal market access obligation (section 7.2) and of non-reciprocated multilateral entitlements (section 7.3). Section 7.4 assesses how reasonably rational policymakers can be assumed to efficiently organize the enforcement regime. Section 7.5 concludes by comparing the institutional design of the hypothetical EBC-version of the WTO (the WTO*) with the contemporary trade regime. Since our characterization of the hypothetical first-best contract between selfish policymakers may clash with what some pundits deem fair, morally just, or normatively desirable, we briefly assess whether the politically efficient archetype of the WTO* is also a “good” contract with respect to general welfare and the needs of non-signatory parties, such as consumers and producers. Our analysis paints a mildly positive picture of the WTO* in this respect: Since the hypothetical contract makes the world trading system more stable, more inclusive and promotes more participation and cooperation, the WTO* results in a globally more efficient contract. Various non-signatory parties will benefit from this improved contractual framework. Section 7.6 revisits and concludes the compliance/rebalancing debate discussed in Chapter 6.
7.1 The “trade game”

Before characterizing the hypothetical contract, let us give a complete overview of the “trade game” that reasonably rational trade policymakers are confronted with at the outset of their contract negotiations. Consider Chart 7.1: Similar to Chart 3.10 above, but adapted to the WTO context, it illustrates the trade-offs and constraints that trade negotiators have to take into consideration when sitting down to design the EBC (see also Chart 4.4 above).

Chart 7.1 Trade-offs and constraints in the WTO “trade game”

Source: author, based on Schropp (2007, Figure 1)

Notes: This chart represents (from the perspective of one signatory) a stylized version of the contracting game between countries that conclude a trade agreement based on market access externalities and common minimum standards. $t_1$ to $t_6$ mark different points in time. At $t_1$, the time of the conclusion of the agreement, reasonably rational policymakers draft the optimal governance structure for the incomplete contract. Signatories have to anticipate and bear in mind (i) the nature of the traded entitlements, (ii) the veil of ignorance, (iii) the incertitude inherent in the trade game, (iv) the transaction costs in the contract performance phase, and (v) the enforcement capacity constraints due to the absence of a central enforcer. These five constraints fundamentally determine signatories’ design of the most efficient governance structure. PR stands for “property rule”, LR stands for “liability rule”, exp. dam. for “expectation damages”, TC for “transaction costs”.

Chart 7.1 represents a stylized version of the “trade game” that unfolds after the contracting parties agree to conclude a trade agreement – the WTO*. In stage 1 ($t_1$), various policymakers representing sovereign States agree to overcome inefficient market access externalities, and to reap transaction cost efficiencies by introducing minimum
standard levels. To that end, signatories draft a governance structure: They determine substantive and basic auxiliary entitlements (primary rules), as well as entitlement protection mechanisms (secondary rules) and enforcement provisions (tertiary rules).

In stage 2, depending on the quality of this institutional framework, WTO* Members negotiate the level of mutual commitments, i.e. the depth and breadth of market access concessions, as well as the minimum level of positive integration commitments.\textsuperscript{513} The most reluctant liberalizer thereby sets the extent of trade liberalization commitments.

After the level of cooperation is fixed, the repeated-interaction performance phase of the contract begins at stage 3. Countries start interacting according to the terms of the contract (infinite repetition). At some point in time ($t_3$), incertitude in the environment is resolved: An exogenous or endogenous market access- or non-trade contingency hits and may create room for regret in one or more Members.

In stage 4, depending on the terms of the initial agreement, an affected injurer may be allowed to react to the contingency. If a rule of inalienability is in place, or if the shock was too insignificant to warrant expected compensation payments, the injurer will perform as promised. Whenever the contract provides for an LR of \textit{ex post} discretion, the injurer may default from previously agreed levels of performance at his own initiative. This triggers \textit{intra}-contractual remedies in stage 5. Alternatively, the injuring Member initiates a lengthy renegotiation phase with victims (which is resolved any time between $t_3$ and $t_5$), aimed at buying off the victim’s right to demand performance as prescribed.

A victim country that claims to have detected a nullification or impairment of its WTO* rights can initiate the dispute settlement and enforcement mechanisms in stage 6. An injuring Member refusing to pay its remedies as previously agreed to, or defecting from an entitlement protected by an inalienability rule, also triggers a dispute initiation on part of the victim country.

When drafting their contractual governance structure at $t_1$, the signatories of the multilateral trade agreement must be conscious of, and take into serious consideration, the trade-offs and constraints that unfold during this trade game. The following impediments to contracting eventually determine the form and design of the optimal governance structure (cf. numbering in Chart 7.1):

\textsuperscript{513} For expositional convenience, stages 1 and 2 are separated here (cf. footnote 348 in section 4.3).
(i) Nature of entitlements. The trade negotiators must have a common understanding of the nature of the exchanged entitlements (see section 4.2). The market access entitlement is reciprocally owed between country dyads. Hence, the primary promise between every pair of signatories is to have a balanced exchange of market access in place. The common trade liberalization level (depth and breadth of commitments) leading to the balance is secondary. The market access entitlement encompasses a wide range of industries with numerous tariff lines, or services commitments. It also includes a number of auxiliary entitlements geared towards safeguarding that initial balance.

The numerous multilateral entitlements – minimum standard rules, procedural rules, external rules, rules of transparency and obligations owed to the institution – are not reciprocal in nature: There is no balance to speak of. Instead, non-trade entitlements are owed erga omnes partes – to the entire membership. Many multilateral non-trade entitlements are dichotomous (binary) in nature, that is, partial performance is not possible.

(ii) Veil of ignorance. The WTO contract is concluded behind a Rawlsian “veil of ignorance” (as Ethier 2002; Ethier 2001a; Sykes 1991; and Lawrence 2003 concur). Reasonably rational parties negotiate the WTO* as a long-term commitment. They have no full information of their economic significance in the distant future, of the composition of their comparative advantage, of their role as injurers or victims, of the identity of acceding countries, and generally of how future states of the world will impact their well-being.514

The fact that contracting governments are negotiating trade agreements behind a veil of ignorance is consequential. First, it rules out problems connected with adverse selection (hidden knowledge): If signatories are unsure about their future role, they are not likely to strategically misrepresent it. Second, as Ethier (2001a) and Sykes (1991) point out, ex ante symmetrically uninformed, reasonably rational transactors can be considered to take the choices of a social planner whose aim is to maximize the future well-being of all

---

514 Readers might have doubts as to how realistic the concept of a “veil of ignorance” really is in trade agreements. Yet consider the following illustration: The GATT as one of the fundamental pillars of today’s WTO system was founded in April of 1947. Among its original signatories were Burma, Ceylon, the Republic of Cuba, the Czechoslovak Republic, Lebanon, Southern Rhodesia, and Syria. When signing the Agreement none of the signatories could foresee each others’ economic role 60 years down the road. In fact, some of the above countries don’t even exist any more today (Rhodesia, Czechoslovakia), or took quite different economic trajectories due to political upheavals over the last six decades (Burma, Cuba, Syria, Lebanon). According to economic dogma, completely rational policymakers would have had to take into consideration developments of this sort, and discount them accordingly.
contracting parties: In absence of any information to the contrary, parties anticipate that they will be hit by exogenous and endogenous shocks just as often as any other country. This leads them to craft agreements, and commit to trade liberalization, just as a social planner would do. However, note that the maximand is not general (consumer) welfare, but the welfare of all contracting parties, viz. self-interested policymakers.

(iii) Nature of contingencies and incompleteness. In subsection 4.3.1 above we reviewed the nature of market access contingencies, which we characterized as political support shocks with spillover potential. Similarly, non-trade contingencies were also shown to be political support shocks which prompt signatories to react by means of (partial) defection from initially agreed obligations (cf. subsection 4.3.2). The type of incertitude that besets market access and multilateral entitlements was demonstrated to be such that previously unforeseen political contingencies are revealed asymmetrically to injurers. As shown in subsection 4.3.1.2 and 4.3.2, this pre-destines the WTO* contract to be incomplete: Both kinds of entitlements either display necessary incompleteness or inexorable-type B incompleteness (depending on whether the probability density function of contingencies is known to signatories or not; see Chart 4.5).

(iv) Post-contractual transaction costs – costs of renegotiation and quantification. When deciding on the regime of ex post discretion, reasonably rational signatories must take into consideration the exogenous and endogenous transaction costs that unfold during the trade game. As subsection 3.3.2.2 above showed, the choice between a liability rule or a property rule of default is crucially determined by an inherent trade-off between renegotiation costs (pursuant to a PR) and damage calculation- or quantification costs (pursuant to an LR). The entitlement-specific ex post TC in the WTO will be discussed infra.

(v) Enforcement constraints. The WTO* is a self-enforcement regime, as is the WTO. In the absence of an external enforcer, WTO* Members are restricted to their own enforcement capacity. However, enforcement does not necessarily have to be purely bilateral: Just as individual citizens of a country overcome collective action problems, join forces, and institute a system of collective policing and enforcement (by dispensing resources to establish bureaucracies, police forces and other executive bodies), WTO*

---

515 In a situation with two signatories (where one party is “Home” and the other is “Rest of the World”), transactors will assume the roles of victim and injurer of a protectionist backtracking measure with roughly 50 per cent of the time.
Members are free to bundle their enforcement capacities into certain means of collective enforcement.

7.2 Organizing protection of the market access entitlement

When self-interested policymakers negotiate a reciprocal exchange of market access behind a veil of ignorance, they are motivated by two objectives: To encapsulate the current domestic political support constellation (which determines their favorite – politically optimal – level of trade liberalization), and the drive to design a sustainable institutional system that is fit do deal with an uncertain future. When designing the system of trade policy flexibility and market access entitlement protection, forward-looking policymakers behave like social planners whose job it is to maximize the welfare of all present and future self-interested trade policymakers. The analysis will show that policymakers opt for protecting the market access entitlement with an unconditional liability rule of default.

7.2.1 Focusing on default rules

Exposed to necessary or inexorable-type B incompleteness, forward-looking policymakers are conscious of the fact that it is impossible to anticipate every future contingency, even if they invested infinite time and resources into writing the contract (see our discussion in subsection 3.1.3, point 6). Therefore, reasonably rational decision-makers do not bother trying: Contingencies are private knowledge to the affected party, and cannot possibly be directly anticipated. Negative integration provisions that prohibit certain instruments can easily be replaced by ill-meaning Members with other, previously non-specified instruments – with the same protectionist result. In addition, writing down contingencies and instruments is costly, prone to erratic and ambivalent language, and an additional source of opportunism and disputes.

Rational policymakers will thus prefer to leave the market access entitlement completely non-contingent. They content themselves with living in full contractual uncertainty of the future, and with focusing on the efficient design of ex post discretion rules. Thus, contractual clauses relating to the entitlement to reciprocal trade consist of (i) concession schedules, (ii) rules circumscribing and detailing the market access entitlement (definitions and instructions, non-discrimination titles, rules on non-tariff barriers, exceptions, etc.) and (iii) a single rule of default. The default rule states that whenever a
market access-related disturbance occurs, the fallback rule comes into effect. In short, the exchange of market access at the WTO* is organized as an optimally indefinite sub-contract. This finding is consequential. It allows us to resort to some formal work conducted on the WTO.516

7.2.2 Inalienability or ex post discretion?

When assessing the optimal entitlement protection for the reciprocal market access promise, the first question that needs to be tackled by trade policymakers is whether ex post discretion should be allowed at all, or whether a rule of inalienability is in fact more apt. We argue in favor of discretion in two steps: First, by reviewing the literature that establishes a link between trade policy flexibility and the stability of the world trading system in a repeated-interaction context (subsection 7.2.2.1). Second, taking the risk of breakdown of the multilateral trading order as less of a concern, subsection 7.2.2.2 assesses the incentive compatibility of a rule of inalienability in a one-shot game.

7.2.2.1 If stability of the system is an issue517

A completely non-contingent market access deal negotiated behind the veil of ignorance can be conveniently represented by an infinitely repeated tariff-setting game between two symmetrical, equally uninformed players. Self-interested policymakers from two symmetrical countries (which only differ in their factor endowments) sit down to negotiate a tariff-liberalization agreement to overcome mutual market access externalities. Uncertainty is introduced in the form of a completely unexpected exogenous political support shock ($\theta$) of a finite size. The shock by definition is unwelcome, i.e. unleashes a desire for a protectionist reaction in the affected party.518 $\theta$ is a random i.i.d. shock of a commonly known probability density function $p(\theta)$.519

---

516 For reasons of scientific economy, scholars often model the WTO as a tariff-liberalization agreement where future contingencies are unknown. Contingencies are formalized as protectionist shocks of a commonly known probability density function. In other words, the WTO is modeled as an optimally indefinite market-access contract that is beset by what we defined in subsection 3.1.3 as Type A or efficient, or as Type B or necessary incompleteness (depending on the underlying assumptions of how previously unforeseen contingencies are revealed to players).

517 This section is a synthesis of the two outstanding works of Rosendorff (2005) and Herzing (2005, chapter 3). We give an account of the general intuition of these contributions, and leave the details and intricacies of the models to the interested reader.

518 Another way of seeing it is to assume that the political shock always harms import-competing industries who then exercise their political clout with the self-interested policymaker.

519 i.i.d. stands for “independent identically distributed random variables”. In probability theory, a sequence of random variables is i.i.d. if each has the same probability distribution as the others and all are mutually independent.
Shocks are idiosyncratic to each player, and are temporary in nature: They only affect the signatory for a single period.

The tariff-setting game, in which every signatory sets its tariff simultaneously and independently at the beginning of each period, can now be represented as a prisoners’ dilemma (PD) between these two parties (Home and Foreign, the latter represented by *) with the following payoffs:

<table>
<thead>
<tr>
<th></th>
<th>( C^* )</th>
<th>( D^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C )</td>
<td>( W_C(t_C^<em>, t_C^</em>; \theta) ), ( W_{\ast C}^<em>(t_C^</em>, t_C^<em>; \theta^</em>) )</td>
<td>( W_{\ast D}^<em>(t_C^</em>, t_D^<em>; \theta^</em>) )</td>
</tr>
<tr>
<td>( D )</td>
<td>( W_{D}^<em>(t_D^</em>, t_D^<em>; \theta) ), ( W_{\ast D}^</em>(t_D^<em>, t_D^</em>; \theta^*) )</td>
<td>( W_{N}^<em>(t_N^</em>, t_N^<em>; \theta^</em>) )</td>
</tr>
</tbody>
</table>

Thereby, \( t_C^* \) is the previously agreed Cooperative tariff level; \( t_D^* \) is the Defective tariff, namely the best-reaction function to the other player’s cooperative payoff \( t_{BR}(t_C^*) \), and \( t_N^* = t_{BR}^2(t_{BR}) \) is the non-cooperative Nash-tariff, where both countries apply the optimal tariff against each other;

\( W_C(t_C^*, t_C^*; \theta) \) is each player’s per-period payoff in the Cooperative case;

\( W_D^*(t_{BR}(t_C^*), t_C^*; \theta) \) is the payoff for the Defector;

\( W_S^*(t_C^*, t_{BR}^2(t_C^*); \theta) \) is the “Sucker’s payoff” if the other party defaults;

\( W_{N}^*(t_{BR}(t_{BR}), t_{BR}(t_{BR}); \theta) \) is the non-cooperative Nash payoff that exists in the absence of any trading agreement.

The general payoff structure of prisoners’ dilemmas \( W_D > W_C > W_N > W_S \) holds. This infinitely repeated trade-setting game breaks down in a finite time scale, as shown in Chart 7.2 below:

---

520 The sequence of the per-period trade-setting game is as follows: (i) At the beginning of a period, both countries (Home and Foreign*) experience independent political support shocks that are unobservable to the trading partner. (ii) Both countries determine independently what trade policy they will apply, the options either being setting the cooperative tariff or deviation, i.e. applying the optimally defective tariff. (iii) Both players implement their policies, and the period begins. (iv) At the end of the period, the implemented policies are verified. Any deviation by one country is regarded as a breach of the agreement and will therefore lead to a breakdown of the cooperative regime.

521 Note that in a stage-game with symmetrical and equally uninformed players, the expected per-period payoffs are identical. Thus the (*) representing the foreign country can be dropped, which reduces the examination to one single signatory; the other Member’s payoffs are equivalent.
Chart 7.2 The breakdown condition for the simple tariff-setting game

\[ \theta^{H&R} \]

Source: author based on Rosendorff (2005, Figure 1)

Notes: This graph is similar to Chart 3.6, panel a, above. It shows the injurer’s trade-off in the face of some exogenous shock (\( \theta \)): The injurer can decide between defecting once and forgoing cooperation payoffs in all future periods, and continued cooperation. The vertical axis plots the disutility from cooperation foregone, and the utility from defection. \( H&R \) stands for hit-and-run, the difference in payoff between \( W^D(\theta) \) and \( W^C(\theta) \). \( CV^{PD} \) stands for the continuation value of the repeated PD game. The chart shows the cut-off shock level (\( \theta^{H&R} \)) above which the injurer prefers to defect (\( D \)) instead of cooperating (\( C \)).

Chart 7.2, which we came across before in Chart 3.6 panel a, plots the shock level of the unanticipated state of nature on the horizontal axis, and the injurer’s (dis-)utility from defecting as the dependant variable. The well-known \( H&R \)-curve represents the value of one-time defection (\( W^D_\theta – W^C_\theta \)), whereas the line \( CV^{PD} \) is the continuation value of trade cooperation, which the injurer foregoes by defecting.\(^{522}\) Thus, it can be see that the prisoners’ dilemma stage-game is only sustainable for low realizations of \( \theta \), high discount factors, and for low levels of trade liberalization (which flatten the \( H&R \)-curve). However, whenever the exogenous shock exceeds the threshold level \( \theta^{H&R} \), the contract breaks down. With \( i.i.d. \) shocks, this happens in finite time.\(^{523}\)

\(^{522}\) As shown in Chapter 2 above, the continuation value is the net present value of the per-period cooperation benefits over the reversion to non-cooperative Nash play. The line is flat in \( \theta \) because by assumption future payoffs are independent of the current shock level.

\(^{523}\) The breakdown condition in each period is \( W^D_\theta – W^C_\theta > \delta / (1 – \delta) [W^C – W^D] \), where \( W^D_\theta – W^C_\theta \) is the hit-and-run payoff in the current period (dependent on the shock), and \( \delta / (1 – \delta) [W^C – W^D] \) is the continuation value of the game. \( \delta \) thereby depicts the discount factor.
If the market access contingency ($\theta$) were to be symmetrically revealed to both parties, the game would have a trivial outcome: No party would have the chance to misrepresent the true state of nature. Consequently, the shock-affected party would be allowed to refrain from full contractual performance for exactly one period, given that it pays commensurate damages to the victims of the resulting protectionist measure (Bagwell and Staiger 1990; Ethier 2001a). This simple tit-for-tat result would replicate the Pareto-efficient complete contingent contract.524

Yet, whenever shock contingencies are asymmetrically revealed, this offers opportunities for injurers to strategically misrepresent the real size of the shock, and to engage in excessive defection to the detriment of the victim. In order to forestall the possibility of any such opportunistic behavior, the negotiating parties may decide *ex ante* on an inalienability rule of entitlement protection: They might be of the opinion that the losses to victims generated from excessive “breach” outweigh the danger of breakdown of the system. The WTO literature has firmly established that this is not a good idea, and that the introduction of escape clauses Pareto-dominates a rule of mandatory specific performance.

Rosendorff (2005) shows that the inclusion of an escape clause strategy in the action space of the players helps obtain Pareto-superior long-term cooperation in the shadow of the grim trigger. The author extends the above PD game for a third possible action next to cooperation and defection: Each party can escape its commitments by partially withdrawing from its obligations in the current period, given that it grants compensation to the victim.525 Rosendorff assumes that these damage payments are *exogenously* awarded by a dispute settlement panel, and are *proportional* (but not commensurate) to the harm caused to the victim. The resulting payoff matrix of this extended prisoners’ dilemma game is:

---

524 Note that symmetrical revelation of contingencies replicates Axelrod’s classic result that *tit-for-tat* is the most viable retaliation strategy in an infinitely repeated prisoners’ dilemma (Axelrod 1984; Axelrod and Keohane 1986; Oye 1986).

525 Whether this escape clause is organized as a liability-type opt-out, or a directive to engage in tariff renegotiations (property rule), is inconsequential here.
Thereby, \( D \) (\( D^* \)) represents the damage done to the victim, and \( \pi D(\theta^*) \) and \( \pi D^*(\theta) \) represent the damage award to the victim of a protectionist measure.

Rosendorff (ibid., Proposition 1 at pp. 393) proves mathematically that a pair of “structured defection” strategies (shaded cells) is a Nash equilibrium that Pareto-dominates a strategy without temporary escape: A country defects if its partner defected in any period in the past, and otherwise \( C \) (cooperates (if its preference shock is mild), temporarily \( E \) (caps (if the shock is strong), or \( D \) (defects (if the shock is enormous). In other words, a policymaker who can decide in each period whether to cooperate as promised, to escape the obligation once in return for compensation payments or to defect, will act in a way that is \textit{ex ante} welfare-superior to the 2x2 PD game shown above.

To see why this is so, consider Chart 7.3. Starting from the same initial set-up as the previous chart, Chart 7.3 shows that opening up the opportunity to engage in structured one-time defection can often save the trade-setting game from breaking down. Each country affected by a protectionist shock can balance off its incentive to seize the hit-and-run advantage (thereby exiting the system) against escaping its obligations in the present period only and paying proportional indemnity to the victim. For light shocks, where the (expected) compensation payment is larger than the hit-and-run advantage, the injurer cooperates as promised (area \( C \) in Chart 7.2). At a threshold level (\( \theta^EC \)), escape-cum-compensation yields higher payoffs for the injurer (the area \( EC \)). Whenever the current shock is larger that a cut-off level (\( \theta^EC \)), the opportunistic hit-and-run advantage outweighs the continuation value of cooperating. The injurer exits the trade game.

\[\begin{array}{|c|c|c|}
\hline
 & C^* & EC^* & D^* \\
\hline
C & W^C(f^C, f^C; \theta) , & W^C(f^C, f^C; \theta^*) + \pi D(\theta^*) , & W^S(f^C, f^C; \theta) , \\
 & W^S(f^C, f^C; \theta^*) + \pi D(\theta^*) & W^S(f^C, f^C; \theta) , & W^S(f^C, f^C; \theta^*) + \pi D(\theta^*) \\
\hline
EC & W^S(f^D, f^D; \theta) , & W^S(f^D, f^D; \theta^*) + \pi D^*(\theta^*) , & W^S(f^D, f^D; \theta) , \\
 & W^S(f^D, f^D; \theta^*) + \pi D^*(\theta^*) & W^S(f^D, f^D; \theta) , & W^S(f^D, f^D; \theta^*) + \pi D^*(\theta^*) \\
\hline
D & W^S(f^D, f^D; \theta) , & W^S(f^D, f^D; \theta^*) + \pi D(\theta^*) , & W^S(f^D, f^D; \theta) , \\
 & W^S(f^D, f^D; \theta^*) + \pi D(\theta^*) & W^S(f^D, f^D; \theta) , & W^S(f^D, f^D; \theta^*) + \pi D^*(\theta^*) \\
\hline
\end{array}\]

\[\begin{array}{|c|c|c|}
\hline
\end{array}\]

526 The damage caused to the victim is indirectly dependent on the injurer’s experienced shock, because that contingency shock determines the size of the optimal defection tariff of the injuring country.
Chart 7.3 Stability and breakdown in the escape-clause game after Rosendorff (2005)

Source: author based on Rosendorff (2005, Figure 3)
Notes: This graph pictures the same initial situation as Chart 7.2 above. This time, however, the injurer has the opportunity to choose an escape action, and to compensate the victim for that one-time defection. The \( \pi D(\theta) \)-curve represents an exogenously given compensation function that is proportional to the harm caused to the victim. The trade-off that the injurer faces is thus between cooperating (in situations where the protectionist shock, and therewith the hit-and-run gain, is mild), choose escape (EC) and pay damages (in situations where the proportional compensation is less than future payoffs) or Defect and exit the contract (where the experienced shock is enormous). Since escape behavior reduces the expected per-period cooperation, the continuation value (\( CV^{PD} \)) is assumed to decrease for every future period compared to the simple PD game of Chart 7.2.

Comparing the outcomes of Charts 7.2 and 7.3, two things can immediately be noticed: First, the game of structured defection is more robust against breakdown and exit (ibid., Proposition 2 at p. 395): While under inalienability the injurer remains cooperative in the presence of shocks up to (\( \theta^{H&R} \)) and exits thereafter, injurers staffed with the option to escape only exit pursuant to higher shocks (\( \theta^{EC} \)). The trade game staffed with an escape clause is more stable.

Second, the continuation value of the game is reduced, because real cooperation (i.e. contractual exchange that reaps the transaction efficiencies for which the WTO* was concluded in the first place) happens less often (\( \theta^{EC} \) instead of \( \theta^{H&R} \)). This brings down the expected per-period payoffs and consequently the value of the entire game.

In addition to those directly visible results, Rosendorff proves formally that adding an escape-cum-damage option has additional advantages over a rule of inalienability: A wider variety of countries will join the agreement. Whereas in conventional PD games
only the most patient players (those with relatively high discount rates) will sign on to the self-enforcing agreement, an opt-out clause effectively lowers the threshold level of discount factors necessary to sustain a cooperative outcome. Thus, a wider variety of countries with lower discount factors can enter the agreement \((\text{ibid.}, \text{corollary 1 at p. 395})\).

Furthermore, initial agreement will be easier to strike \((\text{ibid.}, \text{p. 389, and Rosendorff and Milner 2001 at pp. 850})\). Fearon (1998) warns that the longer the shadow of the future stretches in infinitely repeated games, the fiercer the initial bargaining is expected to be. The \textit{ex ante} negotiated terms of the agreement “lock in” the uncertain future distributional gains and losses that occur in the course of the repeated game. This in turn gives rise to strategic hold-out behavior by negotiating parties.\(^{527}\) The inclusion of an escape mechanism, now, significantly reduces this lock-in effect, and makes initial bargaining less fierce and thus initial agreement easier to strike.

Herzing’s contribution (2005, chapter 3) is in many ways an improvement on Rosendorff’s article. Most notably, Herzing overcomes the two main weaknesses encumbering Rosendorff’s paper, namely the \textit{assumptions} that compensation be exogenously awarded and proportional to the damage, and that tariff commitments be fixed.\(^{528}\) Herzing endogenizes both decisions, and so provides for a richer picture of the tariff-setting PD between two countries behind the veil of ignorance.\(^{529}\)

Herzing uses the same \(3 \times 3\) matrix of the extended prisoners’ dilemma but has the two symmetrical parties \textit{bargain} over the right size of compensation and the \textit{ex ante} mutual trade liberalization. In other words, he replaces \(\pi D(\theta^*)\) and \(\pi D^*(\theta)\), respectively, for a reciprocally negotiated compensation \(ED(\theta)\). Chart 7.4 illustrates Herzing’s key findings:

---

\(^{527}\) See the discussion on \textit{hold-outs} in footnote 118 (subsection 3.1.1.1), and in footnote 228 and accompanying text (subsection 3.3.2.2).

\(^{528}\) The result of those two exogenous assumptions is that Rosendorff is only concerned with the role of the injurer while neglecting any considerations on the part of potential victims.

\(^{529}\) Herzing’s improvements, however, come at the price of added complexity and incomprehensibility for non-specialists.
Chart 7.4 Stability in the escape-clause game after Herzing (2005)

Source: author, based on Herzing (2005, chapter 3).

Notes: This graph pictures the identical initial situation of Charts 7.2 and 7.3 above, but illustrates Herzing’s (2005) findings. Behind the veil of ignorance, signatories negotiate a trade agreement that never breaks down. This implies a compensation schedule (shaded area) which either compensates the victim country for its expectation damages or amounts to the injurer’s entire continuation value – whichever constraint is binding. The \( ED(\theta) \)-curve represents the victim’s expectation damages (that indirectly depend on the injurer’s shock). Compared with a rule of inalienability, the signatories commit to higher trade liberalization levels under the optimal compensation schedule. This skews the \( H&R \)-curve to the left, and increases the continuation value of the game (upwards shift of the \( CV^{ED} \)-line). In general, more pure cooperative behavior occurs under an expectation damage-backed escape clause (marked by a rightwards shift from \( \theta^{H&R} \) to \( \theta^{ED} \)).

In his highly sophisticated model, Herzing confirms all of Rosendorff’s findings: The enactment of an escape clause regime leads to more stability, less exit, more inclusiveness and less ex ante bargaining. In addition, Herzing puts forth the following results: First, reasonably rational parties may design a framework under which the contract never breaks down (\textit{ibid.}, Proposition 1 at p. 86). This set-up requires the introduction of an intricate compensation scheme triggered by the use of an escape clause. The compensation schedule under which cooperative behavior is always sustained either condemns the injurer to pay the victim’s expectation damages, or to pay a sum equal to the injurer’s entire continuation value of the game (whichever constraint
7 The WTO* as an efficient “breach” contract

is binding; *ibid.*, Proposition 3 at p. 89). In Chart 7.4, this compensation schedule is represented by the shaded area.

Second, the inclusion of an escape clause backed by the optimal compensation scheme facilitates more far-reaching tariff liberalization than under a rule of inalienability (*ibid.*, Proposition 6 at p. 94). In other words, the introduction of trade policy flexibility enhances *ex ante* cooperation. Third, a well-designed escape clause makes pure cooperation more likely (instead of less likely, as Rosendorff claims): Since the injuring country is forced to internalize the externalities it creates, it is hesitant to enact the escape clause too often (*ibid.*, Proposition 5 at p. 93). Thus, pure cooperation is chosen by injurers for higher shocks than under the no-escape regime (see Chart 7.4 where $\theta^{ED}$ lies to the right of $\theta^{H&R}$).

Fourth, an intuitive implication of deeper *ex ante* trade liberalization concessions and more pure cooperation in the face of protectionist shocks is that the expected per-period payoff under an escape clause regime is strictly larger than under a rule of inalienability. $CV^{ED}$ is located above $CV^{PD}$, which indicates a higher continuation value for the trade game with expectation damages (*ibid.*, Proposition 8, p. 96 and Lemma 8, p. 97).

This last finding stands in contrast to the contentions of influential IR scholars (e.g. Downs and Rocke 1995; Rosendorff 2005; Rosendorff and Milner 2001; Setear 1997), who maintain that the inclusion of flexibility instruments in a trade agreement leads to a loss of credibility in the system and to less reciprocal market access concessions than an agreement without escape clauses. This is wrong: As Herzing has demonstrated, a well-crafted escape regime with truly commensurate remedies replicates full compliance by the injurer. Receiving the exact replacement value basically insures the victim of an

---

530 The victim’s expectation damages are indirectly dependant on the political support shock experienced by the injurer, since the latter’s optimal defection reaction is a function of the nature and size of the contingency.

531 In addition, the mutually agreed tariffs are lower for every discount factor countries may be endowed with (*ibid.*).

532 Higher tariff liberalization commitments make defection more attractive and skew the H&R-curve leftwards, as Chart 7.4 shows.

533 “Since in addition to the benefit from increasing possibilities from liberalization [i.e. more *ex ante* tariff commitments], the efficiency-enhancing effect under the optimal compensation costs scheme is sufficiently strong to increase per-period payoffs in relation to when there is no escape clause, an agreement on integrating an escape clause under the optimal compensation scheme will yield an unambiguously better outcome than when there is no escape clause” (Herzing 2005, p. 97).

534 Setear (1997) argues that trade policy flexibility is a step backwards in the process towards greater cooperation, since its relative ease of use increases opportunities for non-cooperation and the likelihood of defection.
escape measure. Hence one should not expect a loss of trust in the system, measured in terms of the continuation value of per-period cooperation.\(^{535}\)

To conclude this discussion on the inalienability of escape: In a non-stationary world a trade agreement concluded for reasons of overcoming a market access-induced PD may break down in finite time under a rule of inalienability. The introduction of an escape clause allowing for temporary deviation from cooperation gives rise to more cooperative behavior, even as shocks increasing the one-period gain from deviating occur. In order to prevent injurers from capitalizing on their private information as to the nature of the protectionist shock, some form of cost must be incurred each time the escape clause is enacted. The best results are achieved if this cost equals the expectation damages incurred by the victim country.

7.2.2.2 If stability of the system is not an issue

If we are not concerned with the stability of the world trading system, and tacitly assume that the agreement is enforceable because every Member has a vital interest in not letting the world trading system break down, it is even more straightforward to demonstrate the Pareto-superiority of an escape clause over a rule of mandatory specific performance in the tariff-setting game.\(^{536}\)

---

\(^{535}\) Rosendorff’s contention (2005, p. 396) that the inclusion of an opt-out mechanism lowers the value of the agreement for its members is an outflow of his reductionist modeling – not of the logic of escape clauses. Intuitively, agreements with an optimal flexibility mechanism adhering to the proportionality principle of efficient “breach” suffer no loss in cooperativeness, and thus do not generate less welfare for two reasons: First, there are instances in which decision-makers use opt-outs where before they had exited the agreement for good. This fact per se has intrinsic value. Second, using the opt-out mechanism is no more than witnessing efficient “breach” at work: In regimes without escape provisions, policymakers must decide between the continuation value of the game, and the temptation to cheat once and then revert to non-cooperation forever after. That means they are confronted with the one-off choice of staying in the agreement or aborting it. Yet under a commensurate compensation scheme policymakers are faced with a new kind of trade-off which now reads: Will a breach be efficiency-enhancing for the general welfare, or will it not? Since victims get fully insured by the contractual escape, they do not value the opt-out game any less. Injurers, hoping to achieve gains (net of compensation payments) certainly value the escape game more than not having this option. The opt-out scheme should thus be strictly welfare enhancing, not welfare deprecating. It is hence a codified Pareto principle.

\(^{536}\) The underlying assumption hereby is that the WTO exists in the “shadow of the grim trigger”: The threat of a grim-trigger punishment and thereby retreating to a highly protectionist past tacitly supports the system’s equilibrium path. Since no Member has an interest in returning to a non-cooperative Nash equilibrium, “all actual trade disputes, punishments, and defiances [sic!] of DSP rulings have been just parts of the equilibrium path” (Ethier 2001a, p. 4). Neglecting concerns over system breakdown brings with it significant modeling convenience: Issues of non-performance, remedies and ex ante commitment are not constricted by the narrow confines of an infinitely repeated trade-setting game. Instead, they can be modeled within a one-shot game, whose outcomes are assumed to be supported by the threat of reverting to a grim-trigger punishment. This allows for more structure in the model (additional decisions and actions).
To see this, let us revisit the formal illustration of incomplete contracting from subsection 3.3.2.1. The general model introduced above is convenient at this point, since it is exactly applicable to our situation of an optimally non-contingent incomplete contract. We assume that two risk-neutral, self-interested and reasonably rational players negotiate the trade deal behind a veil of ignorance. They design a governance structure for a single-transaction, one-shot PD game (or, equivalently, but more realistically, a series of unrelated future periods). In the presence of necessary incompleteness, the agreement is rationally organized as a non-contingent, perfectly indefinite contract that consists only of a description of the level of market access concessions (assume tariff cuts for simplicity), and the agreed-upon rule of escape-cum-remedies (see section 7.2.1 above).

After the details of the governance structure are sealed, the two parties negotiate the (discrete) reciprocal level of tariff liberalization, whereby the most reluctant liberalizer sets the common standards of commitments applicable to both parties. Let us define as level of cooperation, agreed to ex ante by each player. Here, can be understood as the trade liberalization level generated from reciprocal tariff cuts. , whereby 1 equals full trade liberalization. The level of mutual cooperation is endogenously determined (negotiations are not modeled), whereby the most reluctant liberalizer sets ;

\( \theta \) as unforeseen contingency that hits exactly one player. Behind the veil of ignorance the players are ignorant as to which party is affected. \( \theta \in [1, \infty] \) is an exogenous protectionist political support shock of some magnitude. The player affected by the shock may experience regret, and consequently assumes the role of the injurer. The revelation of the shock is private knowledge to the affected party.

\( B \) as “breach” set, that is, \( \{ \theta | \text{the contract will not be performed} \} \);

\( \Delta \) as damage measure (a monetary compensation, for simplicity) payable by the injurer to the victim;

---

537 See Ethier (2001a) and Sykes (1991, Appendix A). Both models demonstrate that the introduction of an escape clause into an incomplete (but enforceable) contract Pareto-dominates a rule of inalienability. However, both models assume that contingencies are symmetrically observable.


539 \( \theta \) is a random i.i.d. shock. The probability density function of \( \theta, p(\theta) \), is common knowledge. \( p(\theta) \) is exogenous, differentiable and stable over time.
The WTO* as an efficient “breach” contract

$W^C(c, \theta)$ as value enjoyed by the injuring country if it gets hit by the exogenous shock, but *performs* from the contract as promised (*Cooperation pay-off*);

$W^D(c, \theta)$ as value enjoyed by the injuring country if it gets hit by the exogenous shock and *escapes* from certain contract obligations (*Defection pay-off*);

$W^D(c, \theta) - \Delta$ as value enjoyed by the injurer, given “breach” and subsequent damage payment;

$V^C(c)$ as value enjoyed by the victim country if the injurer performs his obligations as promised (*Cooperation pay-off*);

$V^D(c)$ as value enjoyed by the victim country if the injurer *escapes* from certain contract obligations (“Sucker’s payoff”);

$V^D(c) + \Delta$ as value enjoyed by the victim country, given “breach” by the injurer and subsequent damage payment.

$V^N(c)$ as value enjoyed by the victim in the pre-contractual non-cooperative past (“Nash” payoff).540

Behind the veil of ignorance, that is, unaware of which party is going to be hit by a protectionist shock in any given period, each policymaker aims at maximizing her expected welfare $E[Z(c, \theta, B(c))]$:

$$E(Z) = \frac{1}{2} \left[ \int_B W^C(c, \theta) p(\theta)d(\theta) + \int_B W^D(c, \theta) p(\theta)d(\theta) \right] + \frac{1}{2} \left[ \int_B V^C(c) p(\theta)d(\theta) + \int_B V^D(c) p(\theta)d(\theta) \right],$$

where $B$ is the “breach” set and $B^\sim$ the non-“breach” (i.e. performance) set. Note that by maximizing $E[Z(\cdot)]$ each signatory *ex ante* acts as if it were maximizing the common welfare, or the general efficiency of the entire game. This is consequential, because it gives us a social planner’s approach to the issue. The benchmark for the two reasonably rational policymakers is thus the unachievable Pareto-efficient complete contingent contract (CCC), which maximizes the sum of the expected values of the contract to the injurer and the victim.541

Whenever reacting to a large protectionist shock is apt to generate joint welfare gains – which is indubitably the case in a tariff-setting game (think of the shock as a balance-of-

540 As in every PD, $W^D(c, \theta) > W^C(c, \theta)$, and $V^C(c) > V^N(c) > V^D(c)$.

541 See Mahlstein and Schropp (2007, section D).
payment crisis, e.g.) – the CCC will prescribe non-performance. More precisely, the Pareto efficient “breach” set equals $B_{opt}^{\delta}(c)$, where

$$B_{opt}^{\delta}(c) = \{ \theta \mid V^D + W^D \geq V^C + W^C \},$$

or equivalently

$$B_{opt}^{\delta}(c) = \{ \theta \mid W^D - W^C \geq V^C - V^D \}.$$ 

Thus, in the hypothetical world of a CCC, the injuring country’s regret contingency (the anticipated protectionist political support shock) must be large enough such that its regret $(W^D - W^C)$ outweighs the harm done to the victim $(V^C - V^D)$. Wherever shocks exist that satisfy this constraint, non-performance is permissible, indeed obligatory. Since the CCC is the first-best contract, it also induces the optimal reciprocal trade liberalization level $(c^\ast)$.\textsuperscript{542}

In the absence of complete rationality and foresight, the CCC must remain a distant ideal for the two signatories. In the presence of incertitude, the contracting parties must decide whether \textit{ex post} discretion should be permissible or not. A rule of mandatory specific performance, now, is no more than a prohibition of any kind of \textit{ex post} non-performance. The breach set is the empty set, $B^{IR}(c) = \{ \}$, whereby $IR$ stands for “inalienability rule”. It is intuitive to see that under a rule of inalienability an injurer (the better nomenclature here would be “party affected by an external shock”), barred from seizing \textit{ex post} regret contingencies, will experience a loss of expected welfare every time a political support shock of substantial size occurs. This condemns the injurer to engage in \textit{under-“breach”}. Inefficiently little room for escape will naturally reduce injurers’ \textit{ex ante} willingness to liberalize. Since behind the veil of ignorance every signatory must anticipate assuming the role of the injurer at least half of the time, an empty breach set and the resulting \textit{under-“breach”} will equally induce fewer \textit{ex ante} trade liberalization commitments.

\textsuperscript{542} Check subsection 3.3.2.1 and Mahlstein and Schropp (2007, section D). Indeed, $c^\ast$ is the politically optimal tariff for both policymakers, given the symmetry of the game set-up.
The Pareto-inferiority of a rule of mandatory performance becomes even more pronounced if we compare the expected payoffs under the inalienability regime with a liability rule of default accompanied by expectation damages: Under a rule of mandatory specific performance

\[
E(Z^{LR}) = \frac{1}{2} \left[ \int_{\theta}^{B} W^C(c, \theta)p(\theta)d(\theta) \right] + \frac{1}{2} \left[ \int_{\theta}^{B} V^C(c)p(\theta)d(\theta) \right],
\]

since \textit{ex post} escape is prohibited. Under a liability rule the expected payoff for each signatory equals

\[
E(Z^{LR}) = \frac{1}{2} \left[ \int_{\theta}^{B} W^C(c, \theta)p(\theta)d(\theta) + \int_{\theta}^{B} W^D(c, \theta)p(\theta)d(\theta) - \Delta \right] + \\
\frac{1}{2} \left[ \int_{\theta}^{B} V^C(c)p(\theta)d(\theta) + \int_{\theta}^{B} V^S(c)p(\theta)d(\theta) + \Delta \right].
\]

The damage payment \(\Delta\) cancels out and yields

\[
E(Z^{LR}) = \frac{1}{2} \left[ \int_{\theta}^{B} W^C(c, \theta)p(\theta)d(\theta) + \int_{\theta}^{B} W^D(c, \theta)p(\theta)d(\theta) \right] + \\
\frac{1}{2} \left[ \int_{\theta}^{B} V^C(c)p(\theta)d(\theta) + \int_{\theta}^{B} V^S(c)p(\theta)d(\theta) \right].
\]

Since the expectation damage measure induces the optimal “breach” decision \((B^{opt})\) on the part of the injurer, and since it holds that

\[
W^D(c, \theta) > W^C(c, \theta), \quad \text{and}
\]

\[
[W^D(\cdot) - W^C(\cdot)] > [V^C(\cdot) - V^S(\cdot)] \quad \forall c, \theta \text{ under } B^{opt},
\]

the expected payoff under a liability rule \(E(Z^{LR})\) is strictly larger than expected payoffs under a rule of inalienability \(E(Z^{IR})\).

To conclude: Behind the veil of ignorance any signatory will opt for escape-cum-expectation damages instead of a rule of inalienability, since the former strictly Pareto-dominates the latter in terms of expected future payoffs.

---

\(^{543}\) See equation (4) in subsection 3.3.2.1.
7 The WTO* as an efficient “breach” contract

7.2.3 A property or liability rule of escape? A question of transaction costs

The last subsection demonstrated that the provision of a rule of post-contractual escape Pareto-dominates a situation where performance is always mandated, no matter what the contingency. This subsection now delves into the issue of efficient non-performance design in a trade agreement based on market access externalities: Is a liability rule or a property rule of entitlement protection more likely to precipitate efficient ex post performance and optimal ex ante trade liberalization concessions?

As a preliminary factor, signatories will want to know which remedy measure best complements a liability rule. The results generated in subsection 3.3.2.1 feed into the situation at hand, as confirmed by Mahlstein and Schropp (2007), and Schropp (2007, Appendix at section E): Behind a veil of ignorance, and in a situation of an optimally indefinite, non-contingent tariff liberalization agreement, and where political support shocks are asymmetrically revealed, signatories will opt for the expectation damage measure. Expectation damages yield Pareto-optimal non-performance decisions: Not only do these then generate an optimal “breach” set, they also effectuate the politically largest possible reciprocal trade liberalization commitments ex ante. In other words, only remedies amounting to the expectation damages can replicate the outcome of the first-best of the CCC.\(^{544}\)

This leaves us with the sole question of whether the market access entitlement in the hypothetical WTO* should be protected by liability-cum-expectation damages or by a rule of renegotiation. As explained in subsection 3.3.2.2, the answer can be found in the inherent trade-off between the transaction costs of renegotiation and the TC of damage calculation. Indeed, that subsection listed a catalog of criteria that reasonably rational trade negotiators should observe when deciding on the relevant rule of flexibility.

7.2.3.1 Transaction costs of damage calculation

When deliberating whether to put in place an LR to protect the reciprocal market access entitlement, trade negotiators must be wary of the fact that both the original contingency and the damage caused by a protectionist measure are private knowledge to the injuring and victim government, respectively. Ex post facto, that is, after the injurer has performed partial withdrawal of previously made market access commitments, the victim has no incentive whatsoever to reveal its true expectation damages. This is the task of an

\(^{544}\) The intuition behind this result is the same as the one given in subsection 3.3.2.1 supra.
impartial arbitrator. A number of exogenous TC in connection with calculating damages must be taken into consideration:

1. **Cost of arbitration.** Having an arbitration entity in place is costly. Although dispute panels can be recruited to assume the job of calculating damages, the WTO* DSB may have to stock up its technical competence; calculation of the damage is essentially a quantification exercise best conducted by experienced economists.

2. **Monetization of expectation damages.** One distinct difficulty for the WTO* arbitrators lies in putting a “price tag” on the political, inherently subjective, harm incurred by the victim government. This is a complicated but by no means impossible task: Firstly, arbitrators are dealing with a tangible issue, namely a market access balance. Mandating that trade damages and lost market access opportunities be offset by additional access in other industries is a doable task, and probably a reliable proxy for the subjective impairment suffered by victim policymakers. DSU panels have collected ample experience of how to deal with market access infringements in the last 60 years. Secondly, as Kaplow and Shavell (1996a, pp. 726) have shown, an independent arbitrator neither needs to be omniscient nor operate flawlessly in order to be effective. As long as its judgment is neither completely incompetent nor systemically biased against victims or injurers, arbitrators can succeed in producing satisfying outcomes.

3. **Implicit calculation of expectation damages by the injurer.** Before the injurer decides to enact a protectionist measure, it will have to assess the likely circle of victims, and the harm which its actions can be expected to cause to these countries. The injuring Member is faced with the same problem as the arbitrators later on: They have to put a price tag on the likely nullification and impairment to victims’ market access entitlement, and cannot count on honest support by the victim(s). For the injuring government this means investing time and resources in research, prior to enacting policies that it might regret thereafter. On the positive side, the uncertainty involved may prompt injuring governments to opt out only in those cases where the political support shock is definitely large enough to justify escape.

It is noteworthy that endogenous TC, or strategic behavior, on the part of the victim or injurer are not present under a remedy-backed LR. Subsection 3.3.2.2 hinted at the
problem of overinvestment (or over-commitment) by the victim in situations of an asymmetrical revelation of contingencies. Yet behind a veil of ignorance where parties do not know their future role before concluding the contract, overinvestment by the victim is not an issue.\textsuperscript{546}

\section*{7.2.3.2 Transaction costs of tariff renegotiations}

The choice for renegotiation as a trade flexibility tool, on the other hand, brings with it a different set of TC-induced problems:

\begin{enumerate}
\item \textbf{Implicit calculation of damages and non-performance gains by injurer and victim.} When the party experiencing regret and the potential victim(s) sit down to renegotiate the terms of their mutual market access balance, both governments have to invest in serious upfront preparations. As is the case under an LR, the injuring party must spend time and effort researching the likely consequences of its actions for the victim(s). In addition, however, each victim government will want to investigate the potential non-performance gains of the injurer in the hope of appropriating as much of these gains as possible. To that end, victim countries must necessarily detect the original political support contingency which hit the injurer. Eliciting the original contingency and (based on that conjecture) calculating the efficiency gains from non-performance is a complex business, and one which by far eclipses the task that arbitrators have to master under a liability rule of escape. Also, calculation errors by one party may be prone to holding up the entire renegotiation process.\textsuperscript{547}

\item \textbf{Costs of renegotiation.} Even if WTO* Members are benevolent and do not engage in strategic gamesmanship, renegotiation is a costly endeavor. Manpower and other negotiation resources are expended, which could be used more efficiently elsewhere. The
\end{enumerate}

\textsuperscript{545} By explicitly mandating expectation damages that completely insure the victim country and replicate full compliance on the part of the injurer, the WTO* avoids the one systematic error that is prevalent in the current WTO system: As we laid out in subsection 5.4.1.3, WTO arbitrators must be reproached for having misinterpreted commensurate damages as reliance- instead of expectation damages.

\textsuperscript{546} The reason is that behind a veil of ignorance all signatories take decisions in the manner of a social planner. The commitment to trade liberalization is made before the veil of ignorance is lifted. Since over-commitment is inefficient behavior, no rational signatory will choose that option.

\textsuperscript{547} Depending on whether renegotiation takes place in a hub-and-spoke manner with victims sequentially engaging in bilateral negotiations with the injurer, or whether it takes place in a convention-style meeting involving all victims, a calculation error on part of one victim may inadvertently hold out the entire settlement process.
more victims involved in the renegotiations, and the more heterogeneous these players are, the longer the deliberations will last.\footnote{From the experience of renegotiations under Art. XXVIII GATT we know that discerning which parties to involve in tariff renegotiations in the first place is a non-trivial issue: Only those that have an export history into the market of the injurer, or also those that could become important exporters in the near future? As Mavroidis (2007, section 3.6.3) shows, the question of which Members hold “initial negotiating rights”, qualify as “principle supplying interests”, or have “substantial interests”, has proven to be notoriously difficult during tariff schedule modifications under Art. XXVIII GATT. Once this issue is settled, renegotiations can commence, whereby the number and heterogeneity of players are prone to hold up the process, since the renegotiation deal necessarily gets more complex the more numerous and the more different players are.}

3. **Opportunity costs of time.** Renegotiations are also costly in terms of opportunities foregone: Whenever an injurer is hit by a significant regret contingency, every moment that lapses without the affected Member reacting is costly. With every minute devoid of a modification agreement, more efficiency gains from escape are lost. Depending on the nature and size of the political support shock, a negotiation-induced delay can completely frustrate all these non-performance gains. Victims thereby do not necessarily have to be malevolent or opportunistic. It just lies in the nature of renegotiations that the time lapsed between experience of shock and the decision to act is larger than under a liability rule of opt-out.

4. **Double-sided hold-out.** In addition to the exogenous (objective) TC listed above, there are serious endogenous (behavioral) ex post TC involved in the renegotiation game. Since both the regret contingency and the damage caused are private information to injurer and victim respectively, both parties have a strong incentive to misrepresent the true state of nature. A victim government, on the one hand, will try to misconstrue its likely damage so as to appropriate as much as possible of what it thinks are the injurer’s gains from non-performance.

The injuring Member, on the other hand, is very likely to engage in reverse hold-out: Since it has proprietary knowledge, the injurer has an incentive to misconstrue both the actual size of the exogenous shock and the extent of expected efficiency gains from non-performance.\footnote{Mahlstein and Schropp (2007), in a tariff-setting model with common knowledge of damage done to the victim (when the original contingency is private information for the injurer), formally show the non-trivial efficiency losses that entail sequential bargaining under a renegotiation provision of flexibility.}

This double-sided hold-out may lead to a war of attrition, both parties mistrusting each other’s assertions, and consequently rejecting the other party’s settlement offers. Wars of
attrition take very long to resolve and are hence very costly.\textsuperscript{550} Neither party will be able to convince the other that it is telling the truth. For fear of establishing a reputation as a soft negotiator, neither party wants to come out as a “loser” of this game, either. In addition, Fearon (1998, pp. 278) argues that those distribution fights, if resolved at all, are paradoxically won by the party with the lowest stakes in the matter, i.e. the country with the lowest discount factor (see also Rosendorff and Milner 2001 at pp. 850). Knowledge of this last fact is not good news for injurers. Anticipating such a disappointing result, injurers may refrain from participating in welfare-enhancing renegotiations right away, since in the marginal case their entire \textit{ex post} gains from non-performance could be appropriated by victims.

5. Redistribution disputes. A final drawback of a PR of flexibility is the occurrence of turf wars between multiple victims. Once victims (think to) have an idea of the size of the injurer’s non-performance gains, they will want to secure the biggest possible share of that pie. They do so by holding out the other victims by deliberately procrastinating the renegotiation process. A strategic victim thereby hopes that its competitors settle early and for less, so that itself can secure the lion share of these spoils.

7.2.3.3 Conclusion: \textit{A rule of liability exceeds a rule of renegotiation as default rule of the market access entitlement}

The presence of \textit{ex post} transaction costs impedes WTO* signatories and renders impossible a truly efficient system of policy flexibility in the market access argument. The liability rule accompanied by expectation damages entails much lower post-contractual TC than a rule of renegotiation. Although the maintenance of a competent WTO* arbitrator is costly, and the monetization of expectation damages not a trivial issue, an LR of default surpasses the alternative of tariff renegotiations by far: A key advantage of the LR is that it can be enacted immediately without any opportunity losses. Another important factor is that granting unilateral opt-out does not give way to strategic behavior, whereas under a PR both the party experiencing regret and \textit{all} potential victims of a suggested measure have an incentive to misrepresent the truth, to play for time, to procrastinate or to engage in other opportunistic strategies. This is a recipe for trouble: It undermines the spirit of togetherness and cooperation, makes mischief and is generally apt to sour the relations between contracting parties. Since the WTO* (and more so, the actual WTO) is a relational contract that crucially relies on the

\textsuperscript{550} See above discussion in subsection 3.3.2.2 (especially in footnote 234).
benevolence and goodwill of its Members, mechanisms that bring out opportunistic traits in signatories must be avoided at all costs.

In conclusion, a property rule of flexibility produces higher costs *ex post*, which prompts less efficient non-performance decisions by potential injurers.\(^{551}\) Anticipating this, signatories behind the veil of ignorance will adjust their *ex ante* trade liberalization concessions accordingly. The provision of an LR-cum-expectation damages on the other hand, though not a perfect tool, prompts relatively more efficient decisions by both injurers and victims.

### 7.2.4 Specifics of the default rule

The liability rule is the single default rule protecting the market access entitlement. This subsection will discuss the specifics of this rule of default. In particular, it assesses whether the enactment of the LR should be connected to some sort of conditionality, and how issues of non-violation should be treated.

#### 7.2.4.1 The level of conditionality – enactment threshold and application scope limitations?

Should the LR be linked to any kind of precondition, or enactment threshold? From the perspective of efficient “breach”, this would seem a foolish idea. First, enactment conditions are sunk costs, and as such do not bear any inherent value other than that of a signaling device: The injurer signals its resolve to comply in the future and to return to the cooperative path once the shock has subsided (Rosendorff 2005 at footnote 20; Rosendorff and Milner 2001, p. 831). However, paying the victim its expectancy under the optimal DR should be enough for the injuring country to demonstrate its cooperative zeal. Second, a high level of contingency is an additional burden for the injurer on top of expectation damages payable to the victim. But everything *in extenso* of expectation damages violates the precept of efficient “breach”, and reduces injurers’ willingness to default in situations where it is actually efficient to do so. Sunk-cost-pre-conditionality is tantamount to demanding over-compensation from the injurer. Over-compensation leads to insufficiently little non-performance, or *under-“breach”*, and ultimately to inefficient

---

\(^{551}\) In anticipation of later costs connected to a PR, injurers will refrain from choosing non-performance in the marginal non-performance decision. These are situations where the ideal CCC would prescribe *ex post* escape, but the injurer, in anticipation of *ex post* costs connected to this decision, will just abstain from engaging in contractual default.
ex ante investments behind the veil of ignorance. Therefore, parties must be expected to refrain from prescribing any sort of enactment threshold.552

Is it sensible to limit the application scope of the liability default rule? Mandating “MFN’ed” non-performance is not advisable: Protectionism should be allowed on a selective, discriminatory basis, just as the accompanying remedy payments are bilateral and commensurate. Since the victim(s) are compensated for their loss(es), non-discrimination would only expand the circle of victims, and therewith the damages to be shouldered by the injuring Member. A maximum period of escape (as in Art. 7.1 SGA, for example) does not seem reasonable, either: As long as injurers comply with paying expectation remedies, their escape is efficient and therefore deserves to be endorsed.

However, two limitations on using the liability opt-out are crucial: First, protectionist backtracking in reaction to market access shocks should be limited to tariff barriers only. Tariffs are the most direct measure with the fewest externalities and dead-weight losses (Krugman and Obstfeld 1994, chapter 9). Also, import taxes are a transparent measure (readily observable by all parties), and easily adjustable.

Second, the damage remedy should consist of compensation only. Tariff compensation offers by the injuring Member to the victim country/countries politically and economically Pareto-dominate the countermeasure of retaliation: Compensation avoids the negative political-economic implications of the suspension of concessions, such as the small-country bias, the economic dubiety, and the contract-defeating nature of retaliation.553 More importantly, tariff compensation does not drive down the negotiated market access balance (a second-order effect of retaliation), but instead keeps it on the same – presumably politically optimal – level that was initially negotiated.554

552 In the same vein, it should be evident that special concessions granted to the injurer are equally ineffective. Art. 8.3 SGA currently confers a three-year grace period for injurers invoking the safeguard of Art. XIX GATT. There is no logical reason why compensation payments by the injurer should set in with a latency of 3 years. This is another violation of the efficient “breach” principle.

553 See our discussion on the weaknesses of the countermeasure of retaliation in subsection 5.4.1.3.

554 Compensation remedies in the form of monetary fees, as championed by some authors (Barfield 2001at pp. 131; Bronckers and van den Broek 2005, pp. 109; Davey 2005b; Pauwelyn 2000, pp. 345; WTO 2004, chapter VI.D; O’Connor and Djordjevic 2005) seem less attractive from the point of view of self-interested policymakers: Although monetary fines are more fungible than tariff liberalization offers, it should be remembered that the currency of the WTO contract is political support, not consumer welfare. To policymakers, seeing new market access opportunities may be worth more than monetary fees. Also, monetary fees may imply negative political-economic consequences for the injuring government, such as public outcry (“selling indulgences” to trade rivals) or having to involve the domestic legislative.
7.2.4.2 Non-violation claims and the default rule of flexibility: Two concepts apart?

As stated in subsection 4.3.1.1, WTO(*) (shorthand for WTO and WTO*) Members may inadvertently close their partners’ market access at times. For that situation, the current WTO Agreement has non-violation claims in stock, though they have hardly ever been used. Reasonably rational policymakers can be expected to staff the WTO* with a workable NVC that can be appealed to whenever a country feels that its market access balance is out of sync subsequent to any kind of non-reported protectionist escape – be it inadvertent or not. The affected country approaches a WTO* arbitrator, who consequently assesses if and how much harm occurred. An NVC is substantially equivalent to a DR, with the sole difference that the initiative does not come from the injurer, but from the victim’s side. Thus, in contrast to the current system where the relationship between NVC and Art. XXVIII is somewhat dubious (cf. subsection 5.4.1.1), under the WTO* they emanate from the same basic concept of liability default.

7.2.5 Conclusion: An unconditional liability rule as optimal protection of the market access entitlement

The optimal institutional design which rational trade negotiators will negotiate in ignorance of their future role in the trade game consists of two components: An optimally indefinite, completely non-contingent definition of the reciprocal market access obligation, and a single rule of default. In the presence of contractual incompleteness of the necessary or inexorable-type B sort, neither contingency measures nor negative integration provisions add anything to contractual efficiency – future states of nature are left undescribed. The mutually granted right to compete in trade partners’ markets is protected by a simple liability rule of protection: The provision features an unconditional opt-out clause, and a stipulation of commensurate tariff compensation amounting to the expectation damage measure.

This single rule of default, accompanied by the expectation damage remedy, is able to replicate the outcomes of the complete contingent contract: Escape from the previously agreed market access level (“breach”) is only performed in situations where it is globally efficient. The injurer internalizes all the harm caused to the victim, and makes the latter indifferent between performance and escape. The injuring Member appropriates all efficiency gains from non-performance, and thus has an incentive to seize regret whenever it occurs. Yet the institutional design precludes the injuring Member from
enacting any inefficient redistributive measure, i.e. opportunistic \textit{ex post} backtracking. This safeguards optimal \textit{ex ante} commitment by all parties behind the veil of ignorance.

This unconditional \textit{intra}-contractual LR is contractually fixed within the GATT/GATS Agreements, but not within the DSU. For the liability-rule regime to work, WTO* Members need binding third-party arbitration, the procedures of which have to be contractually specified. Whether the arbitration shall be performed by WTO panels or the AB, or some other body of the WTO Secretariat is inconsequential. What is important is that this arbitration function is not to be conflated with any adjudicatory role or court-like procedures. Arbitration is not about assessing right or wrong, or settling disputes. It is about finding the right price for a legitimate escape from previously agreed contractual market access obligations.

7.3 Organizing the protection of multilateral entitlements

The next step of our hypothetical bargain exercise is to assess how reasonably rational policymakers are organizing the protection of the various multilateral entitlements that are exchanged in the WTO(*). Before the true state of nature is revealed to the WTO* Members, they are ignorant about the nature and size of possible non-trade contingencies, about which country gets hit, for how long and to what extent. Therefore, again, reasonably rational signatories search for the institutional design that maximizes the common welfare of all participating trade policymakers.

7.3.1 Focusing on default rules

For the same reason that applies to the market access entitlement, it is not rational for policymakers to strive for contractual completeness in a situation of \textit{necessary} or \textit{inexorable-type B incompleteness} that all non-trade entitlements are beset with (cf. subsection 4.3.2 above). Hence, trade decision-makers can be expected to concentrate on the drafting of default rules for the protection of multilateral entitlements. Forward-looking and vigilant drafters examine \textit{every} multilateral entitlement, and assign it its own default rule.
7.3.2 The optimal design of default rules protecting multilateral entitlements

Not all multilateral entitlements are of similar type; there are major qualitative differences. Some entitlements are minor procedural obligations, some are far-reaching commitments to abide by certain minimum standards of regulation, others are external codes of conduct integrated into the contract. Negotiating policymakers should classify non-trade entitlements in three groups: Group 1 consists of entitlements where *ex post* discretion is impermissible for moral or systemic reasons. Those entitlements are to be protected by mandatory specific performance, i.e. a rule of inalienability. Group 2 are those entitlements whose infringement is not welcome, but causes only minor nuisance to the community as a whole. These are to be protected by liquidated damages. Group 3 consists of all other multilateral entitlements which are reasonably protected by a property rule of flexibility. We shall explain each in turn.

7.3.2.1 If *ex post* discretion is impermissible: An inalienability rule of default

As discussed in subsection 3.3.1 above, rules of inalienability are apposite in situations where *ex post* non-performance is immoral, contract-annihilating, or welfare-depreciating. All these situations can occur with respect to some multilateral entitlements:

*Ex post* default is immoral with respect to some external entitlements, such as *ius cogens* norms. Peremptory norms of international law supersede every treaty provision in international law. Sovereign countries must not “contract around” them. They must not deviate from these entitlements, either. Backtracking is also immoral with respect to the voting rights of Members: No injuring party should be able to infringe upon the voting rights owned by other Members. In the same vein, no Member should ever be allowed to relinquish its voting right.

A rule of inalienability has to be put in action whenever contractual escape crowds out the cooperative drive by other signatories. This is especially the case for entitlements where partial performance is not possible, and backtracking is always tantamount to completely nullifying the previously made commitment. In certain cases, some kind of “domino effect” is triggered if a single party defects from its multilateral entitlement:

---

555 It is not possible to partially overstep a deadline; it is not feasible to protect copyrights a bit, to have half a patent office in place, or to only minimally infringe upon *ius cogens*. Compare this to the non-dichotomous (discrete) market access entitlement, where partial non-performance (e.g. temporary protectionist increase of the United States’ tariffs of hot-rolled steel from 10 to 35 percent) is the norm.
The prospect and anticipation of one signatory stepping out of line frustrates the drive of (some or all) other parties to commit *ex ante* to some level of regulation, or some minimum standard. The result of temporary non-performance by an injurer is not *less* regulation, but *no* regulation. Take for example the multilateral obligation to protect patents for 25 years. If that entitlement can legally be reneged upon *ex post*, and country A decides to grant only 2 years of patent protection, A’s neighboring countries will not want to protect any longer than the defector.556 Anticipating several countries to abandon a rule of extended patent protection, no other country will want to commit to the full 25 years. The same logic applies to the protection of copyrights. If countries collectively commit to grant extended copyright to software codes, but, say, China “opts out” of that obligation, other WTO Members must anticipate that many plagiarist companies will settle down in China and undermine the entire system. To avoid a collapse of the regulation from the start, multilateral copyright provisions are best protected by rules of inalienability.

In summary, when drafting the terms of entitlement protection, signatories should think hard about which multilateral entitlements are best protected by a rule of inalienability. In particular, dichotomous, far-reaching, commitment-intensive regulation promises warrant mandatory specific performance at all times so as to avoid a crowding out of *ex ante* cooperation concessions.

### 7.3.2.2 If *ex post* discretion causes minor damage: A liquidated damages rule of default

Following the nature of multilateral entitlements, *ex post* violation is presumably rarely ever welfare-enhancing. The regret contingency must be substantial enough to make up for the harm done to *all* other signatories. Thus, post-contractual non-performance should generally not come too easily for signatories. However, there may be multilateral entitlements whose – possibly inadvertent or negligent – infringement will not cause a great deal of harm to affected Members. Take for example instances of minor misdemeanor, such as an overstepped deadline, an omission to report or notify, or a late payment. These administrative or regulatory offences may be perceived as petty.

---

556 Patent protection creates monopoly rents for the patent holder, and thus prevents economy-wide technological spillovers. Patents hamper technological progress. Consequently, governments may be reluctant to grant extensive patent protection. Offering less protection may give a country a technological head-start in global competition (see e.g. Romer 1996, chapter 3).
infringements in the grand scheme of things. They cause minor nuisance to the system as a whole, but hardly affect any contracting party in particular.

An entitlement protection in the form of an LR or PR is then not warranted, simply because the transaction costs connected to lengthy renegotiations or arbitrated liability easily outweigh the damage done. Yet infringements should not be ignored, either. For that reason, in situations of minor nuisance, a liquidated damage rule of default seems the most sensible and expeditious action. Liquidated damages should be made in cash and be payable to the Organization.557

7.3.2.3 If ex post discretion is permissible: A property rule of default

For all those multiple entitlements which neither fall in the category of immutable protection, nor bagatelle provisions, ex post discretion should be permissible. We find strong evidence that multilateral entitlements of this third group are best protected by a property rule of renegotiation. Consider the following reasons:

1. A rule of liability is a slippery slope. Contracts on the scale of the WTO(*) need some form of institutional spine. Multilateral entitlements are the structural backbone of the WTO(*) contract. Without the unfailing presence of numerous procedural obligations, timelines or general rules of *modus operandi*, predictability and constancy of the international trading regime suffers. If most of these entitlements were “up for grabs” by means of a liability rule, the system would be in danger of destabilization: If no country could ever be sure that its trading partners stick to their obligations owed to the entire membership, and instead had to learn *ex post facto* that a liability-type opt-out occurred, trust and confidence in the contract would suffer. A liability rule for nearly every entitlement would render the idea of a binding agreement futile, and dissolve the very nature of the contract.

2. Strong presumption of opportunism. As stated above, the regret contingency for multilateral entitlements must be significant, since it should exceed the harm done to all signatories that benefit from *erga omnes partes* obligations. The exercise of a unilateral opt-out is generally welfare-depreciating, i.e. opportunistic. Opportunism may happen whenever damages cannot easily be calculated (on this point see next paragraph).

557 This is not to suggest that *any* deadline, statutory guideline, or notification provision should be “up for grabs” using a simple violation-cum-liquidated damages mechanism. Many multilateral entitlements might be outright annihilated by such a provision. The point is merely that reasonably rational contracting parties should carefully assess, which of the many multilateral entitlements should really fall under a relaxed remedy rule.
Signatories can be expected to foreclose any possibility of opportunism upfront by demanding renegotiations instead of unilateral opt-out accompanied by remedy payments.

Also, if the injuring Member is guided by discontent with some systemic defect of the contract rather than opportunistic guile, its desire for ex post non-performance is akin to a modification request. Modifications of systemic flaws are better addressed by renegotiation than by escape-cum-remedy.

3. Liability damages are difficult to measure. This is an important argument against a liability rule of default for multilateral entitlements: It is very difficult to calculate damages caused by a unilateral opt-out of a multilateral entitlement. As pointed out in section 5.2 above (see especially footnote 401 and accompanying text), the quantification of damage caused by an erga omnes obligation entails logical and practical problems: Any encroachment from a multilateral entitlement by one signatory harms the system as a whole. But how can damages to the system be measured?

Assessing damages pursuant to an escape from an erga omnes entitlement is hypothetical, next to impossible to assign and apportion, and difficult to “monetize”: Any WTO* arbitrator would be charged with the non-trivial task of assessing how trade in the international trading system would have evolved had the escaping party performed as promised. Further, the arbitrator would also have to establish to what extent every single contracting party suffered as a result of the non-performance, and how the opt-out affected the competitive relationship between all signatories. Apart from the difficulty connected with the counter-factual nature of that calculation, the damage can be assumed to be profoundly subjective for every victim.558 In addition, every systematic calculation error by the arbitrator is multiplied 150-fold in effect, since all of the 151 WTO(* ) Members (save the injurer) would receive the wrong expectation damage.

In summary, having an arbitrator assess and apportion damages caused by multilateral entitlement infringements is likely to create tremendous discontent among the

558 For example: An arbitrator would have to calculate what the counterfactual level of world trade would have been, had Norway in the Norway – Trondheim Toll Ring case from 1991 publicly tendered the construction work. This calculation should include general equilibrium considerations, second-order ripple effects, and third-party externalities. The arbitrator would also have to assign expectation damages to the U.S., Burkina Faso, Vanuatu and all other WTO* Members and argue convincingly as to why the remedy amounts differ. Things get more difficult if a defection from a multilateral entitlement does not cause palpable harm, but intangible damage: How can the subjective harm to Canada following, say, a refusal by the United States to pay the yearly financial contribution to the WTO be measured?
membership. A PR of default, on the other hand, would avail an impartial bystander of this unmanageable task. Under a renegotiation provision the injurer sits down with all interested parties, explains its problem and negotiates a solution with its trade partners. This provides victim Members with time to reflect on the implications of a possible backtracking measure, and on the expected harm the measure in question is likely to cause them.

4. Probability of strategic victim behavior is small. While damage calculation is notoriously difficult, the transaction costs of renegotiation can be assumed to be minor for cases of escape from multilateral entitlements. Harm done through ill-meaning strategic behavior is of less concern: First, since non-trade contingencies are usually not imminent emergencies that could threaten the existence of countries (or rather, the political survival of the policymakers in charge), injurers are under less time-pressure, and may not easily be held out by victims applying procrastination strategies in order to carve out a better deal for themselves.

Second, although all WTO* Members can potentially partake in the renegotiation of multilateral entitlements, few victims will be interested in participating, and those who do have less interest in behaving strategically. The individual damage to every country is often minor and of little interest to many players (what is the damage to Viet Nam if Chile fails to pay its membership fees to the WTO*?).

Third, if the renegotiation request concerns an important topic, concerned parties will think twice before engaging in strategic gamesmanship: Due to the repeated nature of the trade game, every country must reckon on having to escape from multilateral obligations at some point in time. Holding out the injurer, or vetoing a non-performance request, will be remembered within the membership. And since every Member can take part in every renegotiation, the formerly impeded injurer may easily pay back the “new” injurer commensurably. Thus, the consensus principle of renegotiations could be a blessing: Few signatories will want to have the reputation of being a recalcitrant or opportunistic victim.

7.3.3 Conclusion: Mixed default rules of protection for multilateral entitlements
Reasonably rational trade negotiators ought to concentrate their efforts on designing workable default rules for multilateral entitlements rather than on drafting elaborate contingency rules. We have shown that the exercise of ex post trade policy flexibility should not be made easy for injurers when it comes to multilateral entitlements. It is
reasonable to divide the vast number of non-trade entitlements into three groups: one for cases where *ex post* discretion must be considered as immoral or welfare-depreciating; one for those entitlements where non-performance is a mere nuisance; and one for those where post-contractual non-performance is permissible and protected by a rule of renegotiation.

Thus, the WTO* Charter should include an article comprising of at least three rules of default. Multilateral entitlements throughout the contract can then refer back to this provision. Alternatively, every treaty article listing a non-market access entitlement should also contain in a final paragraph the applicable rule of default. Special attention should be directed to the issue of how entitlements can be protected by the relatively weak liquidated damage rule as well as by the rigid rule of inalienability.

### 7.4 A two-tier system of enforcement

We have characterized the *intra*-contractual entitlement protection rules for the trade and non-trade entitlements in the form of simple default rules. This system of DR ensures that any substantial political support regret experienced by self-interested policymakers during the performance phase of the contract can be seized, *if* doing so enhances the general welfare of the contracting parties. Thus, every WTO* Member acting in good faith can use the contractual safety valves, provided it abides by the rules of non-performance attached to the entitlement it wants to withdraw from. By virtue of the efficient “breach” provisions, no country is forced to use violation of the Agreement in order to liberate itself from untenable commitments.

What remains for this hypothetical trade agreement to deal with are bad-faith and haphazard clashes between countries. Haphazard clashes are provoked by unintentional and inadvertent instances of default, and accidental contractual gaps (textual ambiguities, ambivalent formulations, omissions, erratic provisions). Reasonably rational contracting parties cherish no illusion that their contract will feature these inadvertent gaps. An elaborate system of enforcement should be able to separate good-faith clashes pursuant to inadvertent gaps from bad-faith clashes motivated by sheer opportunistic guile. Clarifying contractual ambiguities and filling inadvertent gaps has an intrinsic positive value for the WTO* membership as a whole: Eliminating haphazard gaps leads to transaction cost efficiencies and makes trading easier and more predictable. The institutional framework should encourage signatories to bring those issues to light, and
under no circumstances should it dispirit Members from openly questioning problematic or erratic provisions.

Reasonably rational trade policymakers should craft a two-tier system of enforcement. Designed as an escalation scheme, enforcement rules deal with extra-contractual behavior in two stages: A dispute stage and a punishment stage.\textsuperscript{559}

The first tier is aimed at eliminating welfare-enhancing good-faith trade disputes, and at solving them in an amicable manner. This dispute stage gives parties the opportunity to demand clarification of the treaty language where they see fit, and to subsequently resolve the dispute harmoniously. Whenever a signatory feels that its entitlements are nullified or impaired by the actions of another Member, it can appeal to a Dispute Settlement Body. The DSB hears the litigating parties, collects the facts of the case, and interprets contractual provisions according to the object and purpose of the entitlement.\textsuperscript{560} In its ruling, a dispute panel clarifies ambiguities and fills inadvertent gaps.\textsuperscript{561}

If it turns out that the claimant indeed witnessed a nullification or impairment of its rights, or that the defendant infringed upon a multilateral entitlement, the DSB in this first phase of enforcement treats the incident like a contractual DR, where the injurer failed to notify the victim(s): For the market access entitlement protected by an LR of default, the arbitrator (after having given the litigants the opportunity to reach a mutually agreed solution) will calculate the expectation damages, including foregone profits during the litigation phase, and the litigation costs incurred by the victim. The injurer is then instructed to settle these remedies in the form of tariff compensation offers.\textsuperscript{562}

For those entitlements protected by a property rule, the solved dispute is handed back to the two disputing parties with a request for them to reach a renegotiation solution, just as

\textsuperscript{559} We stated in subsection 2.1.3 that contractual enforcement always consists of two phases – a dispute or litigation phase, in which issues of enforceability (observability, verifiability, quantifiability) are paramount, and a punishment or remediation phase, in which enforcement capacity plays a crucial role. Many scholars equate enforcement with punishment and thereby ignore (or assume away) that detecting, defining, and recognizing contractual deviation is not at all a clear-cut case.

\textsuperscript{560} We are aware that dispute panels and the AB are required by WTO and international law to primarily interpret ambiguous passages in light of the object and purpose of the provision. However, the WTO* is of such a simple structure that objective of the provision is tantamount to finding out the nature of the relevant entitlement.

\textsuperscript{561} On the gap-filling role and competence of the WTO DSB, see (Keck and Schropp 2007, section D.5).

\textsuperscript{562} Note that tariff compensation amounting to expectation damages is used for intra-contractual escape and in the dispute stage of WTO* enforcement. A resemblance to non-violation claims (see subsection 7.2.4.2) is not coincidental. The dispute phase of WTO* enforcement is actually following the same concept.
is required by the initial property rule of contract entitlement. To that end, disputing (or involved) parties are granted a certain timeframe. But since renegotiation *ex post facto* is a difficult business (the defection has already occurred, so the injurer may have little incentive to acquit the real price of escape), compensation negotiations may break down, or remain inconclusive during the period granted by the DSB. Therefore, after the lapse of the renegotiation period, the arbitrator calculates the expectation damages incurred by the victim(s) from the point of defection to the end of the renegotiation period, and adds half of the non-performance gains (or its best estimate thereof) on top. After all, this sum is equally able to uphold the efficient “breach” principle just as well as renegotiations do. Entitlements protected by inalienability may not be infringed upon, and any defection on the side of the injurer is *ipso facto* interpreted as being driven by bad faith.

The dispute stage of WTO* enforcement is apt to resolve any ambiguity that existed between WTO* Members without introducing any punitive element to the system. Any party that has been driven by good faith up to this point should be satisfied: Injurers may engage in efficient “breach”, and victims may bring actions for nullification and impairment. If thus far a trade dispute has not been resolved, the strong presumption holds that the injurer acted in bad faith from the start.

Hence, in case inalienable entitlements are infringed, or an injurer Member steadfastly refuses to abide by the rules of the game – despite the opportunity to resolve the dispute affably – the *punishment* stage of WTO* enforcement sets in as a second line of defense for the victim. In this stage, enforcement has more “teeth”: First, since the risk remains that non-compliant Members disregard their duty to compensate, the *countermeasure of retaliation* is the ultimate means of enforcement. Retaliation has the incontrovertible advantage of not being controlled by the offender. Second, there is no logical reason for signatories to stay within a “rebalancing” logic, or to keep a dispute bilateral. In order to protect the previously agreed rules of the game, to induce compliance with the panel

---

563 The intuition behind adding half of the efficiency gains from non-performance is simple: Under a PR of escape, victims stand a good chance of appropriating some of the efficiency gains from non-performance gains (see section 3.3 above). If, pursuant to an infringement of an entitlement protected by a PR, the arbitrator only awarded expectation damages (which grant *all* the gains from non-performance to the injurer), no injurer would ever choose the route of renegotiation in the first place. Instead, the injuring Member would wait to be sued and receive a *higher* reservation utility from reimbursing the victim with expectation damages.

564 Strengthening the use of cross-retaliation thereby seems apposite. However, in order to make cross-retaliation a workable tool, the DSB’s restrictive, “superficial and inconsistent” (Hudec 2002, p. 90) interpretation of Art. 23.3(c) DSU, originating from the *EC – Bananas* adjudication, must be revised (cf. Schropp 2005 at footnote 47 and accompanying text).
ruling as quickly as possible, and to deter extra-contractual bad-faith behavior by future injurers, punishment in the second stage of enforcement is punitive in nature. The size of retaliation is fixed by a DSB arbitrator who makes sure that the retaliation amount is strictly (and substantially) higher than the victim’s expectation damages.

Two other key ingredients additionally ensure the effectiveness of WTO* enforcement: One is collective enforcement, the other punishment escalation. In collective enforcement, the suspension of concessions leaves the bilateral realm of complainant-vs.-defendant. Instead, retaliation becomes an issue of concern to the entire WTO* membership. Complaining, affected and concerned parties alike pool their retaliation capacities to overcome the problem of constituting too small a market to cause noticeable pain to the perpetrator (Hudec 2002; Maggi 1999b; Pauwelyn 2000). Enforcement, hence, is a real “sanction” to speak of, since it is free from any rebalancing constraint, and does not bear a bilateral notion.

The second ingredient to lend substance to enforcement is an intricate escalation scheme geared at bringing the recalcitrant WTO* Member into compliance: The longer the obstinate injurer refuses to comply with the panel ruling, the more its punishment is ratcheted up. Collective retaliation claims by the membership grow at an increasing rate.\(^{565}\) At the end of the escalation scheme there are additional penalties in the form of the suspension of certain membership rights, such as the right to attend meetings, to use the DSM, or to receive technical assistance.\(^{566}\)

In summary, the WTO* – the WTO re-conceptualized by rational policymakers – needs procedures and rules for dealing with disputes. Contrary to what some WTO scholars suggest, a binding third-party adjudication, such as the DSB, is not epiphenomenal to a liability rule.\(^{567}\) Mechanisms of dispute settlement must be in place to deal with good-faith clashes (due to contractual ambiguity, interpretative problems, unintentional contract infringements) as well as bad-faith clashes (blatant disregard of WTO* rules for opportunistic gains). WTO* Members will thereby negotiate a two-tier system of treaty enforcement: The first protective belt is apt to deal with good-faith disputes. Remedies at

---

\(^{565}\) With every day the injurer fails to comply, victims’ expectation damages increase. The idea of the escalation scheme is to introduce a punitive element, whereby retaliation damages increase over and above the amount of expectation losses. Similar to interest rates, retaliation awards grow as long as the injurer stays non-compliant.

\(^{566}\) The positive aspects of the suspension of membership rights is that they do not entail negative trade effects. Experiences in that area have already been gathered in the IMF, the Montreal Protocol for the Protection of the Ozone Layer and the ILO (Charnovitz 2001; Lawrence 2003).

\(^{567}\) See the arguments of the “rebalancing” view of the WTO introduced in subsection 6.1.1 above.
this stage are weak and strictly commensurate to the damage caused. The second layer of protection is punitive and collective in nature. Ultimately, contract enforcement must protect against extra-contractual behavior, not invite it. Given that there is always an efficient safety valve in place for benevolent injurers, WTO* enforcement must protect the contract with high penalties.

This two-step escalation scheme of enforcement is both fair and effective. It is fair since it protects efficient ex post escape, and does not discourage interpretative disputes over textual ambiguities, contractual ambivalence and contractual flaws. It is effective since it insures victims’ expectancy under the bargain, guarantees the re-establishment of the balance of concessions and punishes unambiguous bad-faith behavior. For ill-meaning injurers, the incentive structure of the game has changed in comparison to the real-life WTO: Violation of the Agreement is no longer a substitute for using contractual flexibility mechanisms. It is no longer a comfortable fallback that injurers can resort to at any time. Perpetrating extra-contractual bad-faith behavior is now a painful experience for injurers. The rational anticipation of functioning punitive retaliation makes prospective injurers amenable to engaging in serious renegotiations and striving for mutually agreed solutions. In light of coercive punishments, letting renegotiations break down is no longer an attractive option.

7.5 The WTO* as an efficient “breach” contract: A better trade agreement?

Based on the findings gathered from previous chapters, Chapter 7 conducted a hypothetical bargain analysis of the WTO, giving rise to the concept of a theoretical benchmark contract, the WTO*. This contract is an image of what the WTO could look like if it were organized along the principles of the efficient “breach” contract. We opened this chapter with a series of questions concerning the WTO*’s governance structure. Let us provide a summary of findings:

The hypothetical bargain analysis of the WTO* contract depicts a bare-bone contract, consisting of the definition and characterization of the exchanged (substantive and auxiliary) entitlements, as well as of one single intra-contractual protection rule for each entitlement. These entitlements are completely non-contingent in the face of the unbridgeable incertitude the contract is cumbered with.
An efficient “breach” of trade commitments emerges endogenously when rational policymakers negotiate the trade agreement. Contractual flexibility mechanisms in the WTO* capture all gains from non-performance: The market access entitlement is best protected by an unconditional liability rule of default and backed by expectation damages. Expectation damages are payable in additional tariff commitments only. They replicate full compliance of the injurer and put the victim in as good a position as if the injuring country had performed. Non-trade entitlements are protected by default rules of inalienability, liquidated damages, or renegotiation, depending on the damage that ex post discretion provokes in victim countries.

Enforcement has a dual role of addressing good-faith disputes triggered by inadvertent contractual gaps, and of safeguarding compliance with the rules of the game. The precursory dispute stage is geared towards precipitating an amicable resolution of the conflict, and features no punitive element. The subsequent punishment stage is aimed at deterring bad-faith injurers from defecting and causing inefficiencies. An escalatory “sanctions” scheme includes collective suspension of concessions or other obligations, which can be enriched and ratcheted up by the temporary withdrawal of some of the injurer’s membership rights.

To conclude this chapter, we briefly compare the current WTO governance structure with that of the hypothetical WTO*, both in terms of organization (subsection 7.5.1) and efficiency outcomes (subsection 7.5.2). This is followed by a discussion of the stakeholders for whom the benchmark treaty WTO* constitutes the “better” contract (subsection 7.5.3).

7.5.1 How do the WTO and the WTO* differ?

With respect to trade policy flexibility and enforcement the current-day WTO and the hypothetical WTO* share important elements: Both institutional frameworks demand strict compliance with the contractual rules and with dispute panel rulings. Both favor mutually agreed solutions over official third-party arbitration; both contracts prefer tariff compensation over retaliation. Yet a consequential difference is that the system of non-performance in the WTO* works. The governance structure of the hypothetical bargain is such that it sets the right incentives in order to put into effect these contractual stipulations. The institutional design of the WTO* is an incentive compatible arrangement which manifests itself in four important organizational differences between what the WTO should look like and what it looks like today:
1. **The WTO* is simpler.** The provisions of the WTO* contract are less circumstantial, yet not less precise than what we find in the WTO today. The WTO* dispenses with anticipating regret contingencies, prohibiting policy instruments, and drafting elaborate contingency measures. Instead, the WTO* makes do with a definition and explanation of the relevant entitlements, one default rule of protection per entitlement and a set of enforcement rules and procedures. Less treaty language and fewer contractual provisions create less room for ambiguity, informal opt-outs, legal loopholes – and therewith fewer possibilities for opportunistic maneuvering.

2. **The WTO* features different rules of default.** A striking difference between the trade policy flexibility design of WTO and WTO* lies in the definition of what constitutes *intra*-contractual behavior. This is particularly relevant with respect to rules of default. As analyzed in subsection 5.1.1 and section 5.2 above, the WTO’s *de iure* protection of market access- and non-trade entitlements is organized by means of renegotiation rules (Art. XXVIII GATT, Art. XXI GATS, Art. X WTO Charter). Our hypothetical bargain analysis showed that it is more efficient for self-interested policymakers to grant a rule of unconditional liability (backed by expectation remedies) to market access entitlements, and to protect multilateral entitlements with a property rule (featuring negotiated remedies), a rule of inalienability, or liquidated damages, depending on the nature of the multilateral entitlement at hand. When it comes to cases where *ex post* discretion is immoral, a strict rule of inalienability should be formulated.

The contractual default rules in the hypothetical WTO* are apt to capture all room for regret, while discouraging opportunistic opt-out. As elaborated in detail in subsection 5.4.1 the same is *not* true for the contemporary system of “breach” and remedy in the WTO.

3. **Nature and calculation of damages are different.** As discussed in subsection 5.4.1, the interpretation of “substantially equivalent” damages by WTO arbitrators is strictly under-compensatory. *Equivalent* damages payable to the victim pursuant to a liability-type opt-out must be interpreted as expectation gains foregone. Hence the expectation damage measure, which puts the victim in as good a position as had the injurer performed, is the only remedy to safeguard efficient “breach” and adequate *ex ante* commitments. This not only requires a novel calculation method on part of arbitrators, but also gives procedural precedence to the countermeasure of tariff compensation over that of retaliation.
4. Violation of the Agreement is not a substitute for using *intra-contractual flexibility anymore*. A consequential difference between WTO and WTO* is that the current arrangement fails to discriminate effectively between permissible non-performance (flexibility tools) and a flat-out violation of the contract. Some Members experiencing regret are barred from acting in good faith due to rigidities in the current trade policy flexibility regime, while violation is often the “cheapest” non-performance solution for injurers. *De iure* flexibility clauses and violation are thus *substitute* mechanisms of escape. This is illogical and dangerous for the entire system.

The hypothetical WTO*, on the other hand, uses enforcement provisions as a second line of contractual defense; it turns trade policy flexibility and enforcement into complementary instruments. The WTO* displays a two-tier enforcement system that effectively protects its Members against ill-meant *extra-contractual* behavior. Given that there are efficient default rules of flexibility in place, and with contractual misunderstandings smoothed out by a preceding dispute phase, the second stage of WTO enforcement protects the contract with coercive collective penalties – without doing harm to the world trading system.

7.5.2 Efficiency edge of the WTO* over the WTO

What would be the likely consequences if the world trading system were organized along the lines of the hypothetical WTO*? We foresee four important advantages:

1. **Clear separation between good-faith and bad-faith; opportunism curbed.** In section 3.5, we inched our way towards a theory of disputes in incomplete contracts. We called attention to the important difference between good-faith and bad-faith disputes. Acts of good faith can be efficiency-enhancing and should be supported by any efficient contract. Bad-faith behavior is opportunistic, and thus by definition welfare-depreciating, and must be forestalled by robust enforcement mechanisms. The difficulty in drawing the line between good- and bad-faith clashes lies not in the presence of contractual incertitude but in deficiently drafted contracts. Section 3.5 distinguished four types of contracting errors: ambiguous and ambivalent language, insufficient language, rigidity, and suboptimal remedies.

The hypothetical WTO*, now, is free from the problems of rigidity and suboptimal remedies: Every *intra-contractual* use of flexibility mechanisms is *per se* well-intentioned, and every well-intentioned escape is possible. Violation of the Agreement is thus either inflicted by contracting errors (ambiguities, ambivalence or insufficient
language) or driven by bad faith. The dispute stage of WTO* enforcement is able to identify and weed out good-faith disputes generated by contracting flaws, and passes malevolent arguments on to the punishment stage.

This is an important advantage over the current system of WTO enforcement. A clear separation between good intentions and bad-faith disputes fully enables and justifies punitive sanctions for recalcitrant WTO Members.668 Coercive punishments, in turn, deter Members from circumventing the contract by using extra-contractual, informal opt-out avenues. Curbing the opportunistic behavior of injurers is hence an important improvement on the current WTO system.

2. More stability, more compliance. As was shown in the course of the discussion of Herzing’s (2005) model in subsection 7.2.2, the introduction of optimal flexibility mechanisms in trade agreements based on market access externalities substantially increases the contract’s resistance to a wide range of exogenous political support shocks. Member exit (be it partial or full) and complete breakdown of the international trading system are less likely to occur under the efficient “breach” principle.

At the same time, an optimal system of trade policy flexibility like the hypothetical WTO* leads to more compliance with the letter of the Agreement (read: performance as promised). The prospect of punitive damages for bad-faith behavior should deter Members from deliberately defecting from the Agreement. Increased compliance should be the natural consequence.669

3. More cooperation, deeper integration. The WTO* is an efficient “breach” contract and as such largely replicates the outcome of the unachievable Pareto-efficient complete contingent contract. The prospect of increased compliance through the inclusion of the optimal escape scheme in a trade agreement will induce all contracting policymakers to consent to the politically optimal level of trade liberalization in the initial negotiations, as was shown in subsections 7.2.2 and 7.2.3. This result, generated in a simple tariff-

668 It is likely that the current punishment for violation of the WTO Agreement is lax (“toothless”), partly because the WTO framers were cognizant of their insufficient handling of the contractual incompleteness. The founding fathers of the WTO may have realized that the current governance structure was insufficiently selective between good-faith and bad-faith clashes. Afraid to enrage well-intentioned injurers by punishing them over and above the trade damage they caused, the drafters opted for the stopgap solution of commensurate punishment for extra-contractual behavior. To substantiate this conjecture, more historical research would have to be conducted, however.

669 This finding is in disagreement with various IR scholars (e.g. Goldstein, et al. 2000; Smith 2000; Yarbrough and Yarbrough 1997, 1987; Goldstein and Martin 2000), who contend that there exists a “domestic political trade-off between treaty compliance and policy discretion” (Smith 2000, p. 138).
setting model, can quite possibly be generalized to include all entitlements traded in the WTO(*).

Policymakers participating in the initial negotiations do not risk much but gain a lot from including additional and more sensitive issue areas (i.e. sectors that are more likely to be affected by political support shocks) into the Agreement. A broader range of trade topics and industries can thus be included into the efficient trade deal of the WTO*.

4. Fewer disputes, increased trust in the system, freer trade and more expected welfare. A trade agreement which is selective between good- and bad-faith behavior, which is more stable, fosters broader and deeper cooperation between trading partners, and induces increased compliance with the rules, will necessarily yield fewer disputes, more trust, higher levels of welfare and freer trade. We explain each in turn:

The WTO* allows its Members to engage in efficient “breach” whatever the underlying contingency. This reduces the DSB case load, since previous opt-outs in the form of violation-cum-retaliation can now be dealt with intra-contractually. So far, violation of the Agreement is being abused as an informal flexibility mechanism for reasons of opportunistic guile, and for want of an intra-contractual alternative. The WTO*, however, dispenses with such behavior. Only interpretative problems in connection with contracting errors can be expected to be brought before the DSB. Since the WTO*’s structure is much simpler than the current-day Agreement (no contingent language, negative integration provisions or contingency measures), the remaining interpretative problems will be less numerous. The workload of the DSB is reduced.

Trust in the system is generated by the injurers’ certainty of being able to seize regret contingencies whenever they occur. Victims can be sure they will not suffer from the exercise of trade policy flexibility mechanisms. Disputes are less frequent, and flat-out non-cooperative behavior rare. The WTO* surpasses the current contractual framework in all these aspects. A virtuous circle is initiated: Trust in the system is rewarded by freer trade (more depth and breadth of ex ante cooperation commitments). A higher ex ante promise leads to more per-period payoff, which increases the expected welfare that parties can hope to reap, which again fosters Members’ trust in the trading order.
7 The WTO* as an efficient “breach” contract

7.5.3 The WTO*: A “better” contract?

Does this mean that the hypothetical WTO* constitutes an objectively “good” contract? There are two points to make: First, the WTO* is the most efficient contract for self-interested and reasonably rational policymakers, given the contractual circumstances, constraints and trade-offs that the bargaining context entails. The WTO* is the first-best treaty that policymakers can achieve. This makes it a “good” contract from the selfish point of view of the negotiating policymakers.

Second, whether the outcome of the WTO* can be referred to as good, fair, equitable or just lies in the eye of the beholder. These terms are not only inherently subjective, but also relative concepts, and this is not the place to delve into a discussion of justice or fairness.\(^{570}\) If by “good contract” the observer means to say that the WTO* should be conducive to the general global welfare of non-signatories – such as consumers and producers worldwide – then we must leave the design of a “good contract” to future research. We can only offer conjectures here: The fact that the hypothetical WTO* makes the contract more stable, more reliable, more incentive compatible and induces country governments to be more compliant, cheat less, liberalize more industries, make bigger tariff cuts and include more issue areas in the trade deal, would suggest that the WTO* contract should at least not be judged as right-out pernicious to global welfare.

However, a couple of caveats have to be taken into account: For one, in a contract like the WTO*, where governments can step back from nearly any entitlement they have previously committed to, the predictability and stability of market access, and a level playing field of competition for all non-signatory actors, must suffer. The exercise of ex post discretion, and the occurrence of possible tariff retaliations, severely affect the international trading activity. Consumers and producers cannot easily foresee which political support contingencies will prompt self-interested policymakers to withdraw from which commitments. This volatility in the global trading order may induce the private sector to invest less, trade less or consume less than would be efficient.

---

\(^{570}\) The benchmark for our construction of the hypothetical WTO* was efficiency, as seen from the perspective of self-interested policymakers. If this concept of Pareto-efficiency clashes with someone’s notions of fairness and justice, they may overlook that both efficiency and fairness are best understood as relative, rather than absolute concepts in this situation. It is the power of the factual – unalterable contextual real-life constraints – that sets the limits for what is ultimately achievable, and hence efficient and just. If one accepts that the ideals of fairness, equity, justice, etc. – just like the one of efficiency – have to obey a set of inevitable constraints, all of the former principles must collapse with the notion of efficiency. If one is not willing to discount the concepts of fairness, equity, or distributional justice for factual parameters, then s/he must accept that these ideals “must take a back seat” (Downs, et al. 1996, p. 386) in the face of real-life constraints, and agents’ selfish preferences in particular.
Also, the optimal *ex ante* trade liberalization levels for policymakers are potentially a far cry from what would constitute optimal general-welfare levels of commitment. Consumers may see trade liberalization concessions from political trade agreements as overzealous or insufficient, depending on the nature of their utility functions.\(^{571}\) Finally, any sort of remedy payment in the form of tariff compensation, or punishment in the form of retaliation taken in response to a trade policy flexibility measure, does not benefit the harmed export sector (nor, for that matter, the sector which is liberalized as compensation). Hence, any rebalancing of commitments occurs with an eye to the concerned policymakers, but not necessarily for the benefit of the involved industries.\(^{572}\)

In summary, non-signatory stakeholders (especially consumers and exporters) generally benefit from the improved contractual framework of the WTO*. Still, more research should be conducted in this area.

### 7.6 Post scriptum: Compliance vs. rebalancing – redux

Chapter 6 exclusively dealt with the compliance-vs.-rebalancing debate in WTO scholarship. We found that the alleged discourse on the object and purpose of WTO enforcement was actually a disguised debate on the nature of the WTO contract and the organization of trade policy flexibility and enforcement.

The analysis of the compliance/rebalancing controversy showed that the two theories do not match well, since rebalancing and compliance scholars are in fact describing two different contracts:\(^{573}\) The rebalancing camp characterizes a pure market access exchange contract and reflects on *intra*-contractual safety valve mechanisms. Compliance proponents portray the WTO as a full-blown multi-entitlement treaty (even as a

\(^{571}\) This, however, equally goes for the current WTO regime, where trade protection is even higher than in the hypothetical WTO*.

\(^{572}\) This flaw is also inherent in the contemporary WTO system, and can only be overcome by means of monetary compensation dispensed to the victim industries.

\(^{573}\) To recapitulate: The rebalancing approach to the WTO reduces the contract to reciprocal market access commitments, which are *de facto* protected by a pure liability rule. Every signatory can opt out of the WTO Agreement anytime and under any circumstance, given the country pays its compensation, or at least does not obstruct retaliatory self-help measures taken by the victim. We found the rebalancing approach to be agnostic over *extra*-contractual rules of enforcement. Rebalancing proponents see no systemic difference between *intra*-contractual trade policy flexibility and dispute settlement/enforcement. The compliance approach, on the other side, fails to distinguish between different kinds of entitlements. In the eyes of compliance proponents *every* contractual right and obligation is protected by a property rule of default (if it is not explicitly listed in the text as an exception or contingency measure). Everything else amounts to a violation of public international law. It is the explicit task of the DSU to coerce the violator into compliance with the rules of the game and the rulings of dispute panels and the AB.
constitution) and are primarily concerned with the international legal bindingness of panel rulings and the enforcement of WTO violations in general. The only direct point of contention between the two views is whether the market access entitlement is protected by a liability or a property rule of default.

Having assessed in Chapter 7 how the WTO would look if it were an EBC, we feel ready to effectively broker between the two rivaling views. By separating issues of trade policy flexibility from those of enforcement, and by isolating the discussion of non-trade entitlements from that of the market access entitlement, we come to the surprising conclusion that both views have it largely right: Both rebalancing and compliance are the fundamental pillars of the WTO contract, whereby rebalancing is the name of the game for the organization of intra-contractual flexibility, while compliance is the lynchpin of extra-contractual enforcement.

Just as compliance advocates claim, observance of the WTO* rules of the game and compliance with dispute panel rulings are the overarching imperatives. Any steadfast refusal to comply with a panel ruling and any bad-faith violation of the WTO* contract is punished under pain of punitive damages in the second phase of the two-tier enforcement scheme. Rebalancing is the other crucial ingredient. When it comes to rules of trade policy flexibility, the efficient “breach” principle is the fundamental precept of entitlement protection. Injurers are forced to internalize all damages caused by their ex post escape. Escape, be it in the form of a liability-type opt-out or in the form of renegotiation, is encouraged for most of the entitlements. Contrary to compliance gospel, the market access entitlement is protected by an unconditional liability rule of default, because the transaction costs involved in arbitration Pareto-dominate those connected to renegotiation. Contrary to rebalancing wisdom, however, most other entitlements are best protected by a rule of renegotiation. Also, contrary to what rebalancing proponents claim, an impartial and competent dispute settlement body is an indispensable constituent of the WTO. It has to act as an arbitrator in instances of contract escape, but also as an adjudicator, gap-filler and calculator of collective “sanctions” where disputes arise.

In summary, the hypothetical WTO* is both a compliance- and a rebalancing contract.

574 It is permissible in all those instances where post-contractual default is not immoral and welfare-enhancing.
8 Towards an efficient “breach” contract: An agenda for reform

Having assessed in the previous chapter what the WTO would look like as an EBC and how reasonably rational governments would optimally design a system of contractual non-performance, we can now proceed to suggesting an agenda for reform. This chapter provides a conclusion to the entire study, precisely because without the detailed analysis of previous chapters this reform agenda would be just another ad hoc change request to join the ranks of the many that WTO scholarship has seen so far. The following reform proposals are built on firm foundations, since they lay out a politically realistic (read: incentive compatible), systemically viable and efficient system of flexibility and enforcement in the WTO.

The following presents ways in which the contemporary WTO could be shifted onto the trajectory of the achievable first-best – the efficient “breach” contract. The results of Chapter 7, which characterized the hypothetical bargain of the WTO*, can be operationalized and put into effect by changing the text of the current contract. We have worked out a “shortlist” and a richer catalog of reform suggestions: The shortlist (section 8.1) contains those improvements towards an efficient “breach” contract which are most urgent and/or possess the largest reform leverage. The extensive catalog (section 8.2) sketches comprehensive and long-term changes and amendments required to develop a coherent and sustainable institutional system of non-performance in the WTO. Section 8.3 concludes and presents a research outlook.
8 Agenda for reform

8.1 The shortlist of reform

Chapter 7 characterized how the WTO would look, if it were rigorously geared towards efficiency from the point of view of self-interested policymakers: Each entitlement would be optimally indefinite, and protected by a single operable rule of default and by the DSU as a second line of defense. We are aware that turning the current WTO into an EBC is a long-term (and probably very idealistic) endeavor. This section presents the three most pressing reform steps which can deliver the largest “reform traction”. These steps are:

(i) The institution of a default rule of liability for the market access entitlement;
(ii) The implementation of a common default rule for all multiple non-trade entitlements;
(iii) A reform of the current system of enforcement.

With this shortlist, which involves Art. XIX GATT (and Art. X GATS), a novel Art. X\textit{bis} in the WTO Agreement, and Art. 22 DSU, the WTO should establish some progress towards an efficient system of non-performance.

8.1.1 Establish a revised Art. XIX GATT\textsuperscript{575}

A crucial first step in the reform proposal is to refashion Art. XIX GATT, which currently serves as a contingency measure for economic emergencies, into Art. XIX(rev), the default rule for the market access entitlement. Modeled on the specifications of the WTO* (section 7.2 above) the liability rule of flexibility should display the following characteristics:

1. No preconditions, and a minimum of constraints on the application scope. In an incomplete contract a default rule of flexibility should not be impeded by a high level of conditionality (cf. subsection 7.2.4.1). Every injurer should be able to enact the revamped Art. XIX(rev) GATT for any reason and at any time, inasmuch as the country is willing to grant tariff compensation amounting to victims’ expectancy. The use of the DR is to be regulated by a minimum of constraining provisions. Apart from the obligation that protectionism be confined to an increase in tariff instruments only, few conditions apply to the enactment of Art. XIX(rev). Resort to ex post escape can happen

\textsuperscript{575} Art. X GATS should be reformed in a similar way for the trade in services. For reasons of parsimony we only discuss the reform of trade policy flexibility here for trade in goods.
at short notice, be discriminatory in nature, be enacted on a temporary or medium term and be invoked following any fathomable contingency.

2. Compensation, not retaliation, as the remedy of choice. For reasons of efficiency, remedies payable to victim(s) are to be made in the form of tariff compensation (see subsection 7.2.4.1). Therefore, any reference to the suspension of concessions, such as in Art. XIX.3(a) of the contemporary GATT, is to be eliminated.

3. Binding arbitration. An important component of Art. XIX(rev) GATT is the introduction of mandatory arbitration that sets in if no swift mutually agreed solution is reached. The WTO arbitration authority, which may be composed of dispute panelists (although acts of opt-out pursuant to Art. XIX(rev) are not to be conceived of as disputes), must be impartial, competent and enjoy general authority.

4. Insert a mandate for expectation damages. Art. XIX.3(a) of the contemporary GATT speaks of “substantially equivalent” damages that the victim of a safeguard measure can assert. As explained in subsection 5.4.1, WTO arbitrators have wrongly interpreted this term as prospective direct trade damages, i.e. remedies roughly satisfying the reliance damage definition. To avoid substantial flaws of this sort, and to institute the incentive compatible remedy of expectation damages (cf. subsection 7.2.3 above), Art. XIX(rev) GATT must put into unambiguous terms that the benchmark for remedies is the expectation damage measure only.

Art. XIX GATT revised according to the above principles could read:

Article XIX(rev) (Special Action on Imports of Particular Products)

1. A Member shall be free to temporarily withdraw or modify concessions in respect of any product by resorting to a tariff measure.

2. Before a Member can take the special action pursuant to provisions of Paragraph 1 of this Article, it shall give notice in writing to the Members as far in advance as may be practicable and shall provide adequate opportunity for prior consultations with those Members having a substantial interest as exporters of the product concerned. In critical circumstances where delay would cause damage difficult to repair, action

---

576 Some passages follow Roitinger’s suggested diction (Roitinger 2004, pp. 194).
577 Please note that the following suggestions do not fulfill the formal requirements of an official amendment proposal. Our representation is driven by reading convenience, not by legalistic revision criteria.
under Paragraph 1 of this Article may be taken provisionally without prior consultation, on the condition that consultations shall be effected immediately after taking such action.

3. If agreement among the interested Members with respect to the action is not reached, the Member which proposes to take or continue the action shall, nevertheless, be free to do so, if it maintains a general level of reciprocal and mutually advantageous market access concessions not less favourable to trade than that provided for in this Agreement prior to such negotiations. To achieve this precept, the Member shall provide means of trade compensation which put the adversely affected Members in as good a position as had the Member applying the measure performed as promised. This includes nullification and impairment that has accrued since the date of the enactment of the measure in question. Compensation is mandatory and shall be consistent with the covered agreements.

4. (a) A standing arbitrator is established. Resort to arbitration can be requested by any Member involved whenever no agreement on trade compensation is reached between the Member applying the measure in question and the exporting parties which are affected by such a measure.

(b) The arbitrator grants tariff compensation awards according to Paragraph 3, Sentence 2, of this Article.

(c) The parties shall accept the arbitrator’s decision as final.

Art. XIX(rev) GATT allows WTO Members to capture any type of regret contingency. Injuring countries can act independently of the confining language of other GATT contingency measures. The term “temporarily” in paragraph 1 is chosen to distinguish Art. XIX(rev) from Art. XXVIII GATT, a relationship that will have to be clarified in the course of a broader revision of the GATT (more on that infra). Paragraph 2 of the modified Art. XIX(rev) GATT leaves room for WTO Members to reach mutually agreed solutions. In renouncing any timelines, and granting any involved Member the right of instantaneous appeal to arbitration, strategic hold-out games by victims or injurers during this negotiation phase are eliminated. Mandatory compensation amounting to (prospective) expectation damages indemnifies victims adequately for their harm.

578 The expression “arbitrator” shall be interpreted according to the specifications laid down in Art. 22 of the Understanding on Rules and Procedures Governing the Settlement of Disputes.

579 Note that Art. XIX(rev) GATT would replace the entire Agreement on Safeguards.
suffered (see paragraph 3 above). Compensation must be in compliance with the general rules of the WTO, but there is no obligation to offer compensatory liberalization on an MFN basis. Compensation is owed to the victim(s) of the measure only.580

8.1.2 Add Art. Xbis to the WTO Agreement

It is not easy to discern a clear default rule protecting the numerous multilateral entitlements in the current WTO system. As we discussed in subsection 5.4.2, it is not at all certain whether Art. X of the WTO Charter actually constitutes a rule of default, and if so, what it really means. Thus, step two of our reform agenda is to put in place an unambiguous default rule that covers all non-trade entitlements. In subsection 7.3.2 we showed that WTO Members should ideally examine every single multilateral entitlement for optimal protection, which can be a rule of inalienability, liquidated damages, or a property rule of renegotiation. However, cognizant of the fact that demanding an immediate enactment of this ideal design would overburden WTO Members in a first stage of WTO reforms, we propose a robust transitional solution. A novel Art. Xbis is to be added to the WTO Charter. That novel article stipulates a general default rule of renegotiation. Every Member which feels concerned is to veto temporary changes of multilateral entitlements.581 Hence, Art. Xbis of the WTO Agreement could read like this:

Art. Xbis (Modification requests and emergency escape measures)

1. Any Member may, by negotiation and agreement with every Member affected or concerned, modify or withdraw a concession of this Agreement or the Multilateral Trade Agreements in Annexes 1, 2, and 3. Exempt from this provision are those concessions whose modifications and emergency escape measures are regulated directly in Annexes 1A and 1B.

580 We believe that authors like Anderson (2002), Pauwelyn (2000), Charnovitz (2001), Schropp (2005) and Roitinger (2004) are wrong in arguing that the wording “Compensation […], if granted, shall be consistent with the covered agreements” (Art. 22.1(3) DSU) means that compensation must be offered on a non-discriminatory MFN basis. The covered Agreements often provide for exceptions to the MFN principle (e.g. Art. XXIV GATT, the enabling clause, Art. 22.2 DSU), yet these provisions are equally consistent with the covered Agreements. This sentence means that compensation must be in line with the general rules of the game, and it excludes prohibited behavior such as quantitative restrictions, VERs, OMAs or export subsidies.

581 Anticipatory and sophisticated Members will realize which multilateral entitlements are rationally protected by a rigid rule of inalienability, and will veto any corresponding request to escape, even if they do not feel directly affected by the requested measure. Bagatelle opt-outs (ideally protected by a liquidated damages rule), on the other hand, will not concern them.
2. In such negotiations and agreement, which may include provision for compensatory adjustment with respect to tariff liberalization in goods and services, the Members concerned shall endeavour to maintain a general level of reciprocal and mutually advantageous concessions not less favourable to the international trading system than provided for in this Agreement and its Annexes prior to such negotiations.

3. Any Member rejecting a modification or emergency escape request mentioned under Paragraph 1 of this Article shall deposit an instrument of rejection with the Director-General of the WTO within a period of 90 days after the beginning of the initial negotiations.

4. Modification and emergency escape requests of a Plurilateral Trade Agreement shall be governed by the provisions of that Agreement.

The newly added Art. Xbis of the WTO Agreement institutes a general rule of renegotiation for any entitlement, except for the market access entitlement, whose protection is regulated separately in the GATT/GATS (see paragraph 1, sentence 2). The Member requesting escape has 90 days to come to terms with all those signatories which are affected by, or concerned about, the *ex post* escape request. Compensation offers by injuring Members are not limited to tariff measures, but can offer higher commitments in other obligations (e.g. higher minimum standards), as the wording in paragraph 2 suggests ("may include..."). Note that explicit consensus is not required for the measure to be enacted; a Member refusing the outcome of the renegotiations must officially submit a letter of rejection (cf. paragraph 3). This way, silent acquiescence can be achieved for minor cases of non-performance.

### 8.1.3 Revise Art. 22 DSU

The third pillar of the reform proposal is a reorganization of the WTO enforcement regime. Subsection 5.4.1.3 discussed the problems inherent in the contemporary WTO system of enforcement, while section 7.4 laid out in detail how enforcement should ideally be organized in the hypothetical WTO*. Our reform proposal is straightforward. Fortunately, the DSU deals with issues of enforcement in Article 22 DSU only. A revised version of that article should display the following characteristic:

**Introduction of hierarchy and chronology of enforcement instruments.** Enforcement under the current Art. 22 DSU displays a chronology of mutually accepted compensation, followed by the complainant’s retaliation request, and (in case of objections against the requested retaliation schedule by the defendant) arbitration.
Chronology, however, is not equivalent to hierarchy. Under the current DSU regime defendants have little interest in either a mutually agreed solution or in serious compensation negotiations (see our discussion in subsection 5.4.1.3, and Schropp 2005, section 3.4). Thus, the defendants’ incentive structure has to be changed. This is achieved by a two-step procedure: As laid out in section 7.4 above, the first stage of enforcement (the dispute stage) is driven by the goal of achieving a mutually agreed solution after the panel report is circulated, or at least an amicable resolution of the dispute. Therefore, if negotiations over a mutually agreed solution break down after a period of 30 days, mandatory arbitration sets in. At this first stage, the arbitrator is confined to awarding tariff compensation amounting to full expectation damages to the complaining party, which the defendant is encouraged to accept.

If the defending party stays recalcitrant, the second stage of enforcement (the punishment stage) sets in. The arbitrator is authorized to award punitive damages. How best to bring the recalcitrant injurer into compliance is the subject to an elaborate escalation scheme, which is to be designed at a later stage of the larger reform agenda (see infra). This two-prong enforcement scheme gives effect to the chronology of compensation and retaliation by establishing an incentive hierarchy of instruments.

An overhauled Art. 22(rev) DSU could read like this:

**Art. 22(rev) (Compensation and Suspension of Concessions)**

1. Compensation and the suspension of concessions or other obligations are temporary measures available in the event that the recommendations and rulings are not implemented within a reasonable period of time. However, neither compensation nor the suspension of concessions or other obligations is preferred to full implementation of a recommendation to bring a measure into conformity with the covered agreements. Compensation is mandatory and, if granted, shall be consistent with the covered agreements.

2. If the Member concerned fails to bring the measure found to be inconsistent with a covered agreement into compliance therewith or otherwise comply with the recommendations and rulings within the reasonable period of time determined pursuant to paragraph 3 of Article 21, such Member shall, if so requested, and no

---

582 Art. 22(rev) DSU in parts uses the same wording as the original Art. 22 DSU. For the ease of reading, we underline novel wording to distinguish it from the current language.
later than the expiry of the reasonable period of time, enter into negotiations with any party having invoked the dispute settlement procedures, with a view to developing mutually acceptable compensation. If no satisfactory compensation has been agreed within 30 days after the date of expiry of the reasonable period of time, any party having invoked the dispute settlement procedures shall refer the matter of compensation to arbitration. Such arbitration shall be carried out as determined in paragraph 3 of this Article.

3. (a) Arbitration shall be carried out by the original panel, if panel members are available, or by an arbitrator appointed by the Director-General and shall be completed within 60 days after the date of expiry of the reasonable period of time.

(b) The arbitrator shall present to the Member concerned a compensation package taking into consideration voluntary compensation concession offers granted under letter (d) of this Article.

(c) The arbitrator shall not examine the nature of compensation but shall determine whether the level of such trade compensation puts the adversely affected Members in at least as good a position as if the Member applying the measure had performed as promised. This includes nullification and impairment that has accrued since the date of request for consultations, as well as litigation costs incurred, and benefits foregone due to the violating measure. Punitive compensation awards that put the Member applying the measure into a worse position than had it performed shall be avoided.

(d) In considering compensation, the Member concerned may provide the DSB and the arbitrator with a compensation package apt to re-establish the general level of reciprocal and mutually advantageous market access concessions not less favourable to trade than that provided for in this Agreement prior to the nullification and impairment. Compensation offers must be in accordance with the rules and regulations of the covered Agreements.

(e) The parties shall accept the arbitrator’s decision as final and the parties concerned shall not seek a second arbitration.

4. If the Member concerned fails to implement the arbitrator’s decision and to bring the measure found to be inconsistent with a covered agreement into compliance therewith or otherwise to comply with the recommendations and rulings within, the matter shall again be referred to arbitration. The arbitrator authorizes the suspension of concessions or other obligations under the covered agreements. Concessions or other obligations shall not be suspended during the course of the arbitration.
5. (a) The arbitrator acting pursuant to paragraph 4 shall not examine the nature of the concessions or other obligations to be suspended but shall determine the adequate level of such suspension.

(b) The level of the suspension of concessions or other obligations authorized by the DSB shall be no less than what is needed to put adversely affected Members in at least as good a position as if the Member applying the measure had performed as promised.

(c) The parties shall accept the arbitrator’s decision as final and the parties concerned shall not seek a second arbitration.

6. [see Art. 22.3 DSU, except for the chapeau, which reads: In considering what concessions or other obligations to suspend, the arbitrator shall apply the following principles and procedures:] 

7. [see Art. 22.5 DSU]

8. (a) Suspension of concessions or other obligations authorized by the DSB can be enacted by the party having invoked the dispute settlement procedures, by Members affected by the measure in question, or by Members concerned with the functioning of the world trading system.

(b) If pursuant to the enactment of suspension of concessions or other obligations authorized by the DSB the Member concerned remains in default and declines to bring the measure found to be inconsistent with a covered agreement into compliance therewith or otherwise to comply with the recommendations and rulings within, the DSB by consensus can decide to withdraw certain membership rights.

Art. 22(rev) DSU sets in place an enforcement hierarchy that actually works: The dispute stage (paragraphs 1-4) leaves enough room to clarify ambiguities and to settle the dispute amicably, whilst the punishment stage sanctions the injuring Member’s recalcitrance and coerces it into swift compliance.

Paragraph 1 is to strengthen the procedural role of tariff compensation by making it mandatory (see Lindsey, et al. 1999; Schropp 2005, section 4.2). Whenever settlement negotiations break down, paragraph 3 prescribes mandatory arbitration, which is basically to proceed as in Art. XIX(rev) GATT (cf. discussion above). Paragraph 3(c) stipulates a “compensation floor” of expectation damages for the complainant by explicitly mentioning retroactivity, litigation costs and profits foregone. This compensation floor allows the arbitrator to include in its damage calculation the victims’
share of the injurer’s efficiency gains from non-performance. As discussed earlier (see footnote 563 and accompanying text), adding some of the spoils from non-performance is important for the protection of multilateral entitlements backed by a property rule. However, paragraph 3(c) also puts a “compensation cap” on the arbitrator’s quantification efforts: Punitive elements are contained in the dispute stage of enforcement. Paragraph 3(d) provides the defendant with the opportunity to voluntarily submit a pre-selected list of compensation commitments.583 When calculating the compensation awards, it is up to the arbitrator to decide whether to take into consideration the compensation package offered and whether trade liberalization in those respective industries is apt to suit the complainant’s rebalancing needs.

Paragraphs 4-8 of Art. 22(rev) DSU specify the punishment stage of enforcement. As opposed to the current design of Arts. 22.2 and 22.3 DSU, we do not interpose the – largely futile – step of a formal retaliation request to the DSB by the complainant. Instead, whenever the convicted defendant Member chooses to ignore its obligation in the dispute stage of enforcement, the arbitrator re-enters the scene immediately. The arbitrator sanctions retaliation awards and determines the mix of sectors to be targeted according to paragraph 6. Punitive damages are possible, with expectation damages as the remedy floor (see paragraph 5(b)). The collective nature of retaliation is regulated in paragraph 8(a), while further escalation is possible under paragraph 8(b).

8.2 Long-term reform proposals

The reform of only three articles in the WTO contract yields substantial mileage. It would pave most of the way towards turning the current system of trade policy flexibility and enforcement into an EBC. WTO Members experiencing a political support shock would prefer to use the readily available DR instead of most (in)formal tools of escape, and especially to the strategy of violation-cum-retaliation. Tariff compensation amounting to the expectation damage remedy would be instituted and would insure

583 The reader might be reminded of Lawrence’s suggestion of “contingent liberalization commitments” (see Lawrence 2003, chapter 5). The present proposal is of a different nature: We do not suggest that Members negotiate their pre-committed sectors upfront in a multilateral setting. Rather, a losing defendant can unilaterally submit an ad hoc list of “liberalizable” sectors to a specific complainant in the dispute at hand. Strategic and political deliberations on the part of the defendant under Lawrence’s and our own proposal will be quite different (not to speak of the practical and organizational consequences that incriminate the entire system under Lawrence’s proposal).
victims optimally. Stricter enforcement would deter opportunistic Members from overtly infringing upon the rules of the game.

However, more work needs to be done to eventually convert the current WTO contract into the WTO*, the contracting ideal of the EBC as portrayed in Chapter 7 above. Various grander-scale reform schemes need to be tackled, and remaining loopholes and inconsistencies of the current institutional framework must be closed in the long term. More research must flow into an endeavor of that dimension. We sketch the broad contours of a grander long-term reform agenda concerning the market access entitlements (subsection 8.2.1), multilateral entitlements (subsection 8.2.2) and WTO enforcement (subsection 8.2.3).

### 8.2.1 Reforming the protection of the market access entitlement

There are three long-term reform endeavors concerning to the market access entitlement:

(i) Make GATT and GATS leaner;
(ii) Overhaul the current system on antidumping and countervailing duties;
(iii) Clarify the relationship between Art. XIX(rev) and Art. XXVIII GATT, and Arts. X and XXI GATS, respectively.

(i) **Make GATT/GATS leaner.** The contemporary WTO is far from the optimally non-contingent contract that the hypothetical ideal of the WTO* stipulates. The GATT, with its appendant Agreements in particular, features various contingency measures and negative integration provisions that must be seen as insufficient and unwise attempts to fight contractual incompleteness (see discussion of at point 5 in subsection 5.4.1.1., and of point 1 of 7.5.1). Goods and services Agreements should become leaner. Trade negotiators should carefully examine whether in light of the unconditional rules of default, contingency measures like Arts. XXII, XVIII or XX GATT are still relevant. In the same vein, negative integration provisions as featured in the TBT, SPS or GPA should be tested for relevance.

(ii) **Revise AD and CvD regimes.** An important objective of the trading community should be a fundamental reform of the AD and CvD codes of the WTO. As stated in subsection 5.1.2.4 above, it seems evident that AD and CvD actions today are predominantly used as opt-out tools for protectionist reasons, and that the recourse to

---

584 This paragraph is adapted from Schropp (2005, section 4.1).
them as “unfair trade remedies” is a barely veiling fig leaf.\textsuperscript{585} In fact, AD and CvD can be accurately described as a property rule of protection granted to the injurer. It is time Members tackle the reform of the two codes in a manner that fits the WTO’s mandate and original intent.

A fundamental overhaul of AD and CvD regimes would consist of (a) an agreed-upon set of core definitions and principles,\textsuperscript{586} (b) a minimum of technical reforms,\textsuperscript{587} (c) a serious reduction of national discretion, a less lenient standard of deference, and a consideration of basic economic reasoning (cf. discussion in subsection 5.1.2.4, especially at footnote 397 and accompanying text). Since both AD and CvD have hardly any economic basis anyway,\textsuperscript{588} it would be best to eventually do away with the two codes completely and integrate AD and CvD into domestic competition law.\textsuperscript{589}

(iii) Clarify the relationship between Art. XIX(rev) and Art. XXVIII GATT. A final long-term project that needs to be tackled by trade negotiators once the Art. XIX GATT is revised is the relationship between the default rule and the provision of permanent modification of tariff schedules (Art. XXVIII GATT).\textsuperscript{590} Under a DR as proposed, an article on tariff renegotiation may be superfluous. WTO Members would benefit from a further examination of this point.

\textsuperscript{585} According to Messerlin (2000 at p. 163), less than 10 percent of all antidumping cases have even a slightest chance of being considered as “predatory-” or “strategic dumping”, the two only economically noxious categories (Willig 1998). According to the authors, the vast majority of AD measures is driven by protectionist motivation, and thus constitutes protectionist escape actions.

\textsuperscript{586} A binding set of definitions and principles could finally give answers to “trivial” questions like: What is dumping and why is it harmful? What are fundamental objectives and justifications for AD- and CvD action? What exactly constitutes remediable “unfair trade”? Nowhere in the AD Agreement or elsewhere in the GATT there is an attempt to define the basic economic and social precepts, principles and objectives that would justify AD action. There is mention of how “dumping” is determined, but not what exactly makes it pernicious. The negotiators simply seem to have assumed that every country will enact its proper AD rules and regulations on similar standards. This assumption, however, proved futile and the international trading system witnesses a severe regulatory and methodological heterogeneity in AD codes today. Given the very strong deferential standard of review of Art. 17.6 ADA, disagreement is predetermined.

\textsuperscript{587} See Schropp (2005) at footnote 59 for explanations.

\textsuperscript{588} “Although economic theory identifies a few plausible scenarios in which antidumping measures might enhance economic efficiency, the law remains altogether untailored to identifying them or limiting the use of antidumping measures to plausible cases of efficiency gain” (Sykes 1998, p. 2).

\textsuperscript{589} Antitrust agencies deal with anticompetitive and monopolistic tendencies on a daily basis and thus would appear to be the obvious candidates for assessing unfair practice in international trade (Barfield 2005; Messerlin 2000). Profound reform proposals of WTO laws on AD and CvD can be found in Lindsey and Ikenson (2003), Hoekman and Mavroidis (1996) or Bown (2002b). Most of these authors discuss the substitution of AD by antitrust regulation. Horlick and Palmer (2002) focus on the relationship between CvD and antitrust. Roitinger (2004 at pp. 193) offers a valuable literature review.

\textsuperscript{590} The same goes for the relationship between Arts. X and XXI of the GATS.
8.2.2 Reforming the protection of multilateral entitlements

With regard to the efficient protection of the various multilateral entitlements, we suggest two long-term reform topics. First, as explained in section 7.3 above, not every multilateral entitlement is best protected by the same default rule. It seems crucial that trade negotiators convene and analyze every multilateral entitlement for its optimal protection rule. As stated in Chapter 7, there are three ways of protecting those non-trade entitlements: by inalienability, by liquidated damages and by a rule of renegotiation. The new Art. Xbis added to the WTO Charter, albeit a robust interim solution, should be redrafted to comprise three sections, one for each default rule. Entitlements in the Agreements should then explicitly refer back to one of those three sections, so that every multilateral entitlement traded in the entire WTO is protected by a clear and unambiguous rule of default.

A second task for WTO negotiators is to clarify (define) the relationship between Art. X of the WTO Charter (dealing with amendments of the contract) and Art. Xbis of the Charter (dealing with modification and emergency measures). It is not yet clear how these two rules can be reconciled sensibly to function side by side.

8.2.3 Reforming the WTO enforcement regime

Once WTO enforcement has been changed by virtue of Art. 22(rev) DSU, in the long run WTO Members should convene to improve the protection of multilateral entitlements. Clearly, if an inalienable entitlement is infringed upon, the entire dispute phase of enforcement must be skipped. In the same vein, the way arbitrators deal with compensation calculations in the dispute stage, pursuant to infringement of a multilateral entitlement, needs to be clarified, elaborated and cast into treaty language. A second important long-term goal is the design of efficient guidelines for collective retaliation. These guidelines should include suggestions concerning which Members/Member groups should participate in the suspension of concessions, at what stage, and to what
8 Agenda for reform

extent. As a third step, signatories should draft guidelines that contain details of the escalation scheme design of retaliation and sanctions.

8.3 Final remarks and future research

This study provided an introduction to incomplete contracts, performed a comprehensive contract-theoretical analysis of the World Trade Organization, established the WTO as a necessarily incomplete contract, assessed the flaws of the contemporary system of entitlement protection, trade policy flexibility and enforcement in the international trading order, conducted a hypothetical bargain analysis of what the WTO could look like if it were properly designed, and suggested concrete steps of reform towards turning the WTO into an efficient “breach” contract.

It braced and clarified the intricate connection between contractual incompleteness, *intra*-contractual trade policy flexibility mechanisms, contract enforcement and Members’ willingness to commit to trade liberalization. We gave substance to the concept of efficient “breach” (better: efficient non-performance), and elaborated what that concept means in the international trade context and how it can be effectuated in the WTO.

We identified the weaknesses of the current regime of trade flexibility and enforcement in the WTO, including the systemic and dynamic consequences thereof. The resultant reform agenda is concrete, substantiated, politically realistic and systemically viable.

A lot of theoretical ground was covered in the course of this study. We are painfully aware that we often painted with a broad brush and at times slurred over issues which deserve greater attention and closer consideration. In order to focus on the most salient questions, and due to the natural limitations of such a study, we left open numerous issues. Among those there are four unresolved which must be passed on to WTO scholarship. These problems would make worthwhile subjects of future research:

---

591 We should think of an arrangement which stipulates that only affected Members retaliate collectively in a first stage. If that proves ineffective (or if only developing countries are the affected victims), large industrialized Members like the U.S., EU and Japan join the ranks of sanctioning countries. Details of these guidelines need to be contractually fixed in the long term.

592 Issues here include the progression of sanctions over time. The longer the offending Member stays recalcitrant the more penalties are ratcheted up. Also, the question must be tackled as to whether (and if so, when) essential membership rights should be suspended in order to bring the injurer into compliance.
1. The rationale for trade agreements. No examination of a contract is complete without a clear understanding of what drives signatories to strike an agreement in the first place. Our analysis in subsection 4.1.2.5 revealed that current trade scholarship is unfortunately far from answering comprehensively why sovereign countries engage in trade cooperation. Whilst economic, political and legal explanations seem able to elucidate facets of the cooperative motivation, scholarship has not captured the whole picture. More work needs to be done that produces testable results as to which of the numerous approaches (or which combination of approaches) best manages to explain countries’ resolve to engage in trade cooperation in the form of a written treaty. Cross-disciplinary work seems a fruitful and promising avenue for future research.

2. Non-trade entitlements. We perceive a relative neglect of non-market access entitlements in trade scholarship. Pauwelyn (2006) stated that the WTO contains more than just reciprocal market access entitlements. It was our aim to put this finding into practice by dismantling the WTO contract into its constituent entitlements. We showed that the WTO is best conceptualized as a bundle of single-issue contracts, consisting of the prominent market access entitlement but also of substantive minimum standard entitlements, as well as of various auxiliary entitlements. While much work in WTO scholarship has concentrated on the reciprocal exchange of market access and tariff liberalization concessions, it is remarkable that research on the multilateral entitlements exchanged in international trade agreements is scarce. Formal economic work in particular should be welcomed. It is high time for political economists to leave behind the prisoners’ dilemma set-up and to dedicate energy into alternative collective-action games of strategy in order to explain other facets of trade cooperation.

3. The nature of trade disputes. As discussed in section 3.5, any pundit whose object of research is “enforcement mechanisms” or “dispute settlement” ought to have a proper theory of disputes in stock. How can anybody propose novel tools and improved instruments for settling disputes, if s/he does not demonstrate an understanding for the very nature of disputes? For WTO scholars, learning more about how disputes arise and why they escalate should be an issue of great importance. We have proposed a few steps towards a more coherent theory of disputes; much more work needs to be done.

593 In fact, as argued in Chapter 6.3.2, the rebalancing approach to the WTO contract, which most economists subscribe to, is exclusively concerned with the market access entitlement.
4. How to measure expectation damages?\textsuperscript{594} Measuring expectation damages will be a tough call for WTO arbitrators: All trade agreements are inherently political deals designed and concluded by self-interested policymakers. Hence, the initial balance of concessions and the entire metric of the WTO is presumably profoundly political. Yet arbitrators can neither observe nor measure the political harm done to a victim government by its trade partner’s unilateral policy adjustment. How can arbitrators ever claim to be able to calculate, in tangible currency, political expectation damages? How then can they credibly assert to be able to re-establish a profoundly political balance of welfare if political expectation damages are unverifiable?

There is no immediate solution to this conundrum. A couple of comments are nevertheless in order: First, the restoration of expectation damages to the injured party is the correct instrument to measure remedies. It would be imprudent to condemn the endeavor of quantification and monetization of efficiency losses just because it seems challenging.\textsuperscript{595} Second, novel approaches to the quantification of victims’ harm have recently been undertaken (see Keck 2004; Breuss 2004; Josling 2004; Spamann 2006; Trachtman 2006). The quality of these attempts remains to be seen.

One promising way of calculating expectation damages could be to check the work of the neighboring fields of competition and antitrust. Competition authorities have a long track-record of applying coherent economic reasoning when defining relevant markets, market potentials and damages, including profits foregone (cf. Neven 2000). Another

\textsuperscript{594} This paragraph draws in parts on Schropp (2005, section 4.4).

\textsuperscript{595} As Mavroidis (2000 at p. 769) concurs: “The fact though that full recovery [i.e. expectation damages] is, in practice, sometimes hard to calculate, does not render the reparation exercise meaningless […] Although assessment of damages is the task of the judge, calculation of the damage is essentially a quantification exercise, that is, essentially the task of the economist.” Ultimately, the calculation of expectation losses is a technical and empirical task to be conducted by specialists. If WTO arbitrators are not prepared to execute economic analyses of this kind today, the WTO should stock up its economic competence rather than drop the task for reasons of practicality. The fact that the DSB could benefit more from the resources of the WTO Economic Research and Statistics Division and puts insufficient effort into economic and econometric reasoning is a different story, and one that could be remedied quite easily.
possibility is more heuristic in nature: WTO scholars could assess how Members actually proceed when engaging in tariff renegotiations under Art. XXVIII GATT. Although parties do not thereby issue exact figures and numbers, they nevertheless come up with tangible results which take into account the expected harm done by the protectionist impact of a tariff bound increase. The insights generated from studying how the involved WTO Members manage to generate objective outcomes of this subjective renegotiation task may well lead to a general quantification framework.
Bibliography


Bibliography


Bibliography


Masten, S. E. 1999. "Contractual Choice." In: B. Bouckaert and G. de Geest, 


Mavroidis, P. C. 1993. "Government Procurement Agreement; the Trondheim Case: 

Mavroidis, P. C. 2000. "Remedies in the WTO Legal System: Between a Rock and a 

Mimeo.


Forthcoming (on file with the author).

pp. 970-85.

Arbor, MI: University of Michigan Press, pp. 629.


Economics 117(2), pp. 571-608.

& Unwin Ltd.


Bibliography


Curriculum vitae

Simon Arnd Benedikt Schropp

Personal data

Date/Place of birth
July 18th 1976; Munich, Germany

Nationality
German

Education

PhD-coursework in international economics at the Graduate Institute of International Studies, Geneva (IUHEI)

Aug. 2002 – June 2004
Graduation from the School of International and Public Affairs (SIPA) at Columbia University, New York, with a Master of International Affairs (MIA)

May 2002
Received Master of Management awarded by the Community of European Management Schools (CEMS)

PhD-coursework in international political economy at the University of St. Gallen (HSG)

Graduation from University of St. Gallen (HSG), Switzerland, with a Master of Arts in Business Administration (lic.oec.)

Sep. 1986 – May 1995
Städtisches Käthe-Kollwitz-Gymnasium, Munich; held German Abitur

Experience

Mar. 2008
International trade analyst with Sidley Austin LLP, Geneva

Research fellow with the Swiss National Research Centre on Competence in Research: Challenges to Democracy in the 21st Century

Sep. 2006 – Apr. 2007
Full-time consultant for the Economic Research and Statistics Division of the World Trade Organization (WTO)

Feb. 2004 – Aug. 2004
Part-time consulting work for the International Centre for Trade and Sustainable Development (ICTSD), Geneva

BrainsToVentures AG, St. Gallen and Munich; Position: Head of Business Development

Social service at Dermatologische Klinik der TU München, Munich
### Scholarships

<table>
<thead>
<tr>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>Doctoral fellowship awarded by IUHEI Geneva</td>
</tr>
<tr>
<td>2002 – 2004</td>
<td>Scholarship program of the German National Academic Foundation (&quot;Studienstiftung des deutschen Volkes&quot;)</td>
</tr>
<tr>
<td>2002 – 2004</td>
<td>ERP-Scholarship by the German Ministry of Trade and Commerce; full scholarship tenable for a 2-year study in the USA</td>
</tr>
<tr>
<td>2000</td>
<td>Erasmus Scholarship</td>
</tr>
</tbody>
</table>

### Publications

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
<th>Details</th>
</tr>
</thead>
</table>