The London School of Economics and Political Science

Why Comply?

Experimental Evidence on the European Stability and Growth Pact's Incentives for Member States

Sarah-Esther Anneliese Willenbücher (née Ciaglia)

A thesis submitted to the European Institute of the London School of Economics for the degree of Doctor of Philosophy, London

27 July 2022

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others.

I declare that my thesis consists of 88,508 words.

Statement of use of third party for programming help

I confirm that the programming of the experiment has been implemented by Dr. Martin Starman. Niklas Zorell provided self-made graphics for the online game.

Statement of use of third party for editorial help

I confirm that my thesis was copy edited for conventions of language, spelling and grammar by Francesca Sweeney-Androulaki.

Abstract

The financial, economic and sovereign debt crisis in European Economic and Monetary Union (EMU) fundamentally challenged the purpose and functioning of the Stability and Growth Pact (SGP). Its purpose is to promote sustainable public finances to back the European Central Bank's goal for monetary stability. The crisis made visible and exacerbated what had already started before the crisis: member states' governments frequently breached the rules and ran high deficits, and many of these breaches remained without sanctions imposed by the other states in the Council. Given the high number of unsanctioned breaches, why would a member state comply at all?

The literature's accounts of a member state's 'decision' for a breach of the rules (such as a lack of economic or political capacity, a bad economic situation, and political power in the Council) can only explain part of it. This thesis suggests 'economic policy ideologies' to be the complementary piece to explain compliance. The SGP was designed in an ordoliberal spirit, but governments' convictions about what constitutes 'good' economic policy may run counter to it.

To explore this idea, this thesis contributes to the academic debate by providing a different theoretical, empirical and methodological perspectives: 1) theoretical through considering the SGP as a mechanism to govern a common-pool resource and with economic policy ideology to endorse the rules as 'appropriate'; 2) empirical through testing competing explanations to better understand how the SGP functions; and, 3) methodological through a novel type of a common-pool-resource experiment which I call 'cultivation game' to study the SGP, and which could also serve to study other cases of co-operation. The results show that economic policy ideologies shape compliance, in particular this is mediated through business cycle developments. This can inform ongoing debates about reforming the EMU's institutional design.

Table of contents

List of figures and tables

List of any material not bound in the copy of the thesis

D1 Datasets on raw data and final datasets

D2 R programming scripts on constructing the dataset and performing the analyses

D3 R model files where the results of the regression models are stored

This material is provided separately through LSE filedrop.

The experiment can be accessed here: http://personal.lse.ac.uk/ciaglia. The original version can be accessed when selecting any occupation other than state officials or politicians in the survey at the beginning. Else, one accesses the short version.

Abbreviations

Commission	European Commission
Council	Council of the European Union on Economic and Financial Affairs
EC	European Commission
ECB	European Central Bank
ECOFIN	Council of the European Union on Economic and Financial Affairs
EDP	Excessive Deficit Procedure
EMU	European Economic and Monetary Union
ESM	European Stability Mechanism
EU	European Union
GDP	Gross domestic product
NGEU	NextGenerationEU
OCA	Optimal currency area
SGP	Stability and Growth Pact
TEU	Treaty on the European Union
TFEU	Treaty on the Functioning of the European Union

Acknowledgements

First of all, I would like to thank Professor Kevin Featherstone and Professor Paul de Grauwe for being my supervisors and providing their deep insights to this thesis and their support to bringing this project to a close after several years of me also working on a job at the same time. I am also grateful for their kindness and humour. I am thankful to Professor Amy Verdun and Professor Iain Begg as examiners of this thesis and for contributing their dedication, insights and reflections to this thesis.

I am grateful for three years of financial support from a doctoral scholarship from the Konrad Adenauer Stiftung (KAS). Moreover, I am thankful to the professors and doctoral students in the KAS' Research Workshop on Social Market Economy for their insights and suggestions.

I am deeply grateful to my family and my friends who were convinced about this project all the way long and also endured this rather long journey with me. I am the first in the family to submit a PhD thesis and I am happy for their belief in this bunch of papers having any use other than my personal joy for research.

I am thankful to the staff and doctoral students at the European Institute for their comments on and curiosity in my project, notably to Professor Waltraud Schelkle as examiner of this project at an early stage who contributed intriguing questions and thoughts. For their dedication in and enthusiasm towards the topic, I thank Caro, Martin, Christoph, Julia, Jo and Max. I am also thankful to the participants of several conferences: at NIAS,[1] the Experimental Summer School in Oxford,[2] the ZEW Public Finance Conference,[3] and the IMEBESS in Florence;[4] and for the LSE Travel Fund and the European Institute for paying the travel costs. Moreover, I thank several other academics for their time and thoughts to advance specific aspects of the project, and the staff from the LSE Library, the Data protection officer, the LSE IMT Webeditors

1 European Union Decision-Making and Challenges to Economic and Financial Governance - Part 2 (co-sponsored by EURO-CEFG), 08-09 December 2015, Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS), Wassenaar (The Netherlands).

2 Essex/Nuffield CESS Experimental Summer School, 29 June to 10 July 2015, Nuffield College, Centre for Experimental Social Sciences (CESS), University of Oxford, Oxford (United Kingdom).

3 ZEW Public Finance Conference "The Future of Fiscal Coordination in Europe", 23-24 April 2018, Mannheim (Germany).

4 The Fifth International Meeting on Experimental and Behavioral Social Sciences (IMEBESS), 03-05 May 2018, European University Institute, Florence (Italy).

and the LSE Behavioural Lab for their technical support. I am particularly thankful for the support and hospitality of Professor Francisco Torres and Professor Annette Bongardt as well as Professor João Pereira dos Santos during my fieldwork in Lisbon.

Finally, I am thankful to my companion Chrysa for spending the time together and occasionally also discussing “more important things in life than the PhD”, as our common supervisor used to say when he saw us squeezed on the stairs to the entrance of the European Institute being immersed in our talks and thought experiments; for sharing my personal zest to clear misunderstandings to facilitate a proper debate, and the idea that everybody shall live how they wish to with respect for others.

1. Introduction

"Breaches of the 3% of GDP reference value for the government deficit have been repeated and persistent in some countries, leading to the conclusion that at least in these cases the implementation of the Pact has lacked sufficient rigour and political will."

European Central Bank 2008: 53

The financial, economic and sovereign debt crisis in European Economic and Monetary Union (EMU) fundamentally challenged the purpose and the functioning of the Stability and Growth Pact (SGP) to promote sustainable public finances. Several member states struggled with high deficits and low market confidence. High borrowing costs and the risk of sovereign default brought the euro area to the brink of breakup. Among other reasons, the blame was attributed to the SGP for not having been able to incentivise sustainable public budget positions before the crisis. A large body of literature concludes that the SGP has failed: on the one hand, because member states[5] frequently breached the budget rules (Eyraud et al. 2017, I refer to this as the 'national-level rule'); on the other hand, because peer pressure failed to incentivise compliance (Schuknecht 2005) and the Council[6] did not enforce the rules sufficiently (Hallerberg 2013, i.e., 'European-level rule'). Nonetheless, in spite of these weaknesses related to the SGP, some member states did comply with the budget rules and some breaching states have been sanctioned by the Council. From this perspective, compliance appears puzzling: why do member states comply at all?

The crisis also made apparent a fundamental misunderstanding about the purpose and usefulness of the SGP among member states. Differences in perspectives referred to the reasons of and the response to the crisis: whether origins of the crisis were outside the scope of the government or to some extent also homegrown; whether

5 I use the term 'member states' to describe decisions by a range of public actors in a member state on both the general government budget and on sanctioning other member states in the ECOFIN Council. While strictly speaking the actors that decide on a breach at the national level (parliament, or in federal states, also the states' parliaments, government entities and local governments) are not the same as those that decide on a breach at the European level (finance ministers), this thesis follows the extensive literature on the SGP and assumes broad similarity, at least for the federal level where the actors come from the same party or coalition that forms the government. For a discussion of the extent to which a government can be considered a unitary actor, see von Hagen (2010).

6 Council of the European Union on Economic and Financial Affairs; in the following, this thesis will only use the term 'Council'. The Eurogroup is the relevant composition of the Council with only euro area member states.

the best response would be state intervention and stimulus polices or fiscal consolidation and structural reforms, so as to put state finances back on more sustainable feet. Moreover, the cure that some member states advised others to pursue was not always implemented at home. Unsurprisingly, this was heavily contested. In this omnipresent misunderstanding, though, what missed the political debate were the different economic circumstances and initial conditions the member states were in, and, even more striking, the different views on how (their particular) economy would work. While this conflict between economic policy ideologies has been a major theme during the build-up of EMU, its acknowledgement was largely absent from the political debate during regular SGP implementation. With this, its revival during the crisis came somewhat as a surprise. Since both sides accused the SGP of having failed to prevent the crisis, either by being too rigid or not implementing it properly, the crisis revealed that the role of economic policy ideology for SGP compliance had been underestimated. For this reason, this thesis seeks to investigate the role of economic policy ideology for SGP compliance. This is a novel avenue for research and this thesis links the European integration literature to the regular implementation of the SGP. In particular, it investigates differences in economic policy ideology as a complementary piece to explain adherence to the SGP rules.

The question about reasons for compliance is once again of high relevance today. While the SGP has been declared dead countless times before, the European Commission (EC) recently resumed the debate about reforming the SGP (European Commission 2021). This shows that fiscal surveillance remains an essential part of the EMU institutional design, even though the pandemic has changed the perspective on public deficits. Unlike in the financial, economic and sovereign debt crisis, this time, the SGP rules were temporarily discontinued at the outbreak of the Covid-19 pandemic with the drawing of the general escape clause to facilitate public spending. The Next Generation EU (NGEU) economic recovery package marked a turning point. While beforehand, the main reforms of strengthening the institutional design focused on negative incentives to abate breaches,[7] now the NGEU provides positive incentives to encourage compliance. A large amount of additional funding (750 bn EUR) is offered to member states for investments, and is financed by borrowing on communitarian

7 with financial assistance programmes and later institutionalisation of conditionality with the foundation of the ESM and increasing ‘quasi-automatism’ of sanctions in the EDP

terms (AAA-rated, European Commission 2022b). The numerous crises in the past decade seem to have changed the overall perspective on the role of the state towards increasing state intervention to address the crises (cf. Constâncio 2020). In an environment of abundant money supply and huge demand for investment, the 'old-fashioned' idea of fiscal sustainability seems almost anachronistic. However, with the risk at the doorstep of increased inflation and reduced growth, voices calling for fiscal prudency are rising, and challenge increased state intervention (Alcidi and Gros 2019). Already high stocks of debt in some member states bring the focus back on debt sustainability and puts fiscal surveillance in the spotlight.

This emphasises the need to understand under which conditions member states comply with the SGP. The literature claims that 'political will' is needed for compliance, and several authors call for 'ownership' of the rules (Buti 2006). Some scholars consider the breaching of rules as being the result of unfavourable conditions, such as economic and political capacities (Buti and Pench 2004, Hansen 2015, Johnston and Regan 2016), and call on more assertive political will to implement compliance. Since budgetary policy is not a common European competence, member states ultimately cannot be forced to implement specific policies. Hence, in the quote from above, the European Central Bank (ECB) concludes that 'political will' is required as ultimate motivation to comply. Similarly, even Buti,[8] the then Director at the European Commission responsible for the assessments of compliance, claims that "successful implementation of the […] Pact will depend on political will" (Buti 2006: 19). Accordingly, the literature has named the SGP rules "the dog that would never bite" (Heipertz and Verdun 2004: 765).[9] The argument is that it is unlikely that sovereign states would subscribe to strict limits of public finances and instead advocate to retain a political decision on this in the Council (cf. also Schure and Verdun 2008).

'Political will' appears too vague as an explanation, as it can mean many different things and this is difficult to grasp and to operationalise. In this thesis, I seek to investigate in economic policy ideology as one aspect to better substantiate this notion. 'Economic policy ideology' is understood as a set of beliefs about the means-end relationship of economic, fiscal and monetary policies to address a particular

8 Marco Buti was Director in the Directorate-General for Economic and Financial Affairs at the European Commission (DG ECFIN) at the time and is today this DG's Director-General.

9 Similarly, Gros et al. (2004) called the EC proposal to reform the SGP in 2004 "the dog that lost its bark".

situation. Notably in an environment of high complexity and uncertainty such as the economy, beliefs help policy-makers to "evaluate costs and benefits" (McNamara 1998: 4). In this thesis, the focus is on 'economic policy ideology', which in some countries is part of the wider cultural heritage and interlinks, for instance, with the prevalent conception of the state. Building on rational choice accounts complemented by 'beliefs' (Sen 1977, Hall 1989, Ostrom 1990, Hodgson 2006), I assume that governments act based on rational considerations and ideational perspectives on economic policy issues. Moreover, I view the SGP through the lens of Institutionalism (overview see Hall and Taylor 1996 and Schmidt 2010), by which an institutional design shapes government interaction. Accordingly, 'political will' could reflect convictions about what constitutes 'good' economic policy, both at the national and the European level.

The SGP was designed in an ordoliberal spirit, but convictions about what constitutes 'good' economic policy diverge among governments – and may run counter to the ordoliberal principle. This has been a theme in the literature on EMU integration from the start of EMU (McNamara 1998, Dyson and Featherstone 1999, Heipertz and Verdun 2010). Since the sovereign debt crisis, a revival of ideological arguments has taken place aimed at revealing the best economic strategy to lead out of the crisis (Hall 2012, Brunnermeier et al. 2016): fiscal stimulus for a quick economic recovery or structural reforms to pave the ground for long-term growth and to prevent future crises. In particular, the discussion about the design of the financial support programmes sharpened the line of conflict between calls for austerity to safeguard public finances and to avoid sovereign default, and calls for monetary financing, or, on the one hand, an outright bail-out to free a member state from 'unsubstantiated' financial market pressure, while on the other, structural reforms to enable sustainable economic growth and deficit financing to avoid economic contraction and stimulate economic recovery (Hall 2012, Ryner 2015, Bongardt and Torres 2016, see Featherstone 2016 for discussion on the third bail-out programme for Greece). This indicates that the differences stemming from economic policy ideology unfold especially in times of economic distress, i.e., times of scarcity of public funding and high demand for state action.

This thesis tries to contribute to advancing the understanding of whether adherents to different economic policy ideologies could comply with a one-size-fits-

all rule, such as the SGP under certain conditions. With enforcement relying on political agreement in the Council, and in the absence of an impartial enforcement body, the question is: Could the SGP manage to incentivise fiscal sustainability extrinsically, or would it necessarily need intrinsic motivation, i.e., a conviction that constraining fiscal policy is part of 'good' economic policy?

To explore this idea, this thesis contributes to the academic debate by providing a different theoretical, empirical and methodological perspective: 1) theoretical through considering the SGP as a mechanism to govern a common-pool resource, and with economic policy ideology to endorse the rules as 'appropriate'; 2) empirical through testing competing explanations to better understand how the SGP functions; and 3) methodological through a novel type of a common-pool-resource experiment, which I call 'cultivation game',[10] to study the SGP and which could also serve to study other cases of co-operation.

Specifically, as regards the theoretical contribution, I conceptualise the SGP as a non-hierarchical institution to govern the common-pool resource, 'Euro's monetary stability' (Schelkle 2017: 42ff.), and amend it with overlapping private goods. The common-pool resource overlaps with the private goods of each of the members, i.e., the domestic economy. Extraction from the common-pool resource is, hence, not just based on preferences towards the common-pool resource, but also put in the context of its use and applicability for the domestic economy. At times, members might perceive engaging in maintaining either the common or the private good as a trade-off, i.e., preserving monetary stability and stimulating economic growth, while for others both go hand in hand. I argue that perceiving this as a trade-off depends on the economic policy ideology. In an environment where the use of the common-pool resource is rival and non-exclusive, collective action problems arise to maintain the good (see definition in Ostrom et al. 1994: 6, Olson 1965, Hardin 1968). In the absence of a third-party enforcer, the members need to agree to and endorse the rules to limit the use, and to define the allocation of the common-pool resource. For this to happen, it needs a 'common understanding' (Ostrom et al. 1994: 327) among the members that the rules are 'appropriate' (p. 329).[11] reflecting their needs and expected benefits from

10 In the following, I use the term 'game' to refer to the experimental design and the term 'experiment' to refer to the research method.

11 Ostrom (1990) presents a perspective on rational choice that entails norms as driving the evaluation of costs and benefits: "Individuals are perceived as weighing expected benefits and

co-operation. In the context of the SGP, I suggest that 'appropriateness' emanates from sharing the underlying economic policy ideology. For the purpose of this thesis, I break it down into 'stimulus'- versus 'prevention'-oriented economic policies.

Secondly, the literature puts forward several hypotheses regarding capacity issues, which, however, have not yet been shown to provide exhaustive explanations for rule adherence. I seek to test the following ones in my analyses. The most prominent hypothesis states that a breach is likely to happen in an economic downturn (cf. Hansen 2015). Several studies suggest that the member states' size or heterogeneity, as regards their economic and political impact on the others, shape motivations to comply (Beetsma and Giuliodori 2004, Buti and Pench 2004, Herzog 2004, Irlenbusch and Sutter 2006, Ca´Zorzi et al. 2011, Ohr and Özalbayrak 2012). Additionally, scholars refer to a 'North-South' divide between member states, assuming different levels of capabilities or preferences regarding the continuous realisation of sustainable public finances (Barbone and Poniatowski 2013, Johnston and Regan 2016, Magone et al. 2016). Some scholars argue that the threat of getting sanctioned is not strong enough, as the voting rules facilitate forming non-punishment alliances through reciprocal voting (so-called 'sinners' solidarity'; Irlenbusch et al. 2003[12]). All in all, the literature focuses on explaining breaches, while compliance is expected to happen as a default. However, this disregards, for instance, principled compliance or compliance despite a difficult economic situation, which an experimental study could investigate explicitly.

Thirdly, this thesis develops a novel experimental research design for studying the SGP under laboratory conditions, the 'cultivation game', complementing the two established common-pool-resource games, the investment game (see Ostrom et al. 1994, Deadman et al. 2000) and the request game (Suleiman and Rapoport 1988, Budescu et al. 1995). The development of the experiment forms a genuine part of the thesis. Given that there are no official data on voting in the ECOFIN Council, I cannot use real-world data to study the functioning of both the budget and the voting rules

costs in making decisions as these are affected by internal norms and discount rates. Using this concept of rational action, one predicts that individuals will select strategies whose expected benefits will exceed expected costs" (p. 193). This is distinct from the 'logic of appropriateness' that March and Olsen (1996) formulated later.

12 The term was first used by Otmar Issing at an EU Summit in 1996 (as quoted in Irlenbusch et al. 2003: 647).

and their interplay. An experiment allows this thesis to test competing explanations in a controlled setting. It sheds light on the conditions under which breaches happen. The suggested experiment simulates the SGP in a simplified game about cultivating an apple orchard together with others. It mirrors an SGP-like economic setting and sanctioning procedure. The experiment is based on a common-pool resource that overlaps with the private goods of 19 farmers[13]. The gains from the common-pool resource can be used for advancing the private good. This provides a meaningful context for using the common-pool resource, which goes beyond standard common-pool resource games.

In short, the experimental design is as follows: the participants can harvest from and invest in trees, part of which are exclusive and part of which are shared with others. Harvesting more from shared trees than allowed corresponds to the deficit rule. Participants are asked to sanction such behaviour, which in turn corresponds to the voting rule. The participants choose between 'stimulus'- and 'prevention'-oriented investments to cultivate the field, to co-operate on cultivation with others, and to seal deals to avoid sanctions. When handling scarce resources that are shared with others, participants reveal economic preferences. Playing over several rounds verifies whether such behaviour is consistent. The rounds follow business cycle developments. Participants are assigned randomly to one of three different sizes (small, medium, large) reflecting political capacity. Three hundred and twenty-seven participants, from universities in the capitals of four member states (France, Portugal, Germany, and Greece), played the online game, named the 'Apple Tree Game'. The member states were chosen with respect to the position that they took in the economic policy debate during the sovereign debt crisis. The original debate about the design of the SGP was led by France and Germany (Dyson and Featherstone 1999, Heipertz and Verdun 2010, Hix and Høyland 2011: 258ff.) expressing the most prominent opposing views on economic policy: preferring 'stimulus' versus 'prevention'-oriented policies. During the crisis, Portugal and Greece entered the debate on either side (cf. Lourtie 2012, Magone 2016, Tsoukalis 2012, Featherstone 2016). While Portugal endorsed the German position, Greece supported the French viewpoint.

13 At the time of designing the experiment, the euro area still consisted of 19 members and Croatia only joined in 2023.

The purpose of the experiment is to understand what drives adherence with an SGP-like rule for governing a common-pool resource in principle. It is not to seek to explain member states' compliance track record. Instead, the intention is to disentangle competing explanations as suggested in the literature, and to examine the role of 'economic policy ideology'. The main drawback of experimental evidence is usually its applicability to the real world. There are two main risks: the design could be too abstract to fit the real-world case accurately, and be biased by the experimenter's unconscious perspectives on the topic. Secondly, university students are simply not politicians, whose responsibility towards the electorate cannot be captured by a fictitious experiment. To address these concerns, I have tested the design with a diverse audience. I discuss remaining limitations in a dedicated section. University students are a common target group for experiments, the results of which can provide insights for generalisable conclusions in case some caveats are properly addressed (Druckman and Kam 2012). These results should be understood as providing insights into how, in principle, economic preferences relate to compliance in an SGP-*like* environment.

The results show that economic policy ideologies affect rule adherence. Participants who prefer a 'stimulus'-oriented strategy breach the deficit rule more often. 'Prevention'-oriented participants more often vote correctly. Moreover, this thesis finds strong support for the role of the business cycle and 'reciprocal voting'. This supports the findings from the literature. However, the results neither support the 'North-South', nor the 'size' hypothesis from the literature. Finally, co-operation when cultivating the field together with others increases the likelihood to breach both rules. It seems that co-operative behaviour for cultivation also leads to the belief that the governance could follow the individual's interests, instead of rendering compliance an intrinsic motivation per se.

While the limitations of the experimental method in social sciences are widely discussed (see in particular Section 6.4), the experiment provides three notable insights for practical application: Firstly, the SGP-like voting rules provide disincentives for compliance, and reciprocal voting undermines the original purpose of the rules. Secondly, and in particular in the absence of successful extrinsic disciplining, it becomes apparent that fiscal rules would need to be based on common economic thought in order to be followed. Thirdly, besides these two aspects, it is also economic necessity that drives breaches. The EMU institutional framework might need to

account for this with another institution, in order to support the SGP's success. With this, the thesis provides insights on how to amend the rules so as to respond to member states' plurality and to improve the effectiveness of incentives.

The remainder of the thesis is organised as follows. The following chapter describes the puzzle in more detail: it describes the SGP rules, shows the compliance track record of member states and contrasts it with a review of the literature. Chapter 3 develops the theoretical argument and derives the hypotheses. The research design follows in Chapter 4. Given the fact that the experimental design is novel and a genuine contribution of this thesis, it is not part of the research design, while Chapter 5 describes and justifies its design. Chapter 6 provides the empirical analyses of compliance with the equivalent deficit rule and the voting rule. Chapter 7 discusses the results regarding their applicability to the real-world case of the SGP, summarises the results in that view and locates the thesis' contribution to the academic debate. A sketch of avenues for future research concludes the thesis.

2. The Stability and Growth Pact

This chapter presents the case of the Stability and Growth Pact, and reflects on the academic debate on reasons for rule adherence. The first section sketches the institutional design of the SGP, presents the basic rules and the definition of 'compliance'. Secondly, data on deficit and debt levels shed light on the compliance track record at the national level. This section also reviews studies with more sophisticated analyses to investigate whether and how the SGP rules have impacted domestic budgets. It also gathers information on European-level compliance within the limitations mentioned. Section 2.3 discusses the literature's suggested explanations for compliance. These are contrasted with the data presented in Section 2.2. Section 2.4 summarises, identifies gaps and questions.

2.1 The institutional design and the definition of compliance

The main rationale for introducing the Stability and Growth Pact[14] (SGP) was to avoid 'excessive' deficit and debt levels in the euro area member states.[15] These were considered to ensure the Euro's stability, and to complement the prohibition for the ECB and other member states to bail out a country (Art. 123(1) and 125(1) TFEU, also see Artis and Winkler 1997). EMU is characterised by its asymmetry regarding monetary and fiscal policy competences. While the ECB has the competence to set the monetary policy for the Euro, fiscal policy remains a national competence. The SGP is sought to bridge the gap, and to facilitate co-ordination at the intergovernmental level, as long as such competence at the supranational level is not a politically feasible option.

14 The SGP is based on the Treaty on the European Union (TEU), the Treaty on the Functioning of the European Union (TFEU), Protocol No.12 to the TFEU, and the following legal acts, including their amendments: Resolution of the European Council on the Stability and Growth Pact (Amsterdam, 17 June 1997), Council Regulation (EC) No 1466/97, Council Regulation (EC) No 1467/97, and further secondary legislation (see https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/stability-and-growth-pact/legal-basis-stability-and-growth-pact_en).

15 This thesis only considers euro area member states. All EU countries are asked to comply with these rules. However, for euro area member states, this is especially relevant as they share a currency.

The details of the rules have been reformed several times,[16] but the essence, which is to constrain public borrowing, did not change. Given the experimental method used in this thesis, it is not the details but the essence of the rules that are to be meaningfully transmitted into an experimental design. In short, this refers to the deficit rule as the relevant rule that has been used to trigger corrective action, and to the Council's political judgement on the definition of compliance and respective corrective action, which might also go against the spirit of the SGP.

In essence, the SGP asks a member state to implement the rules at two intertwined levels: complying with the deficit and debt rule (which I call '*national level'* in the following), and sanctioning another member state's breach in the Council (which I call '*European level'*). At the national level, member states are asked to keep their general government deficit level below 3% and their debt level below 60% of their Gross Domestic Product (GDP) (Art. 104 TFEU, Protocol 12 to TFEU). There are exception clauses for economically severe times and other emergencies, which may require higher public borrowing. Moreover, over time, both rules have been refined to accommodate specific investment needs, the economic situation, and to promote a reduction path of levels that exceed the reference values. Of both rules, the deficit rule seeks to ensure short- and medium-term fiscal sustainability, and is currently the relevant indicator of the SGP. The debt rule is considered with regard to debt sustainability, which has only become salient since the sovereign debt crisis[17] but has so far not been considered to have triggered corrective action[18]. The debt rule is

16 This section only describes the basic rules, which were not changed with the numerous reforms. The reform in 2005 (see a description in Hayes-Renshaw and Wallace 2006: 288ff., Heipertz and Verdun 2010: 162ff.) introduced the notion of a 'temporary' exceeding of the deficit limit as acceptable if it is due to the economic situation. Further amendments have taken place in the follow-up of the sovereign debt crisis (Two-Pack, Six-Pack, Fiscal Compact). In particular the calculation of the deficit limit has become rather complex with focusing on the structural deficit (for a discussion see Leiner-Killinger and Nerlich 2019). The medium-term objective (MTO) is calculated individually for each country. It describes the level of the cyclically adjusted deficit that is allowed given the potential output and structural elements of the deficit (see Angerer 2015 for a concise description). It was first introduced in the preventive arm with the reform in 2005. Before, the preventive arm prescribed to keep the budget balance close to balance or in surplus. Later reforms put the medium-term objective (MTO) at the centre of the surveillance and, hence, shifted the focus from the corrective arm to the preventive arm. Additionally, the debt rule is now complemented with a downward projection path.

17 Long-term debt sustainability became key to defining a member state's financing problems as liquidity and not as solvency issues. Support for the latter falls under the no-bail-out clause, and would legally not have been possible.

18 The reform in 2011 introduced the debt criterion by which the debt level that exceeds the reference value of 60% has to be reduced by one-twentieth per year. However, this criterion has not yet been used by the Commission and the Council to trigger corrective action.

currently being discussed in the governance reform debate to gain a greater role in the SGP in conjunction with the deficit rule (also see discussion in section 7.2.2).[19]

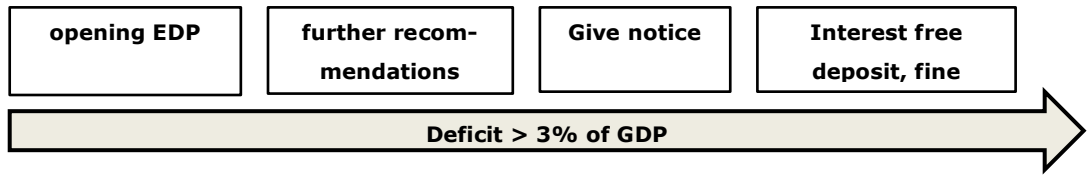
At the European level, the SGP says that a breach has to be sanctioned by the finance ministers in the Council. In case the deficit level exceeds the threshold, the Commission formulates an assessment, on the basis of which the Council decides whether the deficit is to be considered 'excessive' or not.[20] The Council decides without the member state in question (Art. 121(4) TFEU and 126(13) TFEU), and with qualified majority.[21] This opens the Excessive Deficit Procedure (EDP, Art. 126(6) TFEU), which defines steps of escalation in case the respective member state does not correct the deficit in a specific time frame. The ultimate step is a financial sanction in the form of a fine. The figure below describes the main steps. Together with opening the EDP, the Council issues recommendations to the member state (Art. 126(7) TFEU). In case the member state does not follow them within a certain period of time, the Council can decide on further steps. In case the requirements are met, the Council decides to close the EDP (Art. 126(12) TFEU).

19 The debt level is suggested to serve as a qualifying indicator to set the required level of fiscal consolidation.

20 According to the legal basis, the role of the European Commission is that of a neutral examiner. The Council retains full decision power, and can also amend the proposal, though not write one of their own (European Court of Justice C-27/04). The Council cannot be forced to decide for sanctions, not even by the Court of Justice of the EU (Art. 126(10) TFEU).

21 Only those countries having the euro as their currency are allowed to vote (Art. 136(2) TFEU). Originally, the qualified majority for the Council (Art. 238(3)a TFEU, Art. 3(4) Protocol No. 36 to TFEU) is met if 73.9% of the votes, and 10 out of 18 member states (19 minus the one in question) agree. A state can request a majority of 62% of the population. This implies that abstentions qualify as 'no' votes. The number of votes that a member state has is weighted according to its population size: Germany, France, Italy 29; Spain 27; The Netherlands 13; Belgium, Greece, Portugal 12; Austria 10; Ireland, Finland 7; Cyprus, Luxembourg, Slovenia 4; Malta 3 votes (Art. 3(3) Protocol No. 36 to TFEU). Since the Lisbon Treaty, the calculation of the qualified majority has changed, including a transitional period from 2014 to 2016 (for a discussion of the details, see Verdun 2013: 1132, Thomson 2013: 1222f., for the Nice Treaty rules, see Hayes-Renshaw and Wallace 2006: 265). However, in principle, the weighting according to population size combined with the single votes per member state remain. Since the recent reforms, the last two steps of the EDP linked to financial sanctions require reverse qualified majority voting. This means that a proposal is endorsed if there is no qualified majority against it. However, the opening of an EDP and all the other steps still require a qualified majority. This reform sought to abate the power of large member states who before could easily form a blocking minority. Moreover, the reforms sought to strengthen the 'automaticity' of the EDP by prescribing time limits within which a decision must be made. These changes are not relevant for this thesis as it focuses on the essence of the rules, and not the details.

Figure 2.1-1: Steps of the excessive deficit procedure.



Source: own illustration based on Art. 126 TFEU.

Moreover, the SGP has two so-called 'arms': the preventive and the corrective arm. The first consists of a permanent multilateral surveillance process, where the Commission and the Council review whether the rules were met on a yearly basis.[22] The basic rule on the current budget position is to be 'close to balance or in surplus'. It is refined with country-specific medium-term objectives for the structural balance and with fiscal adjustment requirements towards the objective, defined based on a matrix considering the economic development (output gap) and the debt level (below or above 60% of GDP). The corrective arm consists of the EDP as described above.

This thesis considers compliance at both levels to be intertwined and, therefore, to be similar in terms of their political role. While only the deficit and debt rules are legally binding EU law, the idea of compliance at the European level results from the spirit of the rules, but is not a legal rule and cannot be legally enforced. With this, compliance on both levels has a different legal quality. Members in the Council enjoy sovereignty when voting, and this is not constrained by a legally binding rule. Instead, they are free to exert political judgement. This also reflects in the varying degree of firmness that the Council showed towards the rules, as will be discussed below (see Sections 2.2 and 2.3). Nonetheless, the member states are requested to vote on the issue, and are asked to consider the national-level rules for their voting assessment. This is meant with the political notion of compliance. Therefore, this thesis considers both levels' rules to be similar for the purpose of studying their intertwined compliance. For ease of reading, this thesis calls the European-level rule a rule, although it is not a legal rule as such, but one that emanates from the SGP's spirit.

22 Today, this process is detailed in the 'European Semester', where member states submit their budgetary plans in the autumn and the European Commission checks upfront whether the budget risks are non-compliant. Moreover, the Commission issues country-specific recommendations that include suggestions for broader economic policy reforms, which, vice versa, can be taken into account by the member states for their medium-term policy priorities to be formulated in the Stability Programmes to be submitted in the spring.

2.2 Compliance track record

There is a rich literature demonstrating the weak compliance track record with the deficit rule. The debt rule was not much considered until the outbreak of the sovereign debt crisis. In contrast, the individual member states' adherence to the voting rule is hardly documented. There is data for unsanctioned breaches of the deficit rule, but not for who voted on sanctions. The below graphs only consider the period since the non-reversible peg of national currencies to the euro in 1999 until 2008. Afterwards, the financial, economic and sovereign debt crisis took full effect, which lasted a considerable time (see Gros 2017).[23] Some countries were still under post-programme surveillance at the outbreak of the Covid-19 pandemic in 2020 when the activation of the general escape clause temporarily discontinued the rules.

While adherence with rules at the European level can be considered to be purposeful, breaches at the national level might also be incidental in the case of unforeseen events. This might be attributed to forecast errors, overly optimistic forecasts or ex post revisions of statistical data[24] (European Central Bank 2008, Beetsma et al. 2009, von Hagen 2010, also see Eyraud et al. 2017), or else reluctance to implement relevant policy measures (Wierts 2008: 99ff.), which is subsumed under 'political will' (for instance in the run-up to elections, see Afonso and Hauptmeier 2009). It is difficult to disentangle the effects[25] but the European Central Bank (2008) notes that "budgetary plans are seen to have been too optimistic mainly in countries that have incurred excessive deficits for significant periods (Germany, Greece, France, Italy and Portugal)" (p. 63, similar conclusion in Buti and Pench 2004: 1026). The budget remains subject to the decision of the Parliament, and hence can ultimately be considered a purposeful decision.

23 One could argue that the announcement of the ECB's Outright Monetary Transactions programme (OMT) in 2012 marked the end of the crisis. However, governments faced lengthy struggles bringing fiscal positions below the reference levels, and both the Commission and the Council were reluctant to progress with the sanctioning procedure, especially in light of strong domestic opposition in some countries towards the conditionality of their financial assistance programmes. This was the case, for instance, in France and Italy.

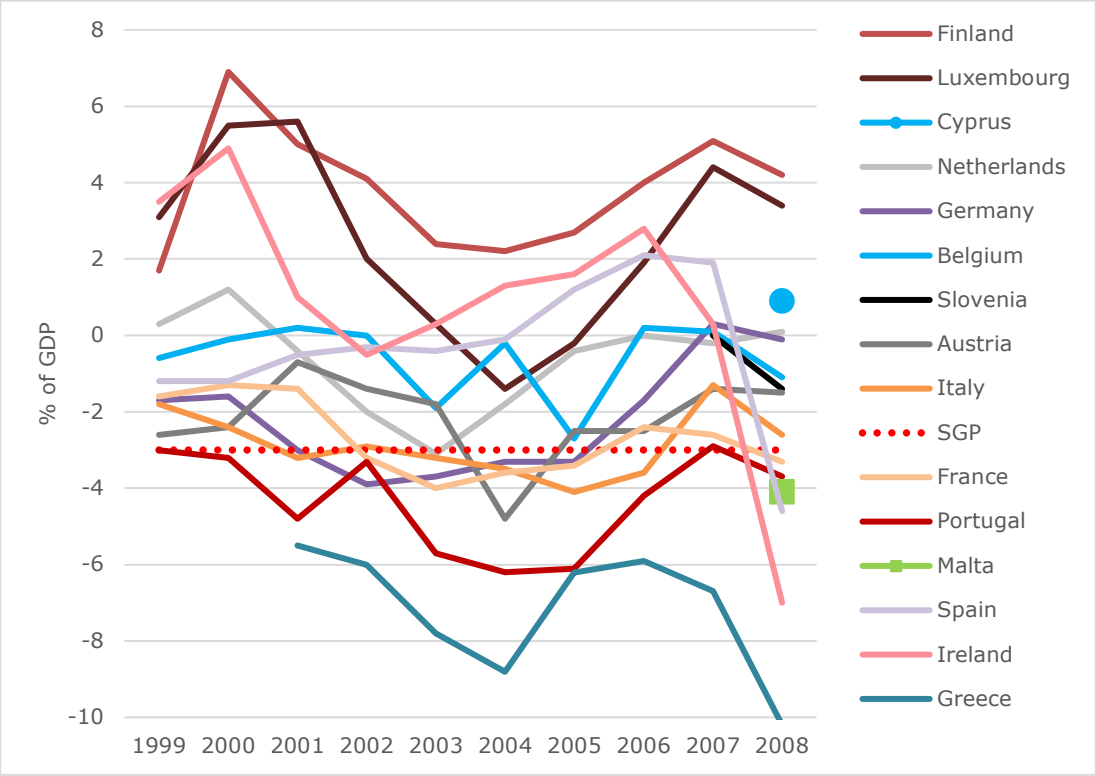
24 Italy and Portugal misreported data (Ngai 2012) but the most significant case was Greece. In 2005, it became known that Greece had misreported deficit and debt levels for years, including during the time when its accession to the euro area had been decided (European Central Bank 2008: 59).

25 This is also the reason why during the reforms of the SGP, the focus shifted from the government deficit to the cyclically adjusted balance, which is considered to reflect the part of the balance that can be attributed to government action, while the cyclical component is considered to be due to unforeseen events.

2.2.1 National-level compliance: deficit and debt levels

Defining a breach of the deficit rule is up to the European Commission's assessment. The figure below depicts the general government deficit levels compared to the 3% threshold.[26] A case is defined as one country-year observation to reflect that a budget applies to one year. The data show that the deficit level has been equal to or above the 3% threshold in 37 cases. This represents 30.3% out of 122 cases. The member states that most often had such deficit levels were Greece, Portugal, Germany, France and Italy.

Figure 2.2-1: Deficit levels in euro area member states – 1999 to 2008.



Source: Eurostat 2021a, General government Net lending (+) /net borrowing (-). The data was extracted in November 2021 and, hence, does not reflect the data available at that time. In some cases, the data have been revised. The graph shows the countries as they became members of the euro area. The red dotted line shows the SGP's reference value of -3% of GDP.

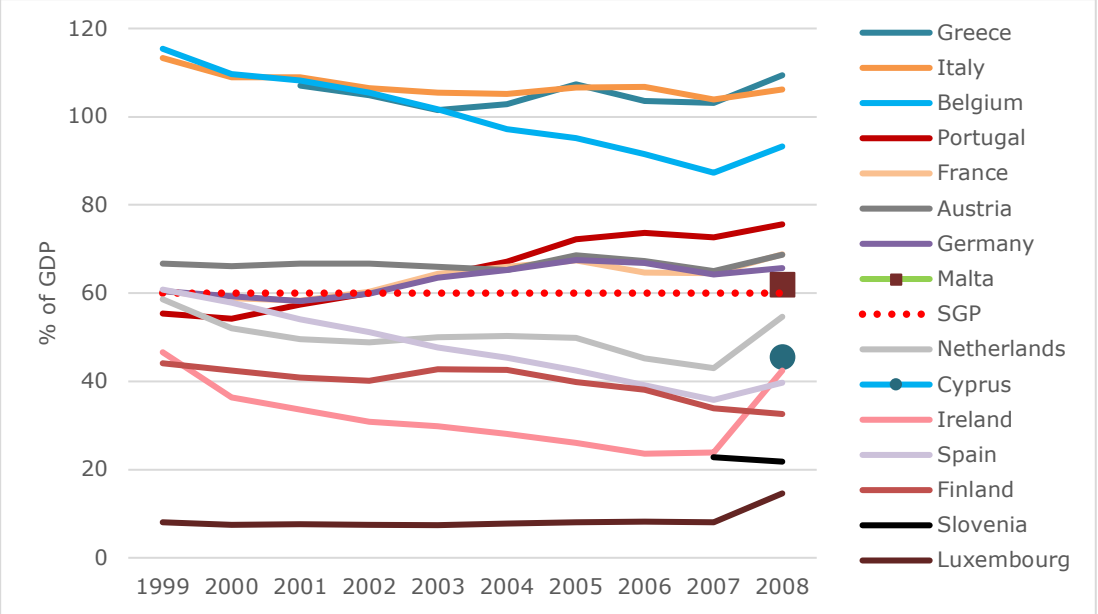
Similarly, more sophisticated research shows that many member states breached the deficit rule (Eyraud et al. 2017). The European Central Bank (2008) shows that Greece, Portugal, Germany and Italy breached the deficit rule most often. For all

26 The European Fiscal Board also uses the numerical data without discretionary assessment by the European Commission for its Compliance Tracker (Larch and Santacroce 2020).

except Germany, data misreporting was common at the time, and was only made public later. Even in the preventive arm, Greece, France, Italy and Portugal never complied between 1998 and 2007 (European Central Bank 2008).

Despite the fact that the debt rule has not been as relevant as the deficit rule, it is discussed here to provide a complete picture. Compliance with the debt rule is defined as levels below 60% of GDP.[27] The figure below shows that only seven countries managed to remain below the threshold. Of those countries where the debt level was significantly above 60%, only Belgium and Italy managed to reduce it (see also Ngai 2012), however none managed to comply (European Central Bank 2016: 53, Chart 1). For Greece and Austria, the level basically did not change over time, while the debt level in Portugal substantially increased. Debt levels also increased for France and Germany, but only during their economic bust times between 2002 and 2005; they declined again afterwards until 2007. Koehler and König (2015) find that after 1999, debt levels declined in most countries, but not in those that had to undertake particular effort to comply with the Maastricht criteria such as Greece, Portugal and Italy.

Figure 2.2-2: Debt levels in euro area member states – 1999 to 2008.



Source: Eurostat 2021b, General government gross debt. The data were extracted in November 2021 and, hence, do not reflect the data that was available at the time. In some cases, the data have been revised. The graph shows the countries as they became members of the euro area. The red dotted line shows the SGP's reference value of 60% of GDP.

27 This holds for the data until 2008. Later reforms introduced country-specific debt reduction paths.

The literature is divided over the question of whether the SGP had a disciplining effect on deficit and debt levels. While some studies find indications of the intended negative effect on deficit and debt levels (Afonso and Hauptmeier 2009;[28] except in Greece, Portugal and Italy: Koehler and König 2015, European Central Bank 2008), others find that the rules on budget outcomes had no disciplining effect (Ioannou and Stracca 2011, Lemmer and Stegarescu 2009, Ngai 2012). In particular, scholars find no effect on constraining expenditure levels (Afonso and Hauptmeier 2009, Hauptmeier et al. 2010). Even more, Lemmer and Stegarescu (2009) find that revenue windfalls were not used for debt reduction but for increasing expenditures, especially "in countries that also have not met their medium-term objectives" (p. 159). Following the disciplining effect of the Maastricht criteria in the run-up to their accession to the euro area (Ngai 2012: with the exception of Greece) and a favourable economic situation (Hallet and Lewis 2004), there was some 'fatigue' (Fatás and Mihov 2003, similar in Hallet and Lewis 2004 and Schuknecht et al. 2011). In particular, Debrun et al. (2004) show that all member states that were in breach of the deficit rule in 2004 would not have been in breach if they had continued with their same structural balances as in 1999. Nonetheless, scholars also do not find a deficit increasing effect of the SGP as was expected due to 'moral hazard' (Ioannou and Stracca 2011, Afflatet 2017). In that regard, Begg (2017) notes that "just testing whether rules are complied with neglects the fact that they often tug fiscal policy towards the target enshrined in the rule […]. In this sense, the presence of the rule has an impact missed by looking only at full compliance" (p. R11).

Instead, scholars point to the role of economic growth and the reduction of borrowing costs for 'incidental' rule compliance. Bailey and Fingland (2008) observe that "rather than indicating the success of the SGP, the decline in levels of public debt since the mid 1990s may be due in part to trends in GDP growth […] and partly to falling interest rates in the Eurozone" (p. 225). Hallerberg (2016) objects, pointing out an indication of the SGP's success in the "comparison of the recessions in the early 1990s and in the early 2000s, with deficits [that were] smaller in the latter period than in the former" (p.150).

28 They consider all EU countries together and not only euro area countries.

As a result, there seem to be some countries which are more prone to breaches than others, such as Greece, Portugal, Italy, France and Germany. In particular for Greece and Portugal, the literature finds no disciplining effect of the rules. Even more, they seem to be at odds with the development in the rest of the euro area, as Hallerberg (2016) notes that in the first decade of the Euro, "there was optimism that the Eurozone was indeed leading to greater discipline […] with the exception of Greece and Portugal" (p.150). France, Italy and Germany seem to alternate between compliance and breaches, while the rest of the member states does not particularly stand out for their breaches track record. In contrast, Spain, the Netherlands (except in 2003), and all small countries with the exception of Malta, always complied throughout the years (except Ireland in 2008). Belgium and Austria had difficulties with the debt rule while complying with the deficit rule (except Austria in 2004).

2.2.2 European-level compliance: sanctioned breaches

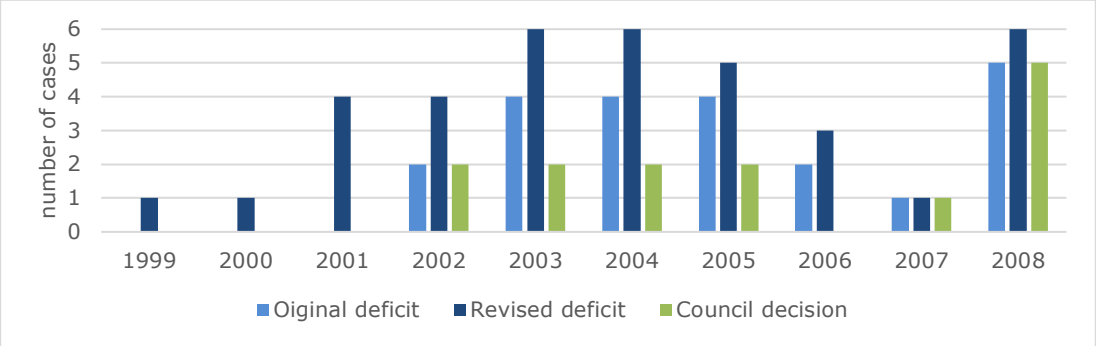
As stated above, there is no information on how individual finance ministers voted in the Council due to a lack of data.[29] Instead, one can look at cases of unsanctioned breaches, which also represent a breach of the rules. A case refers to a country-year observation, whether there was a deficit in that year and whether this had been sanctioned.[30] For the sanctioning, the graph considers the year that the Council decision refers to, which is not necessarily the year in which the decision was made.

In contrast to the above, a breach of the SGP rules is defined by the Council's decision. Both should essentially be the same and only vary due to the evaluation of whether the deficit is considered temporary or the economic situation severe. The literature mentioned above has shown that revised data play a role and therefore, the graph below reports surpassing the 3% deficit threshold, the revised data, and the Council decisions to open an EDP, or to proceed with the relevant steps within an EDP.

29 There is just one incident when votes were reported in the Council decisions, and this was on 23 November 2003 when the decision to open an EDP for France and Germany failed (see Hayes-Renshaw et al. 2006: 176).

30 With this, the statistic does not count individual steps of the EDP procedure, which apply to the same budget year. It rarely happened that several steps were taken in one year. This was only the case for Greece for the budget year 2004 with decisions to "establish inadequate action" (Article 126(8) TFEU) and giving notice (Article 126(9) TFEU).

Figure 2.2-3: Council decisions and 'excessive' deficits – 1999 to 2008.



Source: original deficit: European Commission's EDP reports and proposals (see European Commission 2022a); decisions: Council decisions (see European Commission 2022a); revised deficit: Eurostat in November 2021 (Eurostat 2021a). The graph shows the number of country-year cases and lists the countries as they became member of the euro area.

Table 2.2-4: Council decisions and 'excessive' deficits per country – 1999 to 2008.

Country	Original deficit	Revised deficit	Council decision	Duration of EDP
Germany	4	5	2	2002-2005
France	4	5	2	2002-2004, 2008
Finland	0	0	0	---
Italy	3	5	1	2004-2006
Luxembourg	0	0	0	---
Spain	1	1	1	2008
Greece	5	8	4	2003-2005, 2007-2008
Slovenia	0	0	0	---
Ireland	1	1	1	2008
Portugal	2	9	1	2005-2006
The Netherlands	1	1	1	2003
Austria	0	1	0	---
Belgium	0	0	0	---
Malta	1	1	1	2008
Cyprus	0	0	0	---
Sum	**22**	**37**	**14**	

Source: see Figure 2.2-3. The table shows the number of country-year cases. Countries with no cases are marked in green. The column 'duration' shows the years to which the EDP refers. The procedures that were started in 2008 lasted until after the crisis. The table reports the countries as they became members of the euro area.

The table shows that between 1999 and 2008, there were 22 cases of surpassing the 3% deficit threshold at the time (18.0%), of which a third went unsanctioned (36.4%). Revised statistics reveal even higher numbers of breaches (37 cases, 30.3%), of which two thirds remained unsanctioned (62.2%). France and Greece have experienced two EDPs, while Greece has experienced the longest period in an EDP, i.e., five years,

closely followed by France and Germany with four years. Although some procedures lasted for a long time, serious sanctions had been avoided. There were only three cases where the Council decided to give notice: in 2005 and 2010 for Greece, and in 2006 for Germany.[31] The Council never decided on a deposit or fine.[32] In 2003, a Council decision to give notice to France and Germany failed,[33] which led to the reform of the SGP in 2005. The reluctance of the Council to sanction breaches is also supported by Hallerberg (2013) who finds that the Council weakened the texts of one third of all Commission proposals.[34] Surprisingly, with the unfolding of the crisis in the euro area in 2008, there is a high level of European-level compliance, with all cases of surpassing the 3% deficit level being sanctioned with the opening of an EDP (except for one case with revised data in Portugal).

It is noteworthy that in the event of a member state presenting action and the Commission assessing this to be adequate to reduce the deficit over the medium-term, no follow-up decision by the Council was necessary according to the rules at the time. This was, for instance, the case for Portugal from 2005-2006. Moreover, some cases of revised deficits that went unsanctioned can be associated with misreported data, as highlighted above in the cases of Portugal, Italy and Greece.

2.2.3 Summary

The SGP has not been successful in constraining deficit levels and bringing down debt levels equally in all member states. Moreover, there has been a lack of European-level compliance to enforce the rules. In particular, France, Germany, Greece, Italy and Portugal have shown many cases of breaches, which to a large extent remained unsanctioned. However, despite the large body of evidence for breaches, there is also large compliance at the national level, and to a lesser extent also at the European level. This is puzzling for both levels' compliance given the weak threat of sanctions that the Council has shown, and given that there is no legal risk for the Council not to vote for

31 This refers to the year in which the decision was made and to the budget years in 2004 and 2008 for Greece and 2005 for Germany.

32 Savage and Howarth (2020) consider the financial sanction against Valencia in 2015 for having misreported data to Eurostat as the first instance of the Council deciding on a fine, even though this does not strictly relate to the EDP process.

33 See Schure and Verdun (2008: 479ff.) for a description of the case.

34 He also finds that the member states failed to comply with the weakened texts in half of the cases.

sanctions. Therefore, from a rational perspective, one might have expected to see even more cases of breaches. The literature review sheds light on some explanations.

2.3 Literature review

The first section locates the SGP in the broader academic debate around EMU that focuses on the economic rationale and the purpose of the SGP as economically and politically meaningful. While this literature does not explicitly explain rule adherence, it shows the deeply rooted scepticism towards to the SGP, which might undermine its legitimacy and motivation for enforcement. The following two sections discuss the literature on explanations for rule adherence, which I categorise into 'capacity' and 'ownership' issues.

2.3.1 The SGP as part of EMU: economic rationale and purpose

During the first decade of the SGP, the economic rationale of the pact was at the centre of academic discussions (see Heipertz 2003 for a discussion, see Fischer et al. 2006 for an overview). The lack of economic justification of the numerical limits of 3% and 60% (Eichengreen and Wyplosz 1998, Eichengreen 2003, good overview in Kohler 2007),[35] the inappropriate 'policy-mix'[36] pointing to an institutional design that lacks a 'political union'[37] to ensure a smooth operation of monetary, fiscal and economic

35 The numerical rules were also considered too rigid and of risk to induce pro-cyclical fiscal policy which is seen as growth-hampering (Eijffinger 2003, de Haan et al. 2004, Enderlein 2004, Ferrero 2005, Beetsma and Debrun 2007, de Grauwe and Ji 2014). This, however, was also contested by other scholars who did not find that the SGP induced pro-cyclical fiscal policy (Buti et al. 1998, Deroose et al. 2008). In contrast, the advantage of simple numerical rules is that "they facilitate financial market and public scrutiny of fiscal policies" (Schuknecht 2005: 65).

36 The policy-mix is perceived as ineffective because competences are split between the national and the supranational level (see overview in Collignon 2007). Scholars argue that it lacks a supranational body that could serve as counterpart to the ECB for economic policy taking into account the "aggregate fiscal stance" (Eijffinger 2003: 11) such as a larger EU-level budget (European Commission 1990: 168, Hishow 2007a, 2007b), fiscal transfers or some automatic stabilisers subsumed under the term 'fiscal capacity' (first mentioned in the MacDougall report, European Commission 1977).

37 Cohen and Subacchi (2008) argue that "the problem lies in the governance structure of EMU. Because the euro is a currency without a country, based on an inter-state agreement, participating members find it difficult to speak with a single voice" (p. 1). De Grauwe (2006) summarises that, ideally, "a full political union […would] include[…] a Central European government with the power to spend and to tax, and which is independent of national governments. Such a government is the only institution that can fully back the ECB" (p. 143). Vanthoor (1997) gives a good overview about the history of discussing a 'political union' and argues that the reason why this was not established when founding EMU is the disagreement between France and Germany about what a 'political union' should aim at (also see Dyson and Featherstone 1999: 14, 20).

policy (Eichengreen 1996, Buiter 1999, Catenaro et al. 1999, Collignon 2001, Begg 2003, de Grauwe 2006 and 2013, Sims 2012), and the imperfection of EMU as an optimal currency area (OCA) and accordingly built with a badly chosen group of members (Bean 1992, Beetsma and Bovenberg 2000)[38] were brought forward as arguments of a flawed SGP and EMU institutional design.[39] In particular, criticism targets the overall aim of the monetary union[40] as not having contributed to "high growth and low unemployment" (de Grauwe 2006: 138). In contrast, just before the outbreak of the sovereign debt crisis, Enderlein and Verdun (2009) conclude that "those effects have not translated into the expected decrease of EMU's legitimacy or a widespread democratic deficit of EMU" (p. 491).

The role of the financial market to exert fiscal discipline was seen by some scholars as a more effective mechanism than fiscal rules (Leblond 2006, Goodhart 2006). Some scholars see such a role of markets at risk, especially as it hinges on the credibility of the no-bail-out clause (see discussion in Eyraud 2017: 28ff., Freier and Ciaglia 2019).[41] In contrast, de Grauwe (2011) argues that the EMU design provides too much power to financial markets. Since member states "cease to have control over the currency in which their debt is issued"[42] (p. 1), he argues that the SGP 'invites' financial markets to bet against 'fragile' member states in case of distress and, hence, would not contribute to meaningful fiscal disciplining (de Grauwe and Ji 2012).

The second decade started with the financial, economic and sovereign debt crisis and, accordingly, the focus shifted towards a more encompassing perspective

38 The convergence of member states' business cycles and the level of economic integration in the euro area was considered insufficient, risking the spread of asymmetric, country-specific shocks which the euro area is not equipped to address (for a review of the OCA theory see Mongelli (2002) and Schelkle (2010, 2017: 19-21).

39 Scholars also proposed different SGP designs (Casella 1999, Lindbeck and Niepelt 2005, Schelkle 2005).

40 As was laid down in the European Commission's seminal study (European Commission 1990).

41 Leblond (2006) and Wyplosz (2006) show that the discontinuation of the Pact in 2003 and the European Court of Justice's decision in 2004 had literally no effect on financial markets. In contrast, Kalan et al. (2018) show that the opening of an EDP is associated with an increase in spreads. The fact that sovereign debt is associated with 0% risk weight in banking supervision is considered a disincentive for banks to exert market discipline (Bénassy-Quéré et al. 2018). On the other hand, in 2013 it was introduced that all newly-issued sovereign debt should include a collective action clause (CAC). This signalled to bond holders that their private investment might be 'bailed-in' in case of default – and not necessarily bailed-out by the public. The source of the discussion is that, by establishing the monetary union, the external valuation of the domestic currency by financial markets was abolished.

42 In the sense that the ECB took over the competence from the national central bank, which makes external devaluation of the currency impossible.

that sees the SGP only as one part of a larger institutional framework to govern the monetary union.[43] The Banking Union was created to focus on the doom loop between sovereigns and banks.[44] The European Stability Mechanism (ESM) was established to support member states in liquidity problems and, thereby, to avoid a sovereign default in the currency union.[45] With this, the ESM serves, to some extent, as a lender of last resort purpose, which was at the time complemented by the ECB to (announce to) purchase and accept sovereign bonds for their monetary policy.[46] The main intention was to bring speculative action against member states to a halt, and to prevent contagion effects between member states.

While financial markets assessed individual states on the basis of their credibility to service debt, they also took into account significant contagion effects among several member states in particularly high distress,[47] and were accused of adding 'panic' in light of the tremendous uncertainty in the financial system at the time (de Grauwe and Ji 2012, de Grauwe 2020: 228ff.). Ultimately, this also entailed a risk of break-up of the euro area. At the time, the financial market, together with the rating agencies, were accused of having a short-sighted perspective, which also reflected the

43 Long before its creation, the monetary union was discussed as one part of a much bigger picture of European integration. It was conceptualised as the logical consequence of the single market. Later, the European Commission's seminal study on the design of a monetary union entitled "One market, one money" (European Commission 1990) pictured "rules for national budgets" within an economic union together with the single market and "shock-absorbing functions of budgets" as a separate element next to the monetary union with a single currency and an independent central bank (p. 12).

44 Not only does this include banking supervision being put at the European-level with the creation of the Single Supervisory Mechanism (SSM), the Single Resolution Boards (SRB), the single Resolution Fund (SRF) and the streamlining of supervisory co-ordination in the European System of Financial Supervision (ESFS) and the development of the European Systemic Risk Board (ESRB) to focus on macro-prudential aspects, but it also refers to several legislations like the Banking Resolution and Restructuring Directive (BRRD) that clarifies the 'pecking order' in case of default with the public only as a last resort, and private investors to be involved beforehand. See Schäfer (2016 and 2017) for an evaluation of the introduction of the Banking Union. See Pisani-Ferry (2012) for a discussion of the doom loop.

45 Based on a decision by the member states (the Board of Governors is equivalent to the Council configuration of the Eurogroup), the ESM provides financial assistance and technical support in exchange for fiscal consolidation and structural reforms ('conditionality'). The ESM was preceded by a bilateral support programme for Greece in 2010, the European Financial Stabilisation Mechanism (EFSM) and the European Financial Stability Facility (EFSF). The institution was preceded by an academic debate about a European monetary fund (Gros and Mayer 2010).

46 The OMT was announced, but has not been used yet. For an evaluation of asset purchase programmes, see Belke and Gros (2021) and Havlik and Heinemann (2021).

47 Schelkle (2011) argues that the financial market was "incapable of setting consistent terms of fiscal responsibility; Irish bond issues are attacked both for too little and too much austerity at the same time" (p. 382). It is noteworthy that despite the fact that this affected member states who were already not in a favourable fiscal position and had breached the SGP several times, such as Portugal, Greece and Italy, also Spain and Ireland, who had good fiscal positions and had never breached the SGP before, came under stress and eventually sought financial support from the EU.

high uncertainty about the commitment for governance in the euro area (see Schelkle 2011, discussion see Howarth and Quaglia 2015). Also in that regard, scholars suggest transferring some fiscal sovereignty to the EU level (see discussion in Eyraud et al. 2017). De Grauwe and Ji (2018) argue: "the sovereign bond markets in the Eurozone will continue to be prone to instability" as long as the monetary union lacks "[p]olitical unification […] because the Eurozone has dramatically weakened the power and legitimacy of nation states without creating a nation at the European level" (p. 1).

Accordingly, the academic debate centred around the 'proper' EMU design that is able to absorb both symmetric and asymmetric shocks (see discussion in van Riet 2016, Alcidi et al. 2017). While one side pointed to the SGP as being the critical element in the EMU institutional design to ensure sound public finances and accordingly provide room for manoeuvre in economically bad times (Marzinotto and Sapir 2012), the other side saw the SGP as a co-ordination device for stimulating economic policy in economically bad times (see debate about 'gouvernement économique'[48]).[49] In contrast to the previous OCA theory, and in light of increasing heterogeneity between states due to the crisis, Schelkle (2017) suggests such heterogeneity as representing an opportunity for 'risk-sharing' to alleviate the repercussions of asymmetric shocks, notably through already existing channels stemming from the single market, such as trade and capital flows.[50] Common sovereign debt issuance was discussed as a means to leverage on capacities (de Grauwe 2011, Sims 2012, Pisani-Ferry 2012, for proposals see Delpla and von Weizsäcker

48 While during the crisis, the European Council and the Council (and especially its composition as Eurogroup) developed into the political centre of steering through the crisis (Puetter 2012), the political debate also revolved around a European 'gouvernement économique' as a powerful counterpart to the ECB on the one hand, and a 'finance minister' at the EU level, accompanied by a euro area budget on the other.

49 In the political debate on 'deepening EMU' following the acute crisis resolution, the five-presidents' report (2015) located the SGP in a broader framework to 'complete' the monetary union, including a common macroeconomic stabilisation function (to which access would be conditional) and a treasury. A Franco-German group of scholars sketched an institutional reform to 'reconcile' economic stabilisation and fiscal discipline (Bénassy-Quéré et al. 2018, and 'Italian' response by Messori and Micossi 2018).

50 Additionally, besides fuelling claims for a fiscal capacity with a stabilisation function (de Grauwe 2011), this also resulted in suggestions of a common deposit insurance scheme (Pisani-Ferry 2012), or an unemployment insurance scheme (Dullien and Schwarzer 2005, Dullien 2017, Dolls et al. 2018). The former ranges from some stabilisation element (see Schelkle 2005 for a suggestion), the institutionalisation of automatic stabilisers at EU-level, to investments in designated areas to improve competitiveness and development at EU-level on top of national-level endeavours as was sought with the Lisbon Strategy and Europe2020. Financial market integration is considered to alleviate the severity and differences of shocks across the euro area (Alcidi et al. 2017).

2010, German Council of Economic Experts 2011, Fuest and Heinemann 2017, Gros 2018) and criticised from a moral hazard point of view (Heinemann 2021 suggested a sovereign debt restructuring mechanism).[51] The common worry during the crisis was how to define solidarity without risking moral hazard in the future. This line of conflict was reflected in the political debate of the time, and reflects in the ESM support being conditional on reforms and fiscal consolidation. The SGP rules were reformed to become stricter (Six-pack, Two-pack[52]) and to increase 'ownership' (with the Fiscal Compact[53]).[54] A monitoring scheme for macroeconomic imbalances was established, together with the European Semester where all processes were gathered to follow a cycle of reporting and evaluation ('Six-pack' and 'Two-pack' legislations).

Most recently, with the 750 bn EUR worth Next Generation EU fund (NGEU)[55] and the European Commission's launch of the SGP reform, the academic debate is comparably quiet on the SGP,[56] and instead praises the NGEU for being the long overdue euro area fiscal capacity (de Grauwe and Ji 2021).[57] Other voices, in particular

51 The supporters of the latter criticise the former for trying to artificially put a price on sovereign debt that ignores market evaluation and creates moral hazard for high-debt countries. The supporters of the first criticise the latter for intensifying short-sighted market pressure and ignoring the fact that all debt is issued in the same common currency.

52 Six-Pack: Council Directive 2011/85/EU, Regulations 1173/2011, 1174/2011, 1175/2011, 1176/2011, 1177/2011; Two-Pack: Regulation No 472/2013, Regulation No 473/2013.

53 Treaty on Stability, Coordination and Governance

54 Despite the fact that the reform introduced reverse qualified majority voting (RQMV) for some steps of the EDP, which aimed at increasing the applicability of sanctions and at reducing incentives for strategic interactions in the Council, the reform was criticised either for introducing too much complexity, continuing with the same economic rationale, or not becoming strict enough (Kullas 2011). Seng and Biesenbender (2012) argue that the introduction of reverse qualified majority voting is expected to strengthen the role of the European Commission and, thereby, facilitate sanctioning.

55 Additionally, the SURE programme (Support to mitigate Unemployment Risks in an Emergency) was set out at the beginning of the pandemic to support national unemployment insurance schemes to pay for short-term work schemes. The programme (around 100 bn EUR) provided loans and, therewith, contributed to a European-level fiscal response in the sense of automatic stabilisers. Moreover, the European Investment Bank (EIB) stepped-up its support and the ESM designed a precautionary instrument, which has not yet been used.

56 Scholars agree that the SGP needs to be reformed before the general escape clause expires at the end of 2023 and, ideally, the reform would reduce complexity of the numerical rules and their measurement, simplify exception clauses, address the procyclicality of the rules in good times and the very high debt in some member states (see overview in EU Independent Fiscal Institutions 2021). The idea is to make the rules more easily assessable, also for the public to increase public pressure. Scholars suggest moving to an expenditure rule to reduce issues of pro-cyclicality, discouraging investment and measurement problems of other alternatives (Fuest and Gros 2019). The debate seems to direct towards targeting rather the debt than the deficit level, handing-over the responsibility for the implementation to the domestic level, and increasing oversight by independent fiscal institutions (similar suggestions have been made by de Grauwe 2007).

57 The idea for such an EU-wide fiscal response was preceded by the European Fund for Strategic Investments (EFSI), the so-called 'Juncker-Fund' that was established in 2015 and aimed to leverage private funding for investment projects. It was administered by the EIB. The idea for

in Germany, consider the NGEU as Eurobonds through the back door (Höfling 2021) and together with the other pandemic support measures by the ESM and the ECB the 'end of Maastricht' division of competences (Heinemann 2020). With this, the positioning of the SGP within the EMU institutional design has changed significantly over the years, and is still ongoing. This debate appears crucial to understanding the perceived meaningfulness of the SGP by member states, and their inclination to follow the rules, which will be discussed below under 'ownership' issues.

2.3.2 Reasons for rule adherence

Despite there being a great deal of literature on the SGP, only comparatively few studies focus on explaining reasons for rule adherence. As said above, many studies focus on why the institutional design is flawed, that enforcement is necessarily weak as institutional incentives do not ensure effective peer pressure, and that it ultimately rests on the political willingness of sovereign states. In the following, the review compiles competing explanations for both national- and European-level compliance. For the national level, I further distinguish between domestic considerations for adherence with the deficit rule and strategic considerations based on the perceived threat of getting sanctioned. This refers to incentives stemming from the institutional design and from the European counterparts' willingness to sanction. With this, strategic considerations might outweigh domestic considerations.[58]

In the next sections, I conceptualise the literature around two themes: 'capacity' and 'ownership'. The first relates to a member state having 'ability' in the sense that its conditions and the given situation determine the scope of action and enables compliance to a particular extent. The latter refers to 'political willingness' as the purposefully formulated response by the government. Accordingly, 'capacity' issues appear rather difficult to change, while 'ownership' issues could change in the event of a favourable political majority.

As for capacity, the literature identifies as the most prominent factors the business cycle as a time-variant determinant of a member state's capacity to adhere to

promoting reforms instead of sanctioning the absence of reforms was preceded by the euro Plus Pact in 2011 and the Commission's Reform Support Programme in 2017 (see Dolls et al. 2019).

58 Blavoukos and Pagoulatos (2008) refer to Putnam-inspired 'two-level' negotiations about the reform of the SGP in 2005, and show the strategic nature between both levels.

the rules and a member state's economic conditions and 'size', i.e., strength, as time-invariant or very slowly changing factors. For instance, in economically difficult times or for small, developing or specialised economies, i.e., that happen to be more vulnerable in specific shocks, it is more difficult to comply with the deficit rule. Moreover, the literature assigns differences in general country dispositions, alongside a 'North-South' divide among member states. Additionally, 'capacity' depends on the institutional design's incentives, which might trigger strategic considerations and shape domestic preferences. With this, institutions also determine 'capacity', which could be leveraged politically by the respective perspective on 'ownership'.

'Ownership' of the rules is a prominent theme in the literature, but remains rather vague in terms of its specific materialisation. The overall claim is that it lacks political willingness to adhere to the rules, but the reasons can be manifold. For instance, 'ownership' is mentioned to explain the lack of enforcement at the European level: the rules are said to be lacking a clear ownership by a dedicated body, and the members in the Council do not consistently assume ownership as a collective body given diverging national interests or disregard (Schuknecht 2005). 'Ownership' has also been referred to as a lack of political willingness to implement the deficit rule at the national level (Begg 2017, European Commission 2021). Moreover, this is considered to be more prominent in some member states than in others, and is subsumed under the 'North-South' divide, reflecting policy traditions (cf. Magone et al. 2016). In this thesis, I argue that 'ownership' is linked to economic policy ideology. With constraining public finances, the SGP breathes an ordoliberal spirit, towards which other perspectives such as, for instance, the Keynesian one might find it difficult to take ownership. Chapter 3 will discuss this in detail.

The table below summarises the literature's explanations. Most of them focus on breaches (marked in orange) and the national level (marked in light blue). In comparison, there are only a few on compliance (marked in green) and the European level (marked in blue). This review tries to sketch the most prominent explanations to satisfy the principle of parsimony (see Hancké 2009: 24f.).

Table 2.3-1: Summary of the literature's explanations for rule adherence.

	Aspect	Expectation (to reduce or to increase compliance)	Relevant for compliance at …	
			National level	European level
Capacity	**business cycle**	less compliance	domestic	
	economic conditions:	tight: breach	domestic	x
	- economic development ('wealthy')	wealthy: more compliance	domestic	
	- times of structural reforms	less compliance	domestic	
	- economic spill-overs from others (perceived as positive)	less compliance		x
	- 'North-South' divide (general country dispositions)	south: less compliance	domestic	
	size:	large: less compliance	domestic and strategic	x
	economic and structural arguments:			
	- costs of consolidation (large: high)	large: less compliance	domestic	
	- room for fiscal stimulus (large: high)	large: less compliance	domestic	
	- historical growth rates (small: high)	large: less compliance	domestic	
	- economic role for others (large: high)	Large and small: less compliance	strategic	
	- open vs closed economies & structural reforms and supply-side policies (large: high: rather closed and less supply-side policies needed to increase competitiveness)	large: less compliance	domestic	
	- domestic budgetary institutions (small: commitment, large: delegation)	large: less compliance	domestic	x

	- administrative capacity (small: low)	small: more compliance (voting with EC) or more prone to requests for reciprocal voting		x
	political economy arguments:			
	- vote shares	large: less compliance		x
	- EU law compliance (small: high)	large: less compliance	domestic	
	- domestic power of international organisations (small: high)	large: less compliance	domestic	
	- preference of Council discretion (large) over no or European Commission discretion (medium-sized or small)	large: less compliance; small: more compliance		x
	institutional incentives:			
	- sanctioning procedure being too soft	less compliance	strategic	
	- voting rule	exclude members under EDP: more compliance	strategic	x
	- weighted votes	higher possibility for small states to get sanctioned	strategic	
	- 'volume' of sanctions	more compliance	strategic	
Ownership	ownership (internalise compliance)	more compliance	domestic	x
	domestic structural aspects:			
	- domestic fiscal rules	having fiscal rules: more compliance	domestic	x
	- 'misfit'	less compliance	domestic	
	- electoral cycle	less compliance	domestic	
	- domestic cleavages	less compliance	domestic	

institutional incentives:			
- reciprocal voting	less compliance	strategic	x
- coalition building in the Council (facilitating successful reciprocal voting, i.e., blocking minority)	less compliance within groups	strategic	x
decision-makers' perceptions:			
- 'North-South' divide (policy traditions)	south: less compliance	domestic	x
- government party ideology	left: less compliance	domestic	x
- considering the rules meaningful (economic rationale)	not meaningful: less compliance	domestic	x

Source: own illustration based on the literature review.

2.3.2.1 The role of ‘capacity’

2.3.2.1.1 Business cycle

Among competing explanations, Hansen (2015) finds the business cycle to be the determinant factor for breaching the deficit rule. The effect is also implicitly assumed in the above-mentioned literature on the economic rationale of the SGP that criticises the rules for not sufficiently acknowledging a fiscal response to the business cycle. The reform of the SGP in 2005 was largely welcomed by the academic debate for addressing the temporary country-specific economic situation (Buti 2006, Diebalek et al. 2006, Alves and Afonso 2007, Schelkle 2007, see a comparison of old and new rules on p. 709). The fiscal balance deteriorates in economic downturns due to reduced tax income, increased expenditures for automatic stabilisers and discretionary expansionary fiscal policy. In case consolidation is not a desirable policy option, given its pro-cyclical effects and risks to harm economic recovery in the medium-term, borrowing appears as the only option. Absorbing shocks might be easier for some member states than for others.

2.3.2.1.2 Economic conditions

Tight economic conditions make it more difficult to comply at both levels. If the economy is larger, wealthier, more advanced and more diversified, shocks can more easily be absorbed than in smaller, less wealthy, less advanced and less diversified economies that are more vulnerable in specific sectors and have fewer public resources as leeway to counter shocks. The EU compliance literature shows that compliance at the national level can also be difficult in times of economic distress or structural reforms given that the rules’ constraints might make redistributive policies more difficult (see Börzel 2016). In turn, breaches at the European level, i.e., not voting for sanctions, becomes likely in cases where economic spill-overs from the member state in question are perceived as positive for the domestic economy.

Additionally, the literature identifies differences in general dispositions between member states located in the ‘North’ and those located in the ‘South’, resulting in the latter being more prone to breaching the deficit rule. ‘South’ refers to countries with high inflation, previously soft currency traditions, mixed-market economies with an economic growth model being demand- instead of export-led (according to ‘varieties of capitalism’, Hall and Soskice 2001, Hancké et al. 2007, Hall

2018, Johnston and Regan 2016), and which were "over-indebted" debtors as opposed to "competitive" creditors during the sovereign debt crisis (quotes from Merler and Pisani-Ferry 2012: 1, Matthijs and McNamara 2015, Magone et al. 2016). They had to adjust the most to the Maastricht criteria, with the respective reduction of heterogeneity and harmonisation of business cycles, and are accordingly referred to as belonging to the 'periphery' as opposed to the 'core' of similar countries to form the euro area (Bayoumi and Eichengreen 1993, Verdun 1998: 119,[59] Hix and Høyland 2011: 253) and related windfall gains from euro area membership (Pérez 2019). This also resonates in the EU integration literature that classifies such member states as belonging to the 'periphery' in contrast to the 'core' of countries leading the EMU integration process, such as France and Germany (Heipertz and Verdun 2005: 987, similarly for more integration: Simms 2010). The distinction is also mirrored in the countries' net gains from the EU budget as 'beneficiary' as opposed to 'contributor' member states (Koehler and König 2015, Magone et al. 2016), and in the Council literature's dimension of preferences towards market regulation: "The poorer southern member states prefer extensive regulation of the market, whereas the northern member states tend to support free trade and competition" (Kaeding and Selck 2005: 282). In addition, some member states from the South also struggle with comparably low quality of institutions in general (Sotiropoulos 2004, see, for instance, the case of Greece, Featherstone 2015, or Italy, Quaglia and Maes 2004). Hall (2018) notes that Mediterranean market economies such as Spain, Portugal, Greece and Italy "lack … institutional capacities" to "[operate] the kind of export-led growth models seen in Northern Europe" (p. 11). As a result, scholars assume genuinely different levels of capabilities or preferences for the continuous implementation of sustainable public finances and compliance with the SGP. This also resonates in respective preferences for EMU reform: "In a nutshell, a coalition of predominantly Southern member states prefers fiscal transfers against the opposition of Northern member states, who support fiscal discipline" (Kudrna et al. 2021: 6).

Distinguishing euro area member states into either 'northern' or 'southern', also depends on the perspective and might change over time (de Grauwe and Ji 2018). The distinctions agree with locating Greece, Spain and Portugal in the 'South' and

59 "A core group of countries, i.e., Austria, the Benelux countries and Germany, could possibly form an OCA. But certainly countries such as Ireland, Italy, Portugal, Spain and the UK would not form an OCA with the core group nor with each other" (Verdun 1998: 119).

Austria, Belgium, Finland, Germany, Luxembourg, and the Netherlands in the 'North'. Italy, Ireland and France are particular cases. The EMU integration literature locates Italy and France in the 'core' group of countries. In contrast, the economic literature locates Italy in the 'South', whereas France is a peculiar case. It does not have a tradition of high inflation or a soft currency, but the reliance on exports is comparatively low and accordingly, Hall (2018) locates France in the group of Mediterranean market economies, whereas Johnston and Regan (2016) argue that the "wage developments conform more to those of its export-driven northern neighbours" (p. 324) and classify it as 'northern'. Johnston and Regan (2016) also classify Ireland as 'oscillating demand/export economy' because of several political turns since the end of the 1980s. Magone et al. (2016) summarise several economic and political science distinctions, support the distinction above, and also assign other member states that joined the euro area at a later stage as being either "highly" or "less developed economies" (p. 1). Since this thesis focuses on the economic aspects of the member states and not the political aspects to advance European integration, this thesis follows the economic distinction with Italy belonging to the 'South' and France to the 'North' (following Johnston and Regan 2016, also see discussion in Section 3.2.5).

2.3.2.1.3 Size

Large member states are more prone to breaching the rules at both levels than small or medium-sized ones given their relatively higher economic and political capacities (see Buti and Pench 2004, Herzog 2004, Hansen 2015). Given the track record of France, Italy and Germany, Buti and Pench (2004) conclude that large countries breach the deficit rule more often than the others (with the exception of Greece and Portugal). While they present a neat discussion of factors, they do not come to a conclusive hypothesis of why 'size' could be explanatory. Economic arguments point to the larger costs of fiscal consolidation of large states, which might make it too difficult to implement them, in particular in economic bust times, resulting in a breach. Large states also have more room for fiscal stimulus, which might make it more effective in bust times. Buti and Pench (2004) note that they might confound the role of 'size' with the differentiation between open and more closed economies, where the latter are usually large ones (with the exception of Greece). In general, open economies feel a greater need to follow structural reforms and supply-side oriented policies to increase competitiveness. Chang (2006) argues that "the smaller states possessed less

autonomy due to their size [already before joining the euro area], so the restrictions of the SGP are less constraining than for large states" (p. 114). Moreover, Buti and Pench (2004) note that the observed track record might rather relate to historical economic growth rates, which make consolidation easier, and which correlate with 'size' in the early 2000s: While Germany and Italy have seen low rates (with France ranging a little higher), most small states have seen much higher rates.

From a strategic point of view, large states could leverage their economic role for others. For instance, if a large country serves as an 'economic anchor' for other members, these members could refrain from sanctioning deficits used for fiscal stimulus as they expect to benefit from positive spill-overs (Giuliodori and Beetsma 2004, Herzog 2004, in't Veld 2013). Small countries may benefit relatively more, as well as being the ones that suffer more in an economic downturn of a large member. In turn, a larger state could better manage to internalise negative spill-over effects (Beetsma and Bovenberg 2000: 249). In contrast, small members might benefit from their potential negative-spill overs being 'negligible' for the whole euro area economy (Begg 2003) and choose to breach the rule.

Structural arguments point to the role of domestic budgetary institutions for compliance with the deficit rule and administrative capacity for compliance at the European level. States in which the budgetary competence is delegated to the finance minister, usually experience more discretionary fiscal policies, as opposed to states in which (usually coalitions) governments sign a commitment for budgetary policy beforehand that constrains government action (Hallerberg et al. 2007). Large states usually belong to the former and small ones to the latter (see discussion in Buti and Pench 2004).[60] Another argument relates to administrative capacity (Hodson 2009, Panke 2010, Börzel et al. 2010). If a member state does not have enough resources, or has an inefficient bureaucracy to monitor all other member states, it needs to rely on the European Commission assessment. This could result in European-level

60 Eyraud et al. (2017) also find that large countries are more prone to execution slippages. Beetsma et al. (2009) discuss the finding that those countries in which an independent body produces the official growth forecast show "unbiased growth forecasts" (p. 791) in contrast to states where the forecast is made in the finance ministry. Von Hagen (2010) argues that projection errors are the result of strategic action by the finance minister to avoid sanctioning by the European Commission. He finds that although governments with 'delegations' tend to produce overly optimistic forecasts, they are better at offsetting revenue losses during budget implementation. In contrast, governments with 'contracts' and strong fiscal rules tend to produce overly pessimistic forecasts, but do not manage as well to offset changes during the implementation process.

compliance, i.e., voting for sanctions based on the Commission proposal. However, it might also make such members more prone to requests for reciprocal voting.

Political economy arguments point to the political role of large states. Börzel et al. (2010) find that politically or economically more powerful member states are more likely to violate EU law in general because power makes states "more resistant to external pressure because they have more alternatives to co-operation […] and can more easily pay for reputational or material damages" (p. 1368).[61] Similarly, in large states, where the power of domestic actors prevails, international organisations do not matter as much as in small states (von Hagen 1998: 35 as cited in Buti and Pench 2004: 1029, de Haan et al. 2004). Accordingly, Schure and Verdun (2008) find that large member states prefer Council discretion over defining strict rules. In contrast, medium-sized and small member states, which do not see themselves as having much to say in the Council because of their relatively small vote share, prefer either strict rules or the delegation of discretion to the European Commission. In the context of the SGP, this would mean that large states would prefer a flexible interpretation of the rules and would rather breach the rules. Kleine (2014) describes this as 'informal governance' and finds that small states leave large states more room (even for branches) in order to keep them within the 'institution' (i.e., SGP rules).[62]

2.3.2.1.4 Institutional incentives

Strategic considerations for compliance with the deficit rule stem from a low threat of punishment resulting from the SGP's institutional design. While a low threat of punishment represents a capacity issue which defines the scope of action of a member state for national-level compliance, establishing a credible threat of punishment rests on ownership and compliance at the European level. The section below will discuss the latter. The "lack of credible incentives as much as prospective sanctions" (Begg 2003: 8) would induce non-compliance. The literature identifies several aspects of the institutional design that might in fact result in disincentives, and facilitate breaches through a reduced threat of punishment: a lack of enforcement and the costs of

61 The literature on how the Council operates supports the political role of large member states and finds evidence for them being more likely to voice opposition in the Council legislative process (Mattila and Lane 2001, Hagemann 2006, contributions in Naurin and Wallace 2008, Hosli et al. 2011, Warntjen 2017).

62 Dijsselbloem (2021), former Eurogroup president and Council member for the Netherlands, shares this perspective.

sanctions being too negligible. The former relates to the sanctioning procedure being considered too soft and the voting rule facilitating reciprocal voting. Both hold in particular for large states given the weighted votes. Moreover, the lack of enforcement is attributed to the reluctance of the Council to sanction peers (detailed analyses in Schuknecht 2005). As peers are asked to decide over peers, this risks to create co-dependence between the members. Such co-dependencies could fuel positive externalities and, thereby, breaches. In this case, a majority of favourable members in the Council is needed to exert sufficient peer pressure in the spirit of the SGP (Schuknecht 2005, similar expectation in Buti et al. 1998 and Bräuninger 2004). Whether the Council endorses one or the other perspective also hinges on whether members assume ownership (see Section 2.3.2.2).

The sanctioning procedure has been criticised[63] as being too soft in the sense that "the SGP suffers from inherent legal weakness. Due to the politicised nature of the EDP, the essence of the Pact seems to be not so much a mechanism of 'quasi-automatic sanctions' but rather the institutionalisation of a political pledge to aim for low deficits" (Heipertz and Verdun 2010: 6f.).[64] This also rests on ownership and enforcement at the European level. Additionally, the voting rules do not exclude members that are currently in an EDP, i.e., breaching the deficit rule.[65] This facilitates reciprocal voting among such members to avoid sanctions (so-called 'sinners' solidarity', see Irlenbusch et al. 2003, Irlenbusch and Sutter 2006): "It is hard to imagine that member states on the verge of breaching the deficit criterion sometime in the not-so-distant future will take a tough stance with regard to those countries" (de Haan et al. 2003 as quoted in Buti and Pench 2004: 1029).[66] Moreover, weighted votes

63 This debate refers to 'soft versus hard law' to classify the SGP's elements and understand their weaknesses (Begg 2003, Begg and Schelkle 2004, de Haan et al. 2004, Hodson and Maher 2004). The multilateral surveillance is considered soft law, whereas the EDP is considered hard. Without the 'soft' part, however, there will not be 'hard' sanctions. Soft law is not necessarily weak, but requires a strong commitment from the 'peers' to exert pressure. Enderlein (2004) argues that soft law is the more efficient tool for co-ordinating fiscal policies as it retains member states' autonomy, and would accordingly increase their willingness to co-ordinate. Hodson and Maher (2004) argue that soft law is "the best form of governance for economic co-ordination, especially given the uncertainty over measurement of medium-term fiscal balances" (p. 799).

64 In contrast, Schelkle (2007) argues that the reform in 2005 "is more effectively constraining countries that are officially in 'excessive deficit'" (p. 705) because the wording of the rules was made more applicable.

65 This was suggested by Otmar Issing at an EU Summit in 1996 (as described in Irlenbusch et al. 2003: 647).

66 Buti and Pench (2004) also discuss de Haan's argument (de Haan et al. 2003) that "the probability of being sanctioned depends inversely on the number of countries that breach the SGP rules" (p. 1029; also see the later version of this paper de Haan et al. 2004). This means that the more member

are associated with unevenly favouring large over small member states as they have greater leverage (through more votes) to pass their preferences in the Council (Buti and Pench 2004).[67]

While the probability of enforcement determines the likelihood of getting sanctioned, the volume, type and timing of sanctions determine the costs of a sanction. Ohr and Schmidt (2003) find that the necessary level of costs of sanctions need just be slightly higher than the gains from a breach for the incentivising effect to unfold.[68] The steps of the EDP vary regarding the level of 'costs'. Only the last steps represent pecuniary costs, while the other steps are of a political nature, and could represent a signalling effect to the public, notably to the electorate and to financial markets about the state of public finances (Kalan et al. 2018). De Haan et al. (2004) show that the "perceived size of the SGP's penalty" (p. 251) differs among members of different 'sizes'. Large countries would find sanctions as too low in comparison to the gains from fiscal autonomy. Vice versa, small countries perceive the sanction, which also includes a loss in political reputation, as high and would comply accordingly.[69]

2.3.2.2 The role of 'ownership'

The academic debate's most often named explanation for why a member state would not vote for sanctions, i.e., comply at the European level, is the lack of ownership of the rules (Buti 2006, European Commission 2021).[70] The idea is that a member endorses the rules and acts upon them 'by default'. The lack thereof is attributed to

states are in an EDP, the less likely it is that the Council takes a tough stance on sanctioning another member. While this observation highlights the role of reciprocal voting, and this might have been the case for the years until 2004 (the year of the publication of the article), the year 2008 shows that all member states who breached the deficit rule were put under an EDP. Therefore, it seems that rather the opposite is plausible: As long as it is only a few states that breach the rule, they might be let off the hook. However, if it is many, they would all be sanctioned because of the risk of losing credibility to keep their 'euro area house' in order.

67 Baerg and Hallerberg (2016) show that large countries are more successful at weakening the European Commission's assessments of their annual stability and convergence programmes where they set out the fiscal planning and reforms. However, Eyraud et al. (2017) do not find any indication that the economic size of a member state relates to the length of an EDP.

68 Additionally, the 'volume' of the sanction needs to grow as the probability that a sanction is not realised increases. Eijffinger (2003) argues that the credibility of sanctions also hinges on "the threat of uncertain and delayed sanctions" (p. 11).

69 Similarly, von Hagen (1998) and de Haan et al. (2004) argue that both the pecuniary as well as the political aspects of sanctions would not affect large states as much as small states due to the proportionally lesser "importance of international organisations in domestic politics" (Buti and Pench 2004: 1029).

70 This was also the main argument in the political debate at the time for introducing the Fiscal Compact so as to anchor the SGP rules in domestic constitutional law.

both domestic issues and institutional characteristics that establish the dual-hat that each member in the Council wears as both judge and potentially the one being judged at the same time.[71] At times, domestic interests might run counter to European interests, and this might create moral hazard and induce free-riding, i.e., breaches.

It is generally assumed that the role of ownership for compliance is transmitted from the domestic to the European level. This relates to structural and political arguments. For the former, member states that have fiscal rules in their national legislation are less likely to breach the SGP rules, as they are used to fiscal constraints and endorse their usefulness (cf. Wierts 2008). Even lower-level fiscal institutions, such as delegation or commitment, would serve such purpose (Hallerberg et al. 2007). Furthermore, a policy-specific mismatch between EU law requirements and the member state's current situation, known as 'misfit' in the EU compliance literature, can explain implementation slippages.[72] Such a situation could also be politicised to avoid compliance. Similarly, politically leveraging capacity issues such as the electoral cycle (Hauptmeier et al. 2010, Heipertz and Verdun 2005: 991ff.),[73] or domestic cleavages (Eslava 2011)[74] within the society or the economy can lead to 'ownership' issues.

The literature on how the Council operates supports the perspective that positions which the member states take in the Council originate from domestic preferences (Hagemann et al. 2017, Lewis 2003, Thomson 2011, Høyland and Hansen

71 Scholars have described this in a principal-agent framework and it illustrates the risks of moral hazard due to conflicts of interest and asymmetric information (see overview in Hodson 2009, Schuknecht 2004, Schelkle 2005, Savage and Howarth 2020). The Council can be considered the principal that contracts the agent to perform sustainable public finances. However, being a member of the Council, a member state is agent and principal at the same time.

72 The literature on compliance with EU law in general suggests that domestic objections, be it due to current conditions in the country or political positions of relevant actors, explain delayed or insufficient compliance. The literature suggests either 'veto points' with regard to the domestic policy process (Haverland 2000), or a policy related 'misfit' between domestic policies and the EU law (Thomson 2009) and additionally the efficiency of a member state's administration (Falkner and Treib 2008).

73 Hauptmeier et al. (2010) find that the expenditure ratio increases in the run-up to elections and that deficit spending correlates with the political business cycle (political opportunism). Heipertz and Verdun (2005: 991ff.) show that this was the case for France and Germany in 2003: "the governments had made promises to the electorate that were not in line with the SGP – predominantly tax cuts" (p. 994f.).

74 Eslava (2011) provides a good overview about domestic political cleavages and concludes that the more heterogenous preferences are in the society and the political process, the more likely is it to resolve them with higher deficits. In turn, she notes that "systems where the fiscal decision-making process is more centralized and/or political contexts characterized by less fragmentation are conducive to greater discipline" (p. 665). She adds that recent literature on budget institutions and fiscal rules appear to limit the conflicts, despite the prevalence of heterogenous actors.

2014) and also from strategic considerations (Lewis 2003, Hagemann 2006, Thomson et al. 2009, Aksoy 2012, see discussion in Nugent and Paterson 2010). It is generally argued that domestic interests are rather followed than European interests (Eyraud et al. 2017, the EU compliance literature describes this as 'misfit', Hagemann et al. (2017) find responsiveness in the Council, Schure and Verdun (2008) describe one exceptional case of the Netherlands voting in 2003 against this claim). Accordingly, bilateral reciprocal voting and 'coalition building' for such a purpose would impede assuming ownership in the Council. The literature finds diverse indications for 'coalition building' in general (see contributions in Naurin and Wallace 2008, see a discussion in Bailer et al. 2014),[75] however, it has not yet found a cohesive definition, suggesting that coalitions depend on the policy area (Thomson et al. 2004, Hayes-Renshaw and Wallace 2006: 259-279, Thomson 2011).[76] Some studies find a 'North-South' divide (Kaeding and Selck 2005, Zimmer et al. 2005, Thomson et al. 2009, Mattila 2009, Thomson 2011). Similar to the above, this reflects different preferences for having public finances constrained based on domestic policy traditions. The existence of a left-right dimension is contested, with some studies finding evidence (Hagemann 2006 and 2007, Mattila 2009, Hagemann and Høyland 2008, Kreppel 2013) and others not (Zimmer et al. 2005, Bailer et al. 2014), while Hosli et al. (2011) only find that this dimension is relevant for old member states (before the enlargement in 2004). At the domestic level, conventional wisdom supposes that government party ideology affects budget balances with left-wing ones being associated with spending

75 Despite cleavages, the literature identifies an overall search for consensus in the Council. Interactions in the Council are seen as 'repeated', and thereby fuel a desire for consensus, or at least consensual behaviour among the members. Therefore, despite the fact that in some policy fields where the treaty prescribes qualified majority voting, the members still seek unanimity (Lewis 2000, Mattila and Lane 2001, Hayes-Renshaw et al. 2006, for a discussion see Smeets 2016, Häge 2012). This also seems to be the case in the SGP framework, as no voting records are published but decisions are signed by the president for the Council. Even more, the Eurogroup as the relevant Council composition for the euro area, is described as a rather confidential body (Puetter 2006, Puetter 2012, Hodson 2020). Nonetheless, this does not imply that voting rules and voting weights would not matter. The opportunity to eventually use them presents the default situation and, thereby, shapes expectations. Novak (2013) argues that consensus serves a 'blame avoidance strategy': "Formal voting would disclose the identity of opponents and open them up to criticism" (p. 1092), both in the Council but also in the domestic arena.

76 Some studies find that it seems to relate to categories such as old and new member states (Thomson 2011, Hosli et al. 2011), contributor and beneficiary states (Zimmer et al. 2005, Thomson 2011), or pro or anti-regulation ones (Kaeding and Selck 2005, Thomson 2011). For instance, Koehler and König (2015) find that the SGP managed to decrease debt levels in contributor countries in contrast to beneficiary countries. Hansen (2015) finds that Eurosceptic governments are more likely to breach the deficit rule, and Baerg and Hallerberg (2016) show that member states with a Eurosceptic population manage to weaken the European Commission's recommendation on stability and convergence programmes.

more, and right-wing ones tending to consolidate more.[77] However, as Cusack (1999) shows, this claim has found little empirical evidence and, in contrast, party ideology operates in a more complex environment depending on economic conditions. Hübscher (2016) shows that during a crisis situation, both are equally likely to implement spending cuts, while right-wing governments cut more than left-wing ones. One can assume that party ideology might be too narrow a concept to explain SGP compliance in general.

I argue that ownership issues at the domestic level, as well as the European level, also relate to endorsing the economic rationale of the SGP rules. Considering the rules as meaningful and reflecting 'good' economic policy might help members to assume ownership. Given the fact that the SGP is a 'disciplinarian device' (Schelkle 2005) that does not prescribe easily tangible rewards, the members have to revert to the very purpose of the SGP as establishing a 'reward'. What can be considered a 'reward', as well as costs associated with compliance, have been subject to intense discussions ever since the establishment of the SGP. At the beginning, the purpose of constraining public deficits was to avoid inflationary pressure and pressure on the ECB to monetise sovereign debt (see Eichengreen and Wyplosz 1998, Beetsma and Uhlig 1999, Lindbeck and Niepelt 2005, and Heipertz 2003 for a discussion).[78] With the sovereign debt crisis, the risk of sovereign default became striking with the economic and political repercussions for the euro area.[79] Most recently though, the risk of monetising debt became apparent, with the ECB finding it difficult to end its accommodative monetary policy, which has helped member states' borrowing costs (Belke and Gros 2021, Havlik and Heinemann 2021). This shows the many facets of the SGP's rationale. I argue that sharing the idea of constraining public deficits, and the believed purpose associated with it, establishes ownership. Similarly, in the EU

77 While Hansen (2015) finds that left-wing governments are more likely to breach the debt rule than right-wing governments, he finds no effect on the deficit rule. However, this seems rather to point to country traditions as the debt level reflects legacy, and the deficit level reflects the current government's priorities. Similarly, Bräuninger (2005) finds that despite differences in preferences, spending outcomes of left-wing parties do not significantly differ from those of right-wing parties.

78 The argument is that government budget deficit and debt could cause negative spill-over effects for other countries in a monetary union. These effects might be of an economic nature (demand or supply effects), or a financial nature (distrust of investors, higher interest rates for loans or risk premiums), or affect currency stability (inflation, de- or appreciation of exchange rates) and therefore have an impact on trade flows, the economic development or the budget situation in a member state. The inflation argument was much contested at the time (see Sims 2012).

79 In addition to the risk of default, the role of public debt on economic growth was a prominent debate at the time (Reinhart and Rogoff 2010, Baum et al. 2012).

compliance literature, Börzel et al. (2010) suggest 'acceptance', which Mastenbroek and Kaeding (2006) call 'beliefs' and conclude that properly examining preferences and beliefs can be considered intervening variables between EU law and domestic policies, and more accurately explain 'misfit'. The next chapter will discuss this in more detail as the theoretical basis for this thesis.

2.4 Summary

With a view to the SGP track record (see table below), each of the business cycle, the 'North-South' divide and size cannot explain the whole picture of rule adherence. I pick these explanations to contrast them with the data from Section 2.2 as they are rather easy to compile.[80] Similar to Section 2.2, the period is restricted to 1999-2008, from the non-reversible peg of national currencies to the euro until the unfolding of the financial, economic and sovereign debt crisis. Simply looking at the curves of GDP growth (see figures below) show that economic bust can be associated with higher deficits, especially in 2008. However, 2008 is a special case as the crisis was already at the door step. For the other years, there are also exceptions: for instance, Spain managed to reduce the deficit despite reduced growth rates, while deficits in Italy increased despite increasing growth rates. Moreover, the table shows that there are more cases for which growth decreased rather than increased compared to the previous year, which, however, did not result in a breach.

Secondly, there appears a clear difference according to 'size', though not fully in line with the literature. Indeed, large states show more cases of breaches, and they were proportionally also less often sanctioned than the others. Only with regard to revised deficits are there more cases of breaches for medium-sized countries, which are predominantly attributed to Greece and Portugal. With the exception of Spain, all large countries breached the deficit rule. Medium-sized countries complied, with the exception of Portugal and Greece, which (almost) never complied to the deficit rules,[81] and with short periods of breaches by the Netherlands and Austria. Small countries have never breached the deficit rule, except in 2008. As a result, while the data

80 Checking also the other explanatory factors would require compiling a genuine dataset.
81 Portugal had a deficit of just below 3% of GDP in 2007.

supports the literature's findings for large member states, it does not necessarily support the findings for small and medium-sized member states.

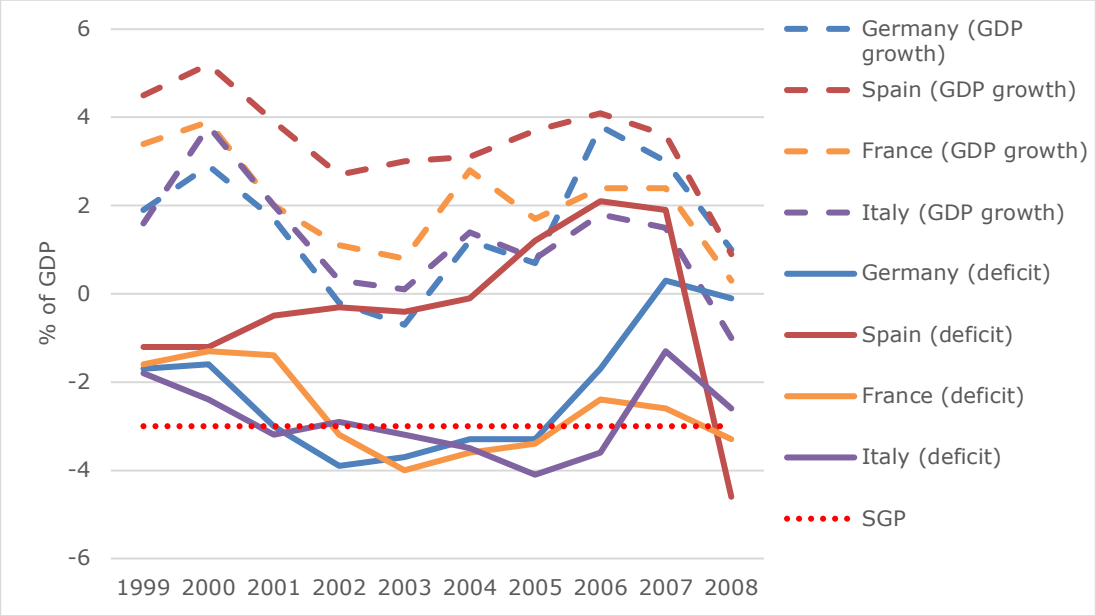
Thirdly, the data show little indication for the 'North-South' divide. With France being a special case, northern and southern countries have practically the same numbers of cases for deficits and decisions. Only as regards the revised deficits were there many more cases in southern member states. The distinction only seems to hold for medium-sized countries where Portugal and Greece breach the rule (South) and the others comply (North). It does not hold for small countries, with Malta and Cyprus complying despite belonging to the South. Neither does it hold for large member states because Germany (North) and France (North) breach the rule (but only during an economic downturn), and Spain (South) complies (which however also had much higher growth rates than the other large states).

Table 2.4-1: Compliance track record and the literature's main explanations per country – 1999 to 2008.

Country	Original deficit	Revised deficit	Council decision	GDP growth less than prev. year	Size	North-South
Germany	4	5	2	6	L	North
France	4	5	2	5	L	North
Finland	0	0	0	4	S	North
Italy	3	5	1	6	L	South
Luxembourg	0	0	0	5	S	North
Spain	1	1	1	4	L	South
Greece	5	8	4	5	M	South
Slovenia	0	0	0	1	S	South
Ireland	1	1	1	6	S	North/South
Portugal	2	9	1	6	M	South
The Netherlands	1	1	1	4	M	North
Austria	0	1	0	5	M	North
Belgium	0	0	0	4	M	North
Malta	1	1	1	1	S	South
Cyprus	0	0	0	1	S	South
Total	**22**	**37**	**14**			
Cases with GDP growth less than previous year and …						
… with deficit	22	37	14			
… without deficit	41	26	49			
Cases according to size						
L	12	16	6			
M	8	19	6			
S	2	2	2			
Cases according to North-South classification						
North	9	12	5			
South	12	24	8			

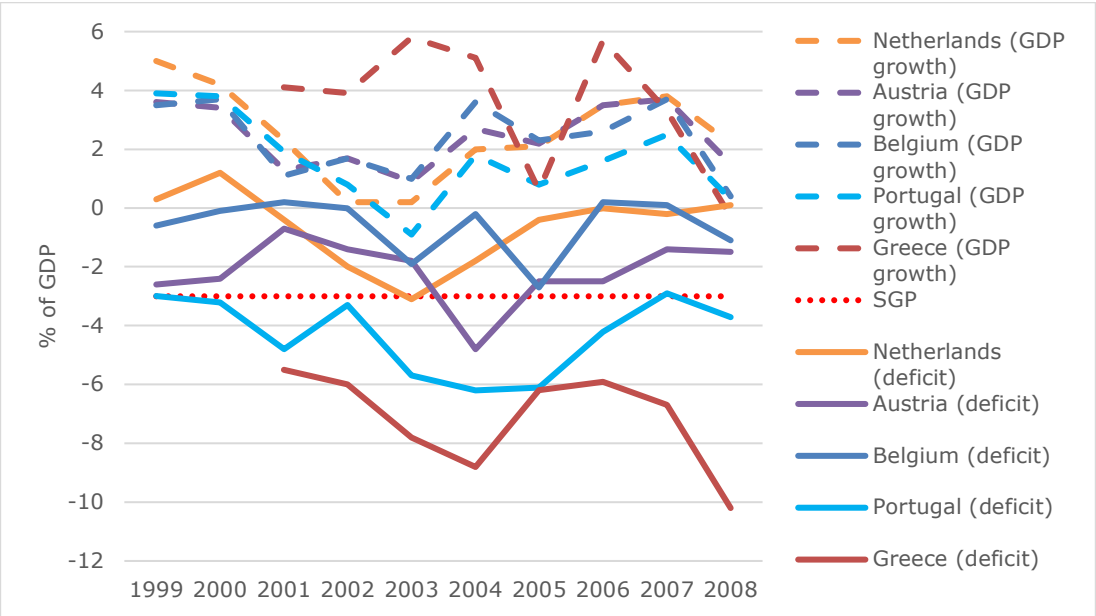
Source: column 1-3: see figure 2.2-3; column 4: Eurostat 2021c; column 5 and 6: literature above. The table shows the number of country-year cases.

Figure 2.4-2a: Deficit and GDP growth in large member states – 1999 to 2008.



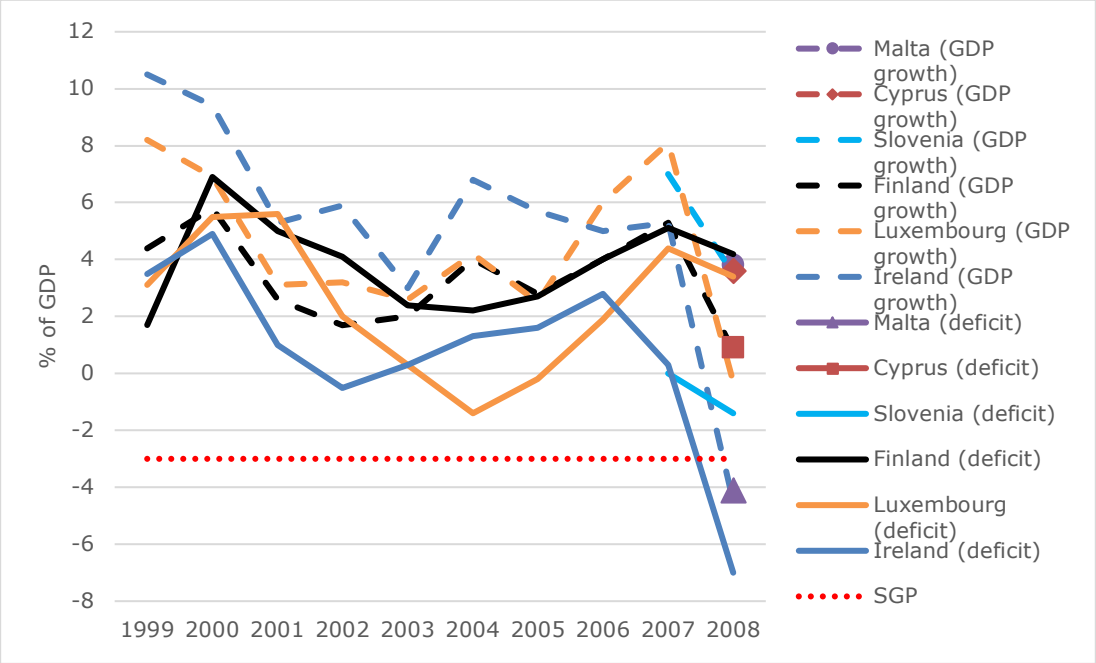
Source: see Figure 2.2-1, Eurostat 2021c. Deficit levels are displayed as % of GDP (solid lines) and GDP growth as year-on-year growth in % of the previous year's GDP (dashed lines). The red dotted line shows the SGP's reference value of -3% of GDP.

Figure 2.4-2b: Deficit levels and GDP growth in medium-sized member states – 1999 to 2008.



Source: see Figure 2.2-1, Eurostat 2021c. Deficit levels are displayed as % of GDP (solid lines) and GDP growth as year-on-year growth in % of the previous year's GDP (dashed lines). The red dotted line shows the SGP's reference value of -3% of GDP.

Figure 2.4-2c: Deficit levels and GDP growth in small member states – 1999 to 2008.



Source: see Figure 2.2-1, Eurostat 2021c. Deficit levels are displayed as % of GDP (solid lines) and GDP growth as year-on-year growth in % of the previous year's GDP (dashed lines). The red dotted line shows the SGP's reference value of -3% of GDP.

It is not so easy to compile data to contrast the literature's other explanations for national-level compliance such as economic conditions, domestic structural aspects, as well as decision-makers' perceptions. This would require in-depth studies. The role of institutional incentives cannot be discussed using observational data as there is no counterfactual data. Moreover, as noted before, there is no data on European-level compliance to contrast them to the literature.

Summing up, the literature puts forward capacity and ownership issues to explain breaches at the national and the European levels. However, it seems that there is no single predominant explanation but rather a combination of several aspects, and I suggest adding another aspect to substantiate 'ownership' by introducing 'economic policy ideology'. The literature suggests that member states breach the deficit rule in economically bad times, economically unfavourable general conditions, structural conditions that might result in fiscal deficits, in the case of large countries, or decision-makers' perceptions. Moreover, strategic considerations stemming from the SGP institutional design might lead to breaches at both levels. Breaches at the European level are associated with reciprocal voting, especially within 'coalitions', positive externalities from the member state in question, to large member states and a lack of

ownership due to having no experience with domestic fiscal rules and relevant decision-makers' perceptions. Compliance is associated with assuming ownership, but the reasons for this remain vague. The literature focuses on explaining breaches, while compliance is expected to happen as a default. However, this disregards, for instance, principled compliance, or compliance despite a difficult economic situation. Contrasting the explanations with descriptive data shows that they seem to interact. Therefore, it is important to study the competing explanations together and disentangle their effect. Moreover, this review shows that there might also be an aspect missing. This thesis introduces the role of 'economic policy ideology' as an aspect of 'ownership'. The next chapter discusses the theoretical concept.

3. Theoretical background: breaches as a collective action problem and economic policy ideology as a link to 'ownership'

In this chapter, I sketch the theoretical background of the thesis. I conceptualise the SGP as a governance mechanism of a common-pool resource, i.e., the common money's stability. This conceptualises rule adherence with the SGP as a collective action problem in a non-hierarchical governance mechanism. Accordingly, limiting the use of a common-pool resource needs to be considered as 'appropriate', and this thesis suggests 'economic policy ideology' to present different perspectives on what 'appropriate' means. Accordingly, considering common rules as inappropriate could undermine legitimacy and, especially in the case of non-hierarchical governance, fuel breaches. In such a case, the SGP's extrinsic incentives for compliance fail to overrule intrinsic motivation for breaches. Section 3.3 presents the hypotheses drawn from this chapter and from the literature review.

3.1 Governing the common-pool resource 'Euro'

One could argue that the SGP is simply an EU law like any other and can, hence, be considered through the lens of the EU law compliance literature. However, on closer examination, it becomes obvious that the SGP is an exceptional kind of EU law because of two aspects. Firstly, with regard to its content, it targets the member states' budget policy, a critical policy that affects other national policies[82] and requires *continuous,* as opposed to *one-off,* implementation. Secondly, with regard to its structure, the SGP requires compliance at two levels: implementing the budget restriction (national level) and sanctioning non-compliant budgets in the Council (European level). In contrast to conventional EU law, where member states assume the role to implement the law and the European Commission assumes the role to enforce implementation, ultimately with infringement procedures and calling on the Court of Justice, here, the Council plays a role for both implementation and enforcement. This puts member states in a dual position, which could easily create conflicts of interest as the literature review has shown. Collective action problems in

82 Jachtenfuchs and Genschel (2013 and 2016) consider it a 'core state power', which gets slowly regulated at the EU level, but through constraining domestic action and not – as would be usual in European integration – creating a genuine European-level competence.

the Council might impede compliance at the European level and, as a result, interfere with enforcement of compliance at the national level already upfront due to a low threat of sanctioning.

Accordingly, this thesis suggests conceptualising the co-operation as a non-hierarchical mechanism to govern the common-pool resource, i.e., the Euro's monetary stability. In particular, I suggest that the common-pool resource overlaps with each member's private good, i.e., the domestic economy. At times, the maintenance of both goods might represent a trade-off, which could create conflicts of interest and result in breaching the appropriation rules for the common-pool resource. This thesis suggests that in such cases, compliance depends on both rational accounts of current costs and benefits, but also ideological accounts of *expectations* regarding the consequences and implications for future costs and benefits. Sharing such beliefs could nourish 'appropriateness' and explain compliance at the European level. In turn, the lack thereof could explain principled breaches.

3.1.1 The 'Euro's stability' as a common-pool resource

Following the legal justification of the SGP, this thesis assumes that the euro and its stability can be considered a common-pool resource for all euro area member states (see Schelkle 2017: 42ff., discussion in Raudla 2010). Its use through government borrowing is rival and basically non-exclusive (Ostrom et al. 1994: 6). It is rival as the capacity to provide money is limited if the ECB's mandate for price stability and the no-bail-out clause shall be followed (Art. 123(1), Art. 127(1), Art. 282(2), Protocol No. 4 to the TFEU). Non-exclusiveness applies as accession to the euro area is restricted, whereas exclusion is very difficult legally (Athanassiou 2009). A common-pool resource faces the threat of the 'tragedy of the commons' (Hardin 1968): rivalry sets incentives to use the good extensively without fearing expulsion as a consequence, and therefore, the good may vanish. Accordingly, a government may want to pursue expansive fiscal policies, but as part of the euro area, they also want the Euro's stability to persist and, consequently, they want the others to follow sound public finances. Hence, individual rationality differs from collective rationality, called moral hazard (Olson 1965). The SGP seeks to prevent the euro from becoming a 'tragedy of the commons'. It limits budget deficits and debt and, thereby, constrains fiscal policies.

The idea to constrain fiscal policy rests on the assumption that the good is rival.[83] Put in very abstract terms, the idea is that there is a specific 'amount' of stable money and that expansive fiscal policies create demand for money. The limit beyond which such demand risks biasing monetary policy is difficult to determine, especially in a monetary union where member states might 'exploit' the common-pool resource to different extents. The SGP's deficit and debt rule seek to present a shortcut to this question based on considerations for fiscal sustainability, in general. In particular, high levels of debt risk driving money demand due to the equivalently high financing needs. Moreover, political pressure might aim to monetise debt through fuelling inflation. According to its independence, the ECB should not experience any explicit or implicit pressure to focus their monetary policy on sovereign debt (Art. 282 TFEU, Protocol No. 4 to the TFEU).[84] Moreover, the assumption of rivalry also rests on the idea that financial markets are not fully efficient. If they were efficient, member states that follow high deficits or debt would pay higher interest rates on their debt, whereas member states with prudent fiscal policies would not see an equivalent increase in borrowing costs (de Grauwe 2020: 228ff.). This builds on the credibility of the no-bail-out clause (Wyplosz 2013, see discussion in Freier and Ciaglia 2019).[85]

Additional to the threat of an 'over-use' for the existence of a common-pool resource, there are two more aspects that relate to the specific nature of the Euro's stability as a common-pool resource. While for usual common-pool resources, limiting extraction is the same as maintaining and contributing to the good, for the Euro's stability, one can conceptualise the three aspects more granularly. Additional to extraction explained above, maintenance also requires sound public finances in order to preserve the euro area as a whole. Constraining fiscal policy also seeks to avoid liquidity problems or sovereign default that would put at risk the overall existence of the euro area. Liquidity problems due to limited access to financial markets are considered problematic in a monetary union, given the no-bail out clause also between member states (Art. 123(1) and 125(1) TFEU), and the lack of a European-level fiscal

83 If the good were considered non-rival, the ECB would provide a public good. There would be an unlimited 'amount' of stable money. Expansive fiscal policies would not result in crowding-out or inflation. This perspective is shared among scholars adhering to Keynesian ideas. This will be discussed in Section 3.2.

84 The independence is also supported by the TFEU that forbids the ECB from purchasing sovereign debt on the primary market (Art. 123(1)).

85 Moreover, the bank-state nexus impedes full efficiency with, for instance, the regulation of 0% risk weight of sovereign bonds held by banks.

capacity. Still, in the current EMU institutional design for large countries, such problems cannot be easily addressed and risk putting pressure on the ECB. Even more critical is a potential sovereign default with political repercussions for the euro area as a whole. The response of monetary policy would need to be delicate, especially in the period up until default is acknowledged and also potentially the exit of the member state from the euro area would be concluded. This would challenge the ECB's monetary policy, financial stability and market operations. With hindsight from the financial, economic and sovereign debt crisis, the institutional design has changed so as to provide financial support to member states in liquidity problems through the ESM and the ECB's Outright Monetary Transactions programme (OMT). With Greece, the euro area also experienced several instances of a liquidity crunch, and, following a loss of access to financial markets, almost a sovereign default and an exit from the euro area. While through different means these issues could have been avoided, in the end, Greece experienced four years of capital controls (2015-2019) and a large share of official debt from different public sources in the EU (77% of total debt, Public Debt Management Authority 2022).[86]

In addition to maintaining and preserving the common-pool resource, I distinguish as a third aspect the contribution to improving the common-pool resource. The idea is that a good economic development in a member state would also reflect in an 'improvement' of the common-pool resource, i.e., expanding the base for stable money and, therewith, the 'amount' of 'stable money'. The ECB's macro-economic assessment, on which their monetary policy, i.e., the provision of the common-pool resource, is based, relates to member states' economic development.

Therefore, this thesis needs to amend the conceptualisation of the Euro's stability as a common-pool resource with overlapping private goods, i.e., each member state's economy. Economic strength can be considered a private good as it is non-rival and exclusive. The government has the sole legitimacy to tax from and redistribute within their economy. The economies overlap with the common-pool resource as they

86 Most recently, the ESM's managing director Regling (2021) even encouraged those member states with a good credit rating not to reduce its debt levels too quickly as there is ample demand for 'safe assets' in the market. This represents the opposite of the previous stance and shows the delicate balance to draw between monetary stability and financial stability and the increasing importance of sovereign debt on the capital market in a low interest rate environment where also social insurance agencies or pension savings products seek investment.

provide the basis for money supply. At the same time, extraction from the common-pool resource happens to the extent that it is used for the domestic economy. In typical common-pool resource settings, the common-pool resource is exogenous, similar to, for instance, the natural accommodation of fish in a lake. Extraction yields direct returns and members would seek ever more individual extraction. In contrast, in the euro area, member states only extract as much as is needed for their economy, and as much as the financial markets would finance at a price that the electorate considers adequate to be financed through taxes.

Summing up, based on the legal conception of the SGP, this thesis conceptualises the Euro's monetary stability as a common-pool resource that overlaps with the private goods of each member state's economy. The SGP rules seek to limit extraction in order to maintain the common-pool resource and to preserve the euro area as a whole. Additionally, the overlapping reflects the member state's stake in contributing to the improvement of the common-pool resource. This particular conceptualisation of the common-pool resource facilitates studying rule adherence with the SGP rules and the collective action problem between member states. It is noteworthy though, that this conceptualisation rests on the legal foundations of the SGP and is contested by one side in the academic debate. This will be discussed in Section 3.2.

3.1.2 The SGP as a governance mechanism

Based on the above conceptualisation, the SGP can be considered a government mechanism of a common-pool resource (Schelkle 2017: 42ff., Collignon 2007, Diessner 2017).[87] As there is no third-party enforcer, it is non-hierarchical self-governance. Accordingly, the reason for non-compliance is attributed to a collective action problem failing to govern the common-pool resource. As the group of member states is rather large and heterogenous, and the governance mechanism is only built on negative incentives such as sanctions, without any explicitly positive incentives like tangible rewards, this makes self-governance even more difficult (cf. Olson 1965). While Collignon (2007) and Diessner (2017) advocate for more EU centralisation to

87 Some scholars already conceptualised the Euro's stability as a common-pool resource. For instance, Artis and Winkler (1997) explicitly refer to governments using 'stable money' and competing for that among themselves. However, they do not explicitly consider the SGP as a mechanism to govern the distributional conflict.

address the collective action problem, Schelkle (2017) argues that collective action problems are not new, and they have also characterised the previous regime before the EMU. Therefore, Schelkle envisages that there will be further co-operation among states (on the 'political market') to address collective action problems, notably by introducing risk sharing elements to substantiate the EMU's institutional design.

In this thesis, I investigate in the role of a common understanding of the meaningfulness and adequacy of the rules for successful self-governance. While, formally, the SGP meets the requirements for a successful non-hierarchical governance of a common-pool resource (e.g., strict rules, limits, monitoring, sanctions) as suggested by Ostrom et al. (1994), compliance at the European level, and hence, successful enforcement of compliance at the national level, requires a 'common understanding' (Ostrom et al. 1994: 327) among the members that the rules are 'appropriate' (p. 329) to reflect their needs and expected benefits from co-operation. In the absence of a third-party enforcer, the members need to initially agree to and continuously endorse the rules, which limit the use and define the allocation of the common-pool resource. One could consider this self-evident, since through self-governance, the members gave themselves these rules. However, continuous implementation in the long-run might become challenging as governments and the member states' conditions change. Nevertheless, the common experience might also enable the members to solve collective action problems as they arise. Ostrom (1990) notes: "When individuals have lived in such situations for a substantial time and have developed shared norms and patterns of reciprocity, they possess social capital with which they can build institutional arrangements for resolving CPR dilemmas" (p. 184). While Schelkle (2017) could be said to be focusing on 'patterns of reciprocity', this thesis seeks to investigate the role of 'shared norms'.

In order to analyse the relevance of 'shared norms' for the case of the SGP, this thesis uses a distinction between 'extrinsic' and 'intrinsic' motivation. While the former refers to any form of institutionalised incentives that seek to induce compliance through application of the rules, the latter instead refers to aspects that members bring by themselves. This either reflects a favourable situation, which allows a member state to comply as costs are lower than gains, or a conviction that compliance is the 'right' thing to do. The former relates to 'capacity' issues put forward in the literature review

above. The latter relates to ideational perspectives of what constitutes good economic policy, which will be discussed in detail below.

The conceptualisation allows this thesis to distinguish between 'extrinsic' and 'intrinsic' incentives for compliance both at the national and the European level. At the national level, extrinsic incentives refer to a credible threat of punishment. The literature review showed that the above-named aspects of successful rules hinge on European-level compliance. While monitoring is supported by the European Commissions' assessments, the implementation of strict rules and limits is in the remit of the Council, and carries some interpretative discretion since the reform in 2005. It is also, according to the literature, subject to strategic interactions. Moreover, the literature points out that there are no immediate and tangible rewards for compliance (which Olson, 1965, calls 'positive incentives', in contrast to 'negative incentives' such as sanctions). They rest on the maintenance of the common-pool resource only, and this might not represent a sufficiently sizable gain in some cases – in particular, in cases where breaches do not represent a credible threat to the maintenance of the common-pool resource. In some cases, enforcement might not be necessary, as compliance can also be 'incidental', reflecting respective intrinsic motivation. This is the case when governments consider that constraining fiscal policies is the 'right' policy choice – irrespective of the SGP. In conclusion, in the absence of successful enforcement, i.e., extrinsic incentives, intrinsic motivation needs to be high enough so that benefits from compliance are considered to be higher than their costs.

At the European level, in contrast, there are legally no extrinsic incentives. One could argue that there are implicit extrinsic incentives through the financial market, which would penalise supporting unsound fiscal policies. However, voting behaviour in the Council is unknown also to the financial market and, hence, cannot serve a signalling effect for prudent or profligate fiscal policies. Secondly, one could argue that the reputation in the Council, also for other policy areas, could serve as an incentive. However, given the salience of budgetary policies, log-rolling with other policy areas might not be straight-forward. This idea rather supports the literature's suggestion of reciprocal voting for mutual breaches. Thirdly, one could argue that peer pressure for compliance could also incentivise compliance by other member states in the Council. However, changes in the situation and preferences in the member states since the establishment of the rules might also affect the Council composition not to

favour such rules anymore and, instead, to have a majority that prefers breaches. In the absence of peer pressure for compliance, compliance by each member state at the European level rests again on intrinsic motivation.

As argued above, intrinsic motivation also builds on perceiving the rules as the 'appropriate' policy choice. This provides the necessary legitimacy that backs the implementation of rules. With this, it ensures principled behaviour within these rules, and also allows for occasional deviations from the rules in difficult circumstances. However, such a normative backing of state behaviour requires two things. First, such intrinsic motivation also requires favourable 'capacities' to some extent. For instance, they have to be strong enough to ensure that public budgets are robust, so that compliance to the rules does not represent much material cost. The conviction must push the perceived benefits of compliance so high so as to outweigh expected economic and political costs. This could be the case when domestic actors and the public favour such policies and these are salient to them. Secondly, almost all large states would need to demonstrate such conviction. The literature review showed that only a few large member states are needed to form a blocking minority and turn the Council towards their preferences of breaching. Therefore, in a self-governance system, i.e., in the absence of extrinsic incentives, compliance rests on 'intrinsic motivation'.

This conceptualisation neglects the role of the European Commission as a third-party enforcer; it will not be included in this study. Legally, the ultimate political decisions about sanctioning remain with the Council. However, the Commission holds the right to present proposals for decision. Moreover, it embodies an important role for the preparation of the economic analysis and suggestions for policy responses. Member states might rely on the Commission's assessment and not conduct substantial analyses of their own. According to the treaties, the Commission should be impartial in their assessments (Art. 17 TEU, Art. 245 TFEU). Until the reform in 2005, the Commission could be considered a strict evaluator of the data and the rules (see, for instance, Chang 2006: 109ff.). It even brought the Council to court for not enforcing the rules. With the reform, however, the Commission was given discretionary power to assess compliance with the deficit rule. From a political economy perspective, one could say that as much as the Commission is dependent on the Council to endorse its proposals, the Council is dependent on the Commission to present a proposal.

Accordingly, scholars investigated the political character of the Commission and consequently, to what extent Commissioners might refrain from presenting specific proposals (Wijsman and Crombez 2017, Killermann 2018, also König and Mäder 2014). [88] Moreover, there is anecdotal evidence that the Commission tries to avoid filing a proposal and to negotiate with the relevant government beforehand (most recently when Italy announced breaching the deficit rule in 2019). Often reiterated is former Commission president Juncker's response when asked why the Commission would not engage in proceedings with an EDP for France: "because it's France" (Juncker 2016). The role of the Commission changed with Juncker's presidency as he continuously emphasised heading a 'political Commission' (Kassim and Laffan 2019). Therefore, one can assume that the Commission would have filed a proposal if it had expected it to be successful. Instead of having a genuine political role itself, the Commission can rather be considered a 'repeater' of member states' conflicts.

3.1.3 Compliance as a matter of rationality and ideology

Based on the literature on EU law compliance, this thesis considers states as actors when studying SGP compliance. To be precise, I consider the current government as temporal representative of a member state. The government is backed by a parliamentary majority to vote on the budget, and the finance minister votes on applying SGP rules in the Council. This thesis views the SGP through the lens of Institutionalism (overview see Hall and Taylor 1996, Nugent and Paterson 2010, and Schmidt 2010), by which an institutional design shapes government interaction. Building on rational choice accounts complemented by shared 'beliefs' (Sen 1977, Hall 1989, Ostrom 1990, Hodgson 2006), namely 'a shared stability culture' (Artis and Winkler 1997: 51), this thesis assumes that governments act based on rational considerations and ideational perspectives on economic policy issues. Firstly, governments are assumed to act rationally and co-operate internationally on the basis of national interests. When establishing an institution, this is the expression of national interests to be safeguarded at international level, hence a country's national and

88 The academic debate has been criticising the Commission for risking their impartiality. The academic debate even makes several suggestions to build a new impartial body to assess state finances that informs, or at least complements, the European Commission's assessment of public finances. To that end, the European Fiscal Board was founded in 2015, together with a network of national fiscal councils, which, since the 'Two-pack', each member state has had to establish.

international interests match.[89] In addition, this thesis argues that continuous compliance with the SGP rules also depends on a specific perception of what constitutes 'good' economic policy. Given that ideational explanations have been observed all over the EMU integration process, it is plausible to assume that they also shape policy convictions that come into play with the SGP's day-to-day politics. Dyson and Featherstone (1999) and McNamara (1998) show such cleavages for the run-up of EMU (see also overview in Hix and Høyland 2011: 253ff.), Heipertz and Verdun (2010) for the creation of the SGP, in specific, and Schäfer (2016) for the creation of the Banking Union. Most recently, the debate during the financial, economic and sovereign debt crisis revealed fundamentally different perceptions of how to best address it (Brunnermeier et al. 2016).

In order to integrate ideational perspectives into the conventional concept of rationality, this thesis supports a perspective by which 'rationality' can be divided into two parts: a rational *process* of decision-making on the one hand, and a rational account of costs and benefits as the *content* of decision-making on the other. A decision-making process is rational as actors build preferences based on complete information, they rank their preferences in a transitive utility function,[90] assessing costs and benefits, and purposefully act to achieve the highest utility. Building on game theory accounts, this process also takes into account what the other actors might do and how this could affect their utility function and related strategy to maximise it. A rational account of costs and benefits includes measurable and comparable aspects, weighed against each other, and irrespective of other kinds of valuations emerging from personal tastes, emotional relationships, concepts of fairness, morals or beliefs.[91]

89 Putnam (1988) described governments as playing 'two-level games' by which "governments seek to maximize their own ability to satisfy domestic pressures, while minimizing the adverse consequences of foreign developments" (p. 434). The negotiation stance of a country is not just based on international issues, but also on domestic politics. Putnam considers an international agreement as passing two levels: the negotiation stage at the international arena, and the ratification stage at the national level. Both levels interact when an international agreement is to be decided. Every Council decision referring to the SGP can be considered such an agreement.

90 As a prerequisite for being able to meaningfully calculate utility functions and to define related strategies, preferences need to be complete and transitive. This means that all options need to be taken into account, that there are no outside options that might change the evaluation, and that between all pairs of options, the agent can define which one to choose or whether they are equal. Transitivity means that all options have to be ordered consistently, and that there are no circle preferences.

91 Gintis (2018) discusses examples where this type of rationality fails to explain behaviour such as in elections and for tax paying. Gintis argues that moral or social considerations, which he calls 'social rationality', also drive assessments of costs and benefits Additionally, the idea of somewhat 'bounded' rationality is also prominent in the literature, and assumes diversion from objective

Research has shown that such 'homo oeconomicus' is difficult to find in practice (see Selten and Ockenfels 1998, van Winden et al. 2008, Anderies et al. 2011, Coleman and Ostrom 2012). Co-operation and also sanctioning non-co-operative behaviour happens more often than expected, in particular based on considerations of fairness.[92] Based on the conventional view of rationality, compliance would depend on an evaluation of quantifiable costs and benefits, and considerations on strategic interactions only. This perspective misses non-quantifiable assessments of costs and benefits such as the salience of a problem to be addressed by politics, and expectations about consequences of policy actions.

The constructivist perspective on international relations offers another relevant aspect for collective action: 'shared beliefs' (McNamara 1998, Schmidt 2010, alongside economists' critiques of the role of rationality: Sen 1977, Hall 1989, Ostrom 1990, Hodgson 2006, Becker et al. 2015). 'Causal beliefs' are considered as describing a means-end relationship of aspects that underlie policy discussions. Beliefs help to "evaluate costs and benefits" (McNamara 1998: 4). This would explain diverging perspectives on costs and benefits, and sharing the same beliefs would bring about coalitions in the Council for or against compliance. The idea is that either members of the government (personal views), the whole government (party ideology), the political elite, experts or other politically-relevant interest groups (Heipertz and Verdun 2004: 771) or the general population (underlying values, culture) hold a set of beliefs that become relevant in the policy process. As described above, the EU law compliance literature also suggested taking 'beliefs' into account when describing a mismatch between domestic policies and EU law (Mastenbroek and Kaeding 2006). The assumption is that there are deeply rooted traditions, collective experiences, or else formed collective beliefs that trickle down to a collective understanding of economic and political phenomena and, accordingly transmit to policy choices to respond to these phenomena. Such an approach is also subsumed under the term 'economic culture' ('Wirtschaftskultur'), used among German scholars, in particular.

rationality due to information asymmetries and intellectual capacities to assess information and rank preferences (see Kahneman 2011). Some scholars acknowledge the existence of values or personal relationships, but try to assess them in light of rational choice (Becker 1996).

92 Moreover, Blyth (2006) cautions against applying micro-economic methods using rationality to any other research question, given the easily confoundable concepts of risk and uncertainty. While the 'homo oeconomicus' could be found in situations where risk is calculable, in situations with uncertainty, expectations would diverge.

Abelshauser (2010) notes: "The political economy of European integration cannot be considered in isolation from the economic-cultural preconditions"[93] (p. 23).

This thesis considers economic policy ideology as a set of beliefs that transmits to economic policy choices. In particular, such beliefs describe the perceived means-end relationship of economic policies and the respective relations between the state and the economy, and between economic, fiscal and monetary policy. The European integration literature suggests that it is plausible to assume that economic beliefs shape policy decisions at the national and international level (McNamara 1998, Dyson and Featherstone 1999, Hall 2012, Matthijs and McNamara 2015, Schäfer 2016). When describing the recent economic and sovereign debt crisis in the euro area, Brunnermeier et al. (2016) speak of a "long-simmering battle over the appropriate economic philosophy" (p. 2). 'Economic policy ideologies' are supposed to drive perceptions of costs and benefits, build preferences and, thereby, affect choices. In particular, ideology could determine whether or not the SGP rules are considered the 'appropriate' economic policy in a specific moment in time. Following the EMU integration literature, this thesis contrasts the two relevant economic policy ideologies: ordoliberalism and a state-interventionist perspective embracing Keynesian economic thought. The next section presents the SGP through the eyes of the ordoliberal spirit that underpins the original SGP. The contestation of the SGP relies on a different concept of the state, money and the economy.

In conclusion, this thesis suggests embracing a rational choice perspective by which the decision-making process can be considered as rational and following a logic of utility maximisation. It also suggests that the assessment of costs and benefits rely not only on rational considerations, but also on ideational perspectives on policy issues (cf. Ostrom 1990: 193).[94] This is distinct from the 'logic of appropriateness' (March and Olsen 1996) as this thesis does not argue that state behaviour follows considerations of what peers consider appropriate. Likewise, it is also not about

93 Own translation, original text: "Die politische Ökonomie der europäischen Integration lässt sich nicht von ihren wirtschaftskulturellen Voraussetzungen lösen." See Klump (1996) for the relevant methodology.

94 Ostrom (1990) presents a perspective on rational choice that entails norms as driving the evaluation of costs and benefits: "Individuals are perceived as weighing expected benefits and costs in making decisions as these are affected by internal norms and discount rates. Using this concept of rational action, one predicts that individuals will select strategies whose expected benefits will exceed expected costs" (p. 193).

convincing domestic actors, nor the other members in the Council, about the believed rightness of European rules (cf. Börzel and Risse 2003). Instead, this thesis argues that economic policy ideology shapes the weighing of costs and benefits according to respective expectations about the purpose and the means of specific policy choices.

3.2 The ordoliberal spirit of the SGP and contestation from the 'left'

The SGP's ordoliberal conception, as sketched above, is contested. The EMU integration literature contrasts the ordoliberal perspective to the state-interventionist perspective, embracing Keynesian economic policy ideas (Hix and Høyland 2011: 258ff., Dyson and Featherstone 1999, McNamara 1998, Brunnermeier et al. 2016). The perception of the purpose of economic policy goes against the SGP's assumptions, and unfolds in particular during a crisis: While ordoliberal policy seeks to prevent crises, adherents of Keynesian ideas seek to stabilise output through stimulating the economy during a crisis. This reflects the intention of policy action and, thereby, beliefs about the 'proper' economic policy.

This thesis uses the distinction between 'stimulus'- and 'prevention'-oriented policies as the essence of both thick economic policy ideologies. This distinction refers to fiscal policy, and is hypothesised to affect the intention to comply with the SGP rules that limit the fiscal scope of action. The sections below describe three aspects of this line of conflict that seem particularly relevant in the SGP context to shed light on the sources of misunderstanding each other when evaluating rule adherence: the understanding of crises, of fiscal policies as a response and of the state's needed scope of action (in addition to the above, for the ordoliberal perspective see Allen 1989, Hasse 2002, Hasse et al. 2002, for the Keynesian perspective see Hall 1989). The fourth section builds the economic policy ideology hypothesis, and the final section describes to what extent it is distinct from the literature's 'North-South' divide.

3.2.1 Understanding of crises and addressing them as end

The underlying dissensus refers to the conception of economic crises. While ordoliberals are of the conviction that crises are to some extent home-grown and, therefore, can be prevented, Keynesians consider crises as a natural phenomenon and therefore an inherent part of the business cycle that the state should target to address

when they occur. Policy discretion allows actors to respond to a crisis in the adequate way when all relevant information is available. They criticise rules for being a useless straight-jacket in such situations because it is difficult to predict what needs to be done in a specific crisis and accordingly design rules beforehand. Ordoliberals argue that they do not intend to predict crises, but to prevent the 'part' which is home-grown, i.e., caused by political decisions (or the absence thereof), notably by misled regulation, government mis-investments or a heated economy that has not been addressed by monetary policy. Accordingly, while ordoliberals may feel a sense of unease about high amplitudes of the business cycle, Keynesians might even see a problem in protracted 'slow' growth, and diagnose a need for government intervention.[95]

Accordingly, both perspectives have different understandings of the role of the state during a crisis. Following the Keynesian perspective, state intervention is necessary to solve the so-called 'savings paradox' as markets are inefficient and instable. When private actors save money in a crisis, the government is the only actor that can spend money to stimulate production and, therewith, private consumption. With this, the state can quickly put an end to economic bust and stabilise employment and wealth. In contrast to ordoliberals, Keynesians do not see a 'crowding-out effect', but a 'stimulation effect' stemming from state action. Therefore, putting limits on public borrowing would lead to pro-cyclical fiscal policies that deteriorate the crisis, slow down economic recovery, and might even risk causing a 'spiral' of deterioration, with low consumption, low prices, low wages, and deflation.

In contrast, ordoliberals see such state intervention as a risk to responsible behaviour of market actors in good times. Anticipating that the state would jump in and pay for the consequences of their decisions, market actors risk being less diligent when making decisions. This might also cause a crisis, and the state's role is to prevent this. Ordoliberals believe that in principle, economic actors know best about their business and investment decisions. Therefore, the state should not interfere with their decisions by 'artificially' stimulating demand. The state would necessarily fail to understand how economic actors would react and whether they would follow or

95 Similarly, in economically normal times, a strong state intervention might be needed to address competitiveness problems through investments. Closing investment gaps to build prosperity for future generations is considered more important than concerns over borrowing.

reverse the stimulation. Moreover, this also risks crowding-out of private demand.[96] Instead, societal goals should be reflected in a regulatory framework that constrains economic action only to the extent that it serves society's wellbeing.

3.2.2 Constraining fiscal policy as means

In general, the very idea of constraining fiscal policies goes against the Keynesian spirit because fiscal policies serve a macro-economic stabilisation function (Sims 2012). The Keynesian perspective objects to the ordoliberal understanding that there is a specific 'amount' of stable money, and that governments would not know how to invest meaningfully and successfully. Instead, they consider government spending to carry a multiplier effect larger than 1 by which 1 euro spent would trigger x euro more to be spent by the private sector. This would stimulate economic growth. Moreover, the government could always create fiat money, and thereby, increase the 'amount' of money. As a result, there would be no crowding-out effect for the private sector. Moreover, with economic policy targeting economic growth, for instance, through investments, there is no need to target a balanced budget in normal times to aim for sustainable public finances. Assuming economic growth of 5% of GDP, the fiscal deficit could be 3% of GDP and the debt level would remain sustainable (see de Grauwe 2007). There are also no concerns over fiscal sustainability as, in theory, sovereign default does not exist because the government could just create money. However, the EMU design misses such features and, therefore, the euro is described as a 'currency without a state' (see van Riet 2016). A lender of last resort would be needed to account for spreads due to liquidity shocks (which are not considered to be based on a country's conditions) and to allow monetary debt financing to account for the missing tool of currency depreciation.

Ordoliberals consider constraining fiscal policy as a means of ensuring sustainable budget positions, and to prepare sufficient scope of action for the duration of a crisis. They say that counter-cyclical policies also need a similar contraction of

96 The SGP shares the monetarist conception of money, whereby one euro can only be spent once. The 'amount' of money with a stable price is considered to increase only based on economic conditions. Public demand for borrowing is considered to bias the 'price' if it does not equivalently serve to increase economic fundamentals. Moreover, it would risk crowding-out private investment given the usually better credit rating of states, and their ability to always raise money through taxation.

government spending in boom times to avoid increased demand and pressure on prices. They doubt that politicians would understand when the time has come to cut back on state action because they are considered opportunistic, for instance, in the run-up to elections because policies such as increasing taxation or reducing government spending are unpopular. In contrast, they argue that in normal times, governments should target a balanced budget to be able to dispose of some leeway of borrowing in an economic downturn. In that respect, a deficit of 3% of GDP as the SGP prescribes should suffice for usual business cycle amplitudes. Moreover, the argument is that sustainable fiscal policies build credibility towards market actors, and a solid foundation for borrowing in exceptional crises to avoid liquidity problems, or even sovereign default.

3.2.3 Identifying the state's scope of action

The argumentation above is based on different concepts of the state. While ordoliberals seek to facilitate private action, a state-interventionist perspective considers the state as the crucial actor to advance the economy and society. While ordoliberals focus on prices as an efficient market mechanism, Keynesians see 'concerted action' as the most effective mechanism to stabilise demand and to advance economic development. The government as the democratic and sovereign leader is given a central role in this action. The state observes the overall development, has aggregate knowledge and the power to steer the economy at the aggregate level. Individual businesses and consumers could not assume this role as they do not have the respective tools. In order for the government to understand what the economy needs, and to collect necessary information about the pressing issues and relevant stakeholder interests and capacities, it needs large corporations and organisations to provide it with information. The government needs discretion over its competences, such as fiscal policies, to fulfil its role as stabiliser, and to respond in a targeted fashion to the peculiarities of the specific economic situation. Rules would artificially constrain the scope of action and bias output to an adverse extent for the economy, the labour force and welfare.

In contrast to the idea of an interventionist state stands the ordoliberal principle of freedom as being necessarily tied to responsibility: "The one who enjoys the

benefits also has to bear the costs of potential implications"[97] (Eucken 1990: 279). For economic actors, this principle of liability means that they enjoy the 'free' use of their property to the maximum extent possible, while they would also be held liable for any adverse effects or harm that arises for other members of the society, or the society as a whole (cf. Quaas 2002). While state interventionists also value responsibility as being tied to freedom, they acknowledge that there might be situations in which economic actors cannot foresee adverse consequences. Accordingly, ordoliberals promote building a policy framework that secures a level-playing field for economic actors. This puts special emphasis on the role of prices, ensuring that supply and demand can set them accurately. For this, it needs strong state agencies to enforce liability of actions. In the entire economic system, rules are considered as prerequisite for credibility and calculability, which are in turn considered critical for economic action. This resonates in the debate between 'rules' and 'discretion'. For the ordoliberals, rules provide stability, especially in times of recovery, to facilitate economic actors' investments. For the state-interventionists, discretion provides stability in times of crises, with the state providing bold action.[98]

3.2.4 Economic policy ideology hypothesis

In summary, the perspective on the common-pool resource, i.e., the Euro's monetary stability, differs along with the respective ideal economic policy, which unfolds in particular during a crisis. From the ordoliberal perspective, monetary stability is considered a major prerequisite for business activity, private investment and economic growth. Low variability in inflation, in addition to refraining from currency devaluation, is considered necessary to reduce transaction costs.[99] Also, anchoring expectations in the long run towards monetary stability provides economic actors with calculability (Abelshauser 2010: 17). In contrast, from a Keynesian perspective, the primacy of monetary policy should be economic growth and employment

97 Own translation, original text: "Wer den Nutzen hat, muss auch den Schaden tragen."

98 Interestingly, for both sides, the principled behaviour towards either rules or discretion anchors expectations. One could even go as far as arguing that their respective perspective is crucial for their respective economies to work. In contrast, promoting one perspective in the other's economy with its specific traditions is likely to fail in the short-term as they have such diverging concepts of the fundamental aspects of how the economy works and what role the state has.

99 Ordoliberals understand prices as the most rational market instrument. Therefore, prices should depend on supply and demand only, and not on biases due to money supply for which the inflation target of a central bank is crucial.

(Abelshauser 2010: 17). Instead of relying on prices as a market instrument, the focus is on bringing together consumption demands and wages, which Keynes considers the main driver of economic development and wealth. Monetary stability is not as important in the short- and medium-term, and the long-term does not help for the present – as Keynes commented "in the long run, we are all dead" (1923: 80). The focus instead is on the state's stabilisation capacity during the business cycle to support economic activity. In short, ordoliberalism believes that markets are self-regulating and stable, and governments are a possible source of instability. Therefore, governments should promote a policy framework that facilitates private action and avoids social harm, in particular to prevent economic shocks. Keynesianism believes that markets are not able to self-regulate to provide full employment and are unstable, which requires the state as a stabilising force. Therefore, in a shock, the government should be able to swiftly react with stimulating demand to quickly facilitate recovery.

Accordingly, this thesis formulates a hypothesis on the role of economic policy ideology for treating the common-pool resource, the Euro's monetary stability, as conceptualised above. Governments that prefer 'prevention'-oriented economic policies over 'stimulus'-oriented economic policies are considered more likely to endorse the idea of restricting budgets and following the SGP rules. They tend to perceive the rules as the 'appropriate' governance mechanism. As a result, compliance at both levels is expected.

> *Economic policy ideology hypothesis (EH)*: Governments that prefer economic policies targeted at stimulus as opposed to prevention are more likely to run higher deficits and are more reluctant to sanction breaching governments.

This thesis does not expect to see economic policy ideologies to cause constant breaches, or compliance, but instead that the effect might be mediated by the business cycle. It expects to see the difference between both orientations as being especially pronounced in economically difficult times when the political strategy for recovery is set out. In economic bust times, 'stimulus'-oriented governments are more likely to breach, while 'prevention'-oriented governments are more likely to comply. In boom times, one expects to see less of a difference between the two, even though 'stimulus'-oriented policies could also happen in boom times to encourage and 'steer' economic change.

This thesis considers two further aspects of personal traits to potentially intervene with measuring 'ideology'. Firstly, risk aversion could be an intervening personal trait that might bias measuring a 'prevention' orientation. Economic policies that reflect risk aversion would result in policies that are similar to prevention, even though the intention may be different. Therefore, it is also important to measure risk aversion in order to be able to disentangle it from a 'prevention' orientation. Moreover, risk aversion and considerations of patience might be confounded. Preferring short-term or long-term endeavours might not necessarily reflect a 'prevention' orientation. Therefore, this study builds a measure that combines risk aversion with impatience, and risk-taking orientation with patience. Both are usually perceived as opposites, allowing this thesis to study indeed risk aversion and not patience. One can assume that in economic policy orientation, a propensity of risk is more salient than a propensity of patience, as this is more relevant for economic policy content. Accordingly, the hypothesis states that risk-averse governments are more likely to endorse and follow the SGP rules:

> *Risk aversion hypothesis (RiH)*: Governments that prefer a risk-taking long-term perspective, as opposed to a risk-averse short-term one, are more likely to run higher deficits and are more reluctant to sanction breaching governments.

Secondly, a propensity for sustainability may intervene with measuring a 'prevention' orientation. A short-term orientation might coincide with preferences for having the capacity for a bold reaction, if necessary. Similarly, a preference for long-term orientation might coincide with preferences to prevent shocks and to provide for the future. Accordingly, the hypothesis states that sustainability-oriented governments are more likely to endorse and follow the SGP rules:

> *Sustainability hypothesis (SuH)*: Governments that prefer a short-term perspective, as opposed to a long-term one, are more likely to run higher deficits and are more reluctant to sanction breaching governments.

Therefore, this thesis assumes that the adherents of either side carry beliefs about a means-end relationship of economic policies. For adherents of ordoliberal policies, this study measures preferences for 'prevention' orientation. Related policies seek to prevent economic shocks, and are reflected in preferences for rules instead of ad hoc discretion. This coincides with preferences for risk aversion and long-term orientation. Policy measures are long-term oriented to build a framework that provides stability

and credibility. In contrast, for adherents of Keynesian economic thought, this study measures 'stimulation' orientation. Related policies aim to stabilise the economy in an economic downturn, stimulate economic recovery, and in normal times, to steer economic development if considered necessary to achieve certain goals. This coincides with preferences for risk-taking and short-term orientation. Policy measures reflect both so as to address a shock swiftly with a determined response and to achieve quick outcomes. Similarly, this is reflected in providing financing to ambitious projects and innovations that would require bold support.

3.2.5 Distinction towards the literature's 'North-South' classification

The economic policy ideology reflects political intentions that might in practice play out differently depending on circumstances and competing policy goals. As described in the literature review, the 'North-South' distinction refers to particular country characteristics stemming from policy traditions and historical economic development. One could argue that these are the result of economic policy ideology in the past and, therefore, the 'North-South' distinction also relates to economic policy ideology. However, this thesis argues that the 'North-South' distinction captures far more aspects that affect economic development, an economy's structure or the position of a currency.

Moreover, there might be other intervening factors. For instance, late developing economies might particularly rely on an interventionist state to quickly advance development. Similarly, Mediterranean market economies might not yet have departed from their "legacy of high levels of state intervention" (Hall 2018: 11), even though this does not necessarily reflect a respective 'economic policy ideology' as sketched above. In particular, countries in which the political system is linked to clientelistic structures might favour a discretionary approach without, though, sharing such economic policy ideology. This might be the case, for instance, for Greece (Katsimi and Moutos 2010, Featherstone 2015). In addition, some member states from the South struggle with comparably low-quality institutions in general, which might result in more discretion (Sotiropoulos 2004, see, for instance, the case of Greece, Featherstone 2015, or Italy, Quaglia and Maes 2004).

Likewise, the classic left-right political spectrum between socialist and conservative ideology appears too broad to fit the economic policy ideology distinction described above (see, for instance, de Simone et al. 2018). While there seems to be similarities with regard to the economic policy intention to focus on wages rather than on private sector returns, even this seems too simplistic as also ordoliberals emphasise the 'social' aspect of economic development. The political ideology plays out differently in member states depending on particular country characteristics, which, for instance, the 'North-South' distinction describes. The fundamental difference is the concept of the state. For instance, a French conservative would advocate for state intervention as much as a French socialist, while a German socialist might advocate for more state intervention than a German conservative, but refrain from too much state intervention and appear even conservative from a French perspective.

To exemplify, the case of France shows that the distinction between both the 'North-South' classification and the economic policy ideology seems necessary. The EMU integration literature depicts the ideational conflict between Germany and France, which however, is not necessarily reflected in the distinction of 'North' and 'South'. Instead, both are described as belonging to the 'core' of the EMU integration process (Heipertz and Verdun 2005: 987). The literature describes France as 'dirigiste', favouring state intervention and Keynesian economic thought (Dyson and Featherstone 1999, Brunnermeier et al. 2016). However, it is not a typical country of the 'South' (see Section 2.3.2.1.2). It does not have a tradition of high inflation, or a soft currency, and was a creditor country during the sovereign debt crisis. Instead, during the EMU integration, it belonged to the 'core' of countries with similar business cycles to Germany and the other countries. In contrast, Germany belongs to the 'North', and is thought to embrace an ordoliberal perspective. Dyson and Featherstone (1999) note that "[f]or German negotiators the postwar ordoliberal economic tradition was a key source of reference. It stressed price stability as the key economic public good that the state must guarantee" (p. 20). Accordingly, Hauptmeier et al. (2010) find that France ran expansionary and even pro-cyclical fiscal policies, whereas Germany ran restricted fiscal policies. This indicates that the ideational debate during the EMU integration and the sovereign debt crisis could also be explanatory for the regular SGP implementation. Therefore, this thesis argues that it is worth studying both hypotheses

separately. Through this, the thesis disentangles 'economic policy ideology' from other characteristics of a member state as depicted in the 'North-South' distinction.

3.3 Disentangling hypotheses on 'capacity'

In order to test the relevance of the economic policy ideology hypothesis, and to disentangle competing explanations, this thesis also studies other hypotheses that were put forward by the literature: economic necessity (the business cycle), economic and political relevance (size), country-specific characteristics ('North-South'), and strategic considerations (reciprocal voting). From the literature on common-pool resources, this thesis uses the standard hypothesis that contributing to the maintenance of the common-pool resource coincides with compliance with its governance rules (co-operation). Additionally, this thesis uses rule adherence at either level as explanatory factors for each other to reflect ownership (European level affects national level) and moral hazard (national level affects European level). This study does not include other factors mentioned in the literature as this would make it too complex. In the following, the first section discusses compliance as the dependent variable, the second section the null hypothesis, and the following sections present the other hypotheses.

3.3.1 Dependent variable: compliance

This thesis assumes that compliance with the voting rules, i.e., at the European level, signals agreement with the rules. Voting for sanctions means that the rules are endorsed and considered the 'appropriate' governance mechanism. In contrast, not voting for sanctions signals disagreement. As discussed above, compliance at the European level needs intrinsic motivation because the member state does not have any other advantage of doing so. There are no immediate financial returns, for instance, or gains other than those attributed to the maintenance of the common-pool resource. Even more, the motivation to vote would be higher than any potential gains associated with engaging in reciprocal voting to circumvent potential future sanctions.

The table below presents a respective typology when combining both level's rule adherence. Moreover, I include the case of abstention from voting, which would signal indifference towards the rules, or outright ignorance. This is important as this would show, for instance, that national-level compliance would be 'incidental' and not

affected by the rules. It just happens irrespective of whether the rules are in place or not. While there is no data to show a difference between abstaining from voting in the Council, or voting against a sanction, this distinction is conceptually relevant to discuss nuances of compliance.

Table 3.3-1: Rule adherence at both levels.

		Voting (European level)		
		Comply	**Breach**	**No voting**
Deficit (national level)	**Comply**	*'respecting'*	*'strategic breach'*	*'indifferent'*
	Breach	*'free-riding'*	*'active breach'*	*'ignorant'*

Source: own description.

Accordingly, this thesis assumes that the combinations of actions reflect the following intentions:

- 'Respecting': The combination comply-comply shows that the rules are endorsed and followed. There is intrinsic motivation to follow the rules.
- 'Free-riding': The combination breach-comply shows that the rules are endorsed, but followed only in specific cases. There is intrinsic motivation to follow the rules, but circumstances might create a need to diverge. It is important to understand whether breaches always happen, or only in specific cases. This would determine whether 'free-riding' on the others' compliance is systematic, or depends on specific circumstances.
- 'Strategic breach': The combination comply-breach shows that the rules are not endorsed, but followed by accident. Compliance with the deficit rule can be considered 'incidental' and there is motivation to go against the rules. Instead, not voting for sanctions may signal engagement in finding allies to circumvent sanctions in case breaches from the deficit rule occur.
- 'Active breach': The combination breach-breach shows that the rules are not endorsed and not followed. There is motivation to go against the rules.

- ‘Indifferent’: The combination comply-abstain shows that the rules are not endorsed and not followed. Compliance with the deficit rule can be considered ‘incidental’ and there is no motivation to follow the rules.

- ‘Ignorant’: The combination breach-abstain shows that the rules are not endorsed and not followed. Breaches from the deficit rule can be considered ‘incidental’ and there is no motivation to follow the rules.

In order to investigate why governments comply at either level, i.e., why they differ in their compliance track record, and why some breaches remain unsanctioned and others not, I need to be clear about what answers I can provide with this thesis. The above typology provides some clear cases. Firstly, as for ‘respecting’, the rules are followed by principle because compliance is built on intrinsic motivation. The opposite is ‘active breach’, where breaches happen because of outright rejection of the rules. Also, in case of abstention, the rules are not endorsed and no intrinsic motivation pushes compliance. However, in case of ‘free-riding’ and ‘strategic breach’, the motivation is not as clear. Moral hazard could be the source of breaches. This means that the preferences for the individual situation (for instance to breach) clash with what they want the others to do (to comply). One can assume that moral hazard is variable and depends on the circumstances. This means that the original strategy to comply could turn into a breach, or vice versa, given specific circumstances. The below hypotheses try to shed light on such circumstances.

3.3.2 Null hypothesis based on rational choice expectations

The null hypothesis states that governments would follow similar fiscal policies in similar economic situations, and variety would only be explained by the different conditions their economies are in at a specific point in time. From a rational choice perspective, one would expect governments to exploit the common-pool resource as well as others’ compliance as much as possible, and to free-ride on their provision of the good, i.e., to run higher deficits whenever they consider it useful – all else equal.

Regarding the voting rule, rational choice would predict that governments would not vote for sanctions as there are no immediate costs but potential gains from reciprocal voting. Compliance only becomes the preferred option in cases where enforcing compliance is considered critical for the provision of the common-pool

resource. This relates, for instance, to times of crises that affect large regions of the euro area. One could also say that this is expected to happen as soon as third actors, such as financial market participants, start to pressure rule enforcement as otherwise their trust would deteriorate. The below hypotheses seek to find explanations for the various ways in which governments respond to the rules.

3.3.3 Economic necessity and political capacity

The main hypothesis from the literature on capacity issues relates to the business cycle, and states that breaches from the deficit rule are more likely in times of economic bust. In bust times, resources become scarce, and a government might need to run higher deficits to satisfy their funding needs. Similarly, one can further assume that such cases of breach will likely remain unsanctioned in order to allow counter-cyclical economic policies and temporary deficit spending.[100] One could expect a government to grant the same flexibility in applying the rules in bust times as they would hope for themselves. Therefore, they would not vote for sanctioning. However, it could as well be the other way around, that even though a government runs higher deficits in bust times, they would not want others to do the same in order to ensure a high maintenance of the common-pool resource. This might only become relevant in times of crises, or symmetric shocks, which affect large parts of the euro area. Therefore, this is not considered for the general hypothesis. In recovery times, the picture might be mixed with some economic necessity remaining, while in boom times, there is no economic necessity left for deficit spending and, therefore, this thesis expects compliance – all else equal:

> *Business cycle hypothesis (BH)*: Governments run higher deficits in bust than in boom times. Governments are more reluctant to sanction breaching governments in bust than in boom times.

The second hypothesis from the literature combines several considerations on economic and political heterogeneity referred to as 'size' of a member state, and concludes that large member states are more prone to breaches at both levels than small or medium-sized ones given their relatively higher economic and political relevance. Such considerations are considered to be static and not easily changeable and,

100 There is a respective exception clause in the rules, which, however, might not suffice in some cases, necessitating a Council decision.

therefore, this thesis considers them as 'capacity' issues. The literature provides a significant variety of arguments emphasising the relevance of large states for the others, which reduces the threat of punishment and, accordingly, incentivises breaches. Given that this thesis can disentangle some of the aspects mentioned in the literature in other hypotheses, it considers 'size' to reflect only two aspects: the vote share in the Council and the economic importance for others. In the euro area, the distinction by vote shares coincides with economic power.[101] Thus, this thesis controls for general economic conditions of a member state, and only focuses on the aspects of size that relate to the opportunity of disposing of political and economic leeway. This conception of 'size' is not straightforwardly applicable to the data presented in Section 2.4, as these cases reflect the entirety of country characteristics that might intervene with the aspect of 'size'. This is the reason why this thesis seeks to disentangle competing explanations.

Accordingly, one can condense such a specific hypothesis from the literature. In case an economic shock also affects other member states and the large state serves as an 'economic anchor', the other states might be willing to refrain from sanctions and enjoy positive spill-over effects from economic stimulus. This expectation reduces the perceived threat of punishment, and incentivises breaching the deficit rule in the first place. Likewise, large states are more likely to breach the voting rule because they can engage in reciprocal voting more effectively, i.e., provide more votes to the counterpart. Even if other states would disapprove of a large state breaching the deficit

101 The literature defines 'size' in economic or in political economy terms. While Buti and Pench (2004) distinguish only between large (France, Germany, Italy, and Spain) and small (all others), there are more nuanced approaches to understanding 'size'. Eyraud et al. (2017) define economic size based on the member state's GDP as a share of the euro area's GDP. The literature studying the Council operation focuses on the number of weighted votes. The number of votes depend on the member state's population size. For instance, Panke (2010: 801) defines as small all member states that have less than the average number of votes (which is 12.78). Hagemann and de Clerck-Sachsse (2007) add a 'medium' category with the Netherlands, Greece, Belgium, Portugal and Sweden (similar: Baldwin and Widgren 2004). This thesis distinguishes 'size' according to the number of weighted votes and distinguishes three sizes. The number of votes reflects the power in the Council, which is the most tangible differentiation. Differentiation, according to economic factors, is more indirect and might also reflect other issues that might not, however, play a role or have different effects on the Council. Additionally, this thesis uses a third size 'medium', as it seems necessary to distinguish the small ones with negligible vote shares and economic impact on the common currency's stability from the medium-sized ones. This thesis keeps a similar distinction of large ones than the literature. It classifies Germany, France, Italy and Spain as 'large', the Netherlands, Greece, Belgium, Portugal, Austria and Finland as 'medium' sized, and Ireland, Lithuania, Latvia, Slovenia, Cyprus, Luxembourg and Malta as 'small'.

rule because of perceived adverse effects to the common-pool resource, they might find it more difficult to find a majority in the Council due to their lower vote shares.

Based on economic arguments, the literature argues that small member states are also more likely to breach the deficit rule because the threat of punishment is low. They are considered as being so economically small that the adverse impact of their deficits could be neglected and, therefore, the Council might not engage in sanctioning them. Behaviour towards the voting rule is difficult to predict as there are two opposing aspects at play. Firstly, they might breach the voting rule in favour of enjoying positive spill-overs from large states. Secondly, Schure and Verdun (2008) point out that small and medium-sized states would prefer compliance with the voting rules to constrain large states' power. Therefore, this thesis does not formulate an expectation towards voting rule compliance.

Medium-sized countries are seldomly explicitly considered in the literature. From an economics' perspective, they are considered 'small' in comparison to large countries. However, their economies cannot be considered negligible to the euro area aggregate. Therefore, this thesis does not formulate an expectation towards deficit rule compliance. Similar to small states, as noted above, one could expect compliance with the voting rules to constrain large states' power:

> *Size hypothesis (SH)*: Governments of small and large member states are more likely to run higher deficits than those of medium-sized ones. Governments of large member states are more reluctant to sanction breaching governments than those of medium-sized or small member states.

The literature's 'North-South' hypothesis builds on conventional wisdom that governments in countries located in the North of the euro area seem more fiscally prudent than their counterparts in the South (for a discussion, please see Sections 2.3.2.1.2 and 3.2.5). Accordingly, countries from the North are more often expected to endorse the SGP rules and follow them than those from the South. The idea is to define their propensity to accept and follow budget limits based on their capacity to implement them, be it economic or political capacity:

> *North-South hypothesis (NSH)*: Governments of member states that are geographically located in the 'South' are more likely to run higher deficits, and are more reluctant to sanction breaching governments than those from the 'North'.

The role of nationality might interact with the 'bust' and the 'size' hypotheses. From a simple look at the compliance track record in Section 2.2, the distinction between 'North' and 'South' seems to hold only for medium-sized countries where Portugal and Greece breach (South) and the others comply (North). The distinction does not seem to hold for small countries where Malta and Cyprus comply despite belonging to the 'South'. It seems that the role of 'size' prevails here. Similarly, the distinction does not seem to hold for large states because Germany and France (North) breach, and Spain (South) complies. However, here the role of the business cycle seems to intervene, as Germany and France breached the rule only in economically bad times and Spain had much higher growth rates than the other large countries, which facilitates compliance.

3.3.4 Strategic interactions

According to the literature on common-pool resources, the willingness to co-operate for the provision of a common-pool resource could have diverging effects on rule adherence. Intuitively, one would expect those who engage in the provision of the common-pool resource to endorse and follow the common rules. However, one could also argue that they might also consider it their 'right' to diverge from them. The hypothesis follows the first argument:

> *Co-operation Hypothesis (CH)*: Those who contribute to the maintenance and improvement of the common-pool resource are also less likely to breach the common governance rules.

As regards the real-world SGP, it is difficult to find specific measures equivalent to defining 'co-operation' in the sense of the common-pool resource literature other than compliance with the deficit rule. All endeavours to increase economic wealth in a member state can be considered as having a positive effect on the common-pool resource in the sense that fundamentals are built for a 'stable' currency.

The 'reciprocal voting' hypothesis builds on the literature's consideration that the SGP's institutional design incentivises reciprocal voting, which might impede rule compliance at both levels. Given that member states in the Council are asked to decide on sanctioning their peers gives rise to forming non-punishment alliances. As a result, (potential) 'sinners' will not sanction other 'sinners', but instead engage in reciprocal voting to avoid sanctions in the present or future. Engaging in such alliances shows

clear disregard for the rules. As explained above, there can be times when such opportunity can turn the original strategy to comply around:

> *Reciprocal voting Hypothesis (RvH)*: Governments that engage in alliances to avoid sanctions are more likely to run higher deficits and be more reluctant to sanctioning breaching governments.

3.3.5 Rule adherence

As a result of the discussion above, this thesis also formulates a hypothesis stating that breaches at one level are also expected to affect breaches at the other level. Governments who often breach the deficit rule either do not seem to care about the common-pool resource, or disapprove of the type of rules. Accordingly, one expects that breaching the deficit rule also affects breaching the voting rule. Likewise, one expects that breaching the voting rule also results in breaching the deficit rule. This is especially likely in cases where governments seek non-punishment allies:

> *Rule adherence hypothesis (RuH)*: Governments that run higher deficits are less likely to sanction breaching governments, and vice versa.

3.3.6 Hypotheses

The table below summarises the hypotheses. This thesis aims to study economic policy ideology as a source of intrinsic motivation for rule adherence. Additionally, the literature-based hypotheses consider the economic business cycle, economic and political size of a member state and related relevance for other member states, country characteristics in terms of a 'North-South' divide, and 'reciprocal voting' in terms of forming non-punishment alliances. The common-pool resource literature provides an explanation based on the provision of the common-pool resource. Finally, there is one hypothesis to reflect the role of compliance with either one rule for adherence with the other rule.

Table 3.3-2: Hypotheses.

Economic policy ideology and personal traits	
Economic policy ideology hypothesis (EH)	Governments that prefer economic policies targeted at stimulus, as opposed to prevention, are more likely to run higher deficits and are more reluctant to sanction breaching governments.
Risk aversion hypothesis (RiH)	Governments that prefer a risk-taking long-term perspective, as opposed to a risk-averse short-term one, are more likely to run higher deficits and are more reluctant to sanction breaching governments.
Sustainability hypothesis (SuH)	Governments that prefer a short-term perspective, as opposed to a long-term one, are more likely to run higher deficits and are more reluctant to sanction breaching governments.
Economic necessity and political importance	
Business cycle hypothesis (BH)	Governments rather run higher deficits and are more reluctant to sanction breaching governments in bust than in boom times.
Size hypothesis (SH)	Governments of small and large member states are more likely to run higher deficits than those of medium-sized ones. Governments of large member states are more reluctant to sanction breaching governments than those of medium-sized or small member states.
North-South hypothesis (NSH)	Governments of member states that are geographically located in the 'South' are more likely to run higher deficits and are more reluctant to sanction breaching governments than those from the 'North'.
Strategic interactions	
Co-operation hypothesis (CH)	Common-pool resource literature: Those who contribute to the maintenance and improvement of the common-pool resource are also less likely to breach the common governance rules.
Reciprocal voting hypothesis (RvH)	Governments that engage in alliances to avoid sanctions are more likely to run higher deficits and are more reluctant to sanction breaching governments.
Rule adherence	
Rule adherence hypothesis (RuH)	Governments that run higher deficits are less likely to sanction breaching governments, and vice versa.

Source: own description.

The purpose of this thesis is to try to disentangle the proposed explanations. This thesis does not expect to see that hypotheses play out purely and would result in constant breaches or compliance. Instead, they might mediate each other. For instance, the bust hypothesis might accentuate economic policy ideology. From a rational perspective, one would expect governments to exploit the common-pool resource (moral hazard), as well as others' compliance (free-riding), as much as possible. It is interesting to see under which conditions behaviour diverges from this rational account. Moreover, this thesis can investigate compliance at the national and at the European level separately.

This departs from the usual literature's account that rule behaviour at both levels would go hand in hand.

3.4 Summary

This thesis argues that the idea of 'appropriateness' in the governance of a common-pool resource reflects in economic policy ideology: The perception of whether the SGP rules are the appropriate governance mechanism to ensure the maintenance of the common-pool resource depends on economic policy ideology prevailing in a member state or specific government. The SGP reflects an ordoliberal spirit, which is easier to endorse by governments or member states that carry a similar economic policy conviction.[102] In contrast, a more state interventionist ideology with embracing Keynesian economic thought would find it difficult to understand the rules as the 'right' economic policy, arguing that the state needs scope of action in order fulfil its steering function and, in particular, its stabilisation function to address the savings paradox during a crisis. In particular, this thesis contrasts economic policies aimed at prevention with those targeting stimulation. This thesis focuses on the intended effect of policy actions, whereas, in practice, there is inherent uncertainty and policies might not yield the intended effect.

Accordingly, breaches from the SGP rules would result from the fact that for some member states, the SGP's ideology is not considered to be 'organically' grown, but 'imposed' on them. Put bluntly, the answer to the research question: 'Why comply?' would simply be, 'because they believe in it'. Compliance with both the deficit rule and the voting rule could, hence, be 'incidental'. Similarly, breaches could be 'systematic' depending on whether the purpose of the SGP is considered part of 'good' economic policy.

With this in mind, this thesis considers economic policy ideology as a part of 'political will' and as building 'ownership' of the SGP rules. Besides capacity issues reflecting economic developments, or opportunistic motivations reflecting strategic

102 We cannot assume that holding specific beliefs that match those of the SGP necessarily coincide with approving the rules as legitimate. These are two different concepts. Sharing the same understanding of economic policies does not necessarily mean that there is also the conviction that rules are needed to institutionalise these beliefs. This thesis just argues that the rules' legitimacy is easier to endorse if the beliefs match. Nonetheless, one can assume that there are few cases where rules would not be considered legitimate despite sharing beliefs.

interactions, ‘political will’ depends on the respective understanding of the real-world phenomena and what policy choices would help to address them. This would explain different reactions to similar economic situations.

It is noteworthy, though, that in contrast to the SGP, this thesis does not value one or other economic policy ideology as ‘better’. Instead, it is about whether fiscal rules need to be based on shared economic thought in order to be followed, or otherwise they would lack understanding and legitimacy.

4. Research design and methodology

4.1 Research question

Based on this theoretical framework, this thesis explores the motivations of a member state to breach or to comply with the SGP's rules at both the national and the European levels. It aims to contribute to the academic debate by 1) considering the SGP's institutional design, as well as the member states' economic characteristics, political power in the Council and prevailing economic policy ideology; 2) understanding compliance at both levels and analysing the resulting strategic interactions; 3) examining not only breaches but also compliance; and 4) providing empirical evidence, all in order to better understand why the SGP could not prevent breaches. The purpose is not to trace back the SGP's history of compliance and breaches, but instead to understand the complexity of adherence and to disentangle the reasons why. From this, one could draw insights into whether such rules could, in principle, establish the intended incentives for all members despite their diversity.

The research question is: Why do member states continue to comply with the rules despite the frequent breaches that the SGP has witnessed? What drives compliance and breaches at both the national and European levels? Or, put the other way around, why would members, which are different in many respects, comply with one and the same rule? Vice versa, why would member states differ in their compliance track record and votes for sanctions? As a result, what might impede the effectiveness of the SGP rules for incentivising rule adherence? Why are some breaches sanctioned, and others not?

With this, compliance at both the national and the European levels are the two dependent variables. Chapter 2 elaborates on the difficulties of measuring them, and of clearly defining breaches and compliance, in particular purposeful decisions for one or the other. Moreover, the unit of analysis is not clear-cut. One could abstract and say it could be country-year cases, as public budgets refer to one year. However, the SGP rules also allow the Council to follow up on continued breaches more quickly, and proceed with sanctioning, even after a few months. This would represent another case as it is a new voting situation in the Council based on a new assessment of the same

year's budget. Using an experiment helps to circumvent this problem and to clearly define a case.

The theoretical argument of this thesis is that economic policy ideology drives SGP compliance. In case it different than the SGP's spirit, it is likely to result in breaches – all else equal (see Section 3.2). As described in detail in Section 3.3, the other independent variables to be studied refer to the business cycle, economic and political heterogeneity, a 'North-South' divide, co-operative behaviour and strategic interactions. These are drawn from the literature (Section 2.3), and this thesis seeks to understand their relative role in a comparative perspective.

4.2 Method

4.2.1 The choice for an experiment and addressing limitations

Experiments are a common tool in political science, seeking to accurately replicate real-world cases, and thereby stand in contrast to simplified models usually used in economics (see Hermann and Ozkececi-Taner 2012, Iyengar 2012, also see Section 5.1 for a review). In particular, they have also been applied to studying state behaviour in international conflicts (Hermann and Ozkececi-Taner 2012). Iyengar (2012) considers them particularly useful to study the impact of nationality in a comparative way, thanks to online recruitment facilities. Ostrom (2006) explicitly recommends using laboratory experiments to study institutions in common-pool resource settings in order to disentangle the effects of competing explanations. If well designed with a view to internal and external validity, experiments allow the researcher to study causal inference.

For the purpose of comprehensively testing competing explanations, this thesis opts for an experiment. An experiment seems especially suited to testing economic policy preferences under laboratory conditions in order to disentangle the effects of other variables. The experiment offers a controlled setting of equal opportunities where participants can reveal preferences. I can manipulate relevant independent variables, such as economic and political heterogeneity and the business cycle, and I can compare different treatment groups using randomised sampling. The unit of analysis, and the definition of compliance and breaches, are clear-cut, allowing this

thesis to study the independent variables precisely. The independent variables are reflected in the treatments and in the choices made in the game. With this, the hypotheses can be studied using the choices and the comparison of treatment groups (see Diekmann 2005: 289ff.).

The experiment allows this thesis to study whether in the same situations and under the same conditions, participants would behave in the same way towards adherence with an SGP-like rule, which is the independent variable. If, for example, participants from different countries do not play in a similar way, their perception of costs and benefits might be different. Accordingly, this enables the research question to be studied from a very specific angle: Do participants make similar decisions under similar conditions (as the SGP expects)? In case there were differences, do they impede the SGP-like rule's smooth functioning? Could, in principle, different economic policy ideologies impede SGP compliance (all else equal)?

For the purpose of studying these questions, this thesis develops a new experimental design, which, by overlapping with private goods, also represents a new type of common-pool resource experiment, which I name the 'cultivation game'. The game has to be sufficiently abstract from the real world in order to avoid biases due to recent experiences, while preserving the SGP's structure and processes. The setting has to be an economic one with a simplified economy that has to be 'governed' by the participant and model the overlapping of a common-pool resource with private goods. Central to this study – and in contrast to previous studies (Irlenbusch et al. 2003, Irlenbusch and Sutter 2006) – is that this experiment puts the pure decisions about 'to comply' and 'to breach' into a proper context by giving them a meaning and basis for decisions. The experiment focuses on the participants' intuitive responses vis-à-vis rules in order to reveal their general tendency to take these rules into account when considering policy options. To the best of my knowledge, there is no similar experiment that I could have employed for this project. Therefore, drawing on experimental literature on common-pool resources, I will develop a novel 'SGP' experiment. It is about individual business behaviour complemented by an SGP-like sanctioning procedure for governing conflicts over the common-pool resource. It is about how to deal with scarce resources that are shared with others. The experimental design needs a great deal of description and justification and is, therefore, a genuine part of the thesis (Chapter 5).

In brief, the core allegory is an apple orchard that is cultivated by 19 farmers together over several rounds. The participant cultivates their own (private good) and shared trees (common-pool resource), which symbolise 'the economy'. The trees grow apples, some of which are needed for fertilising purposes. Weather shocks, that symbolise economic shocks ('business cycle' hypothesis), harm the trees. The participant can harvest apples ('tax the economy') that yield energy points, which can be used, in turn, to harvest more apples and to invest in cultivation items. The participant can choose between different items that either focus on increasing the growth of new apples, or protecting the tree against future shocks ('economic policy ideology' hypothesis). In this way, the participant 'governs' the trees and looks after their prosperity. Harvesting more than half of the apples on shared trees (simplified deficit criterion) corresponds to a breach, and participants can vote on sanctions. To avoid sanctions, participants can seal deals with other farmers ('reciprocal voting' hypothesis). They can also cultivate shared trees together ('co-operation' hypothesis). Participants are randomly assigned as a small, medium-sized or large farmer ('size' hypothesis), and are recruited from four different countries ('North-South' hypothesis). There is no goal of the game given so as not to bias the participants' choices. Participants are told to find their own strategy to cultivate the field.

While the experimental results cannot *explain* actual rule adherence with the SGP, they shed light on the role of hypotheses *in principle* for compliance with SGP-*like* rules. One cannot conclude that it was one hypothesis or another that explains a specific historic case of breaching the SGP. Instead, one could say that under a particular condition, it is more *likely* to observe a breach of the rules. With the experiment, this thesis studies the logic of an SGP-*like* setting, and how particular aspects interact. One can derive insight into the basis of the role of the variables within this comparable setting, which can inform the real-world SGP logic (see Section 6.4). To that end, the hypotheses will be translated into expectations towards the experimental results (Section 5.3.4). Applying an experimental design to a specific policy area and institution, such as the SGP, also makes it difficult to link experimental evidence as facts towards theories that have been developed to explain the real cases (see for a discussion Hancké 2009: 1ff.). This stands in contrast, for instance, to experimental studies that seek to understand human behaviour under particular conditions at an abstract level, such as under which conditions does communication

facilitate co-operation. This also stands in contrast to observational studies that examine real cases in detail. Therefore, one needs to be careful in drawing conclusions from empirical evidence towards the real-world case of the SGP, and also towards the broader academic debate about experiments. A nuanced analysis requires the experiment to fulfil a high level of internal and external validity, and to be precise about the limitations for the analysis. Based on a nuanced justification of the game design (Sections 5.2 and 5.3), Section 5.4 elaborates on internal and external validity in detail. This also includes addressing concerns on how seriously participants would respond in a fictitious context (Section 5.4.2 on social desirability).

In contrast to observational studies, experiments carry a singular advantage: In real-world politics, it is hardly the case that complex situations happen twice in a similar way. Therefore, one aspect might always become more prominent in one particular case than others. With the experiment, though, one can create comparable situations so as to quantitatively analyse them.

4.2.2 Limitations of other methods

In order to study the research question, one could have also pursued other methods, such as quantitative analyses or interviews, which, however face limitations. As regards the former, there is no official record of the member states' voting behaviour in the Council. There is not even any data on what preferences finance ministers expressed in the meetings. Therefore, two essential aspects of SGP compliance cannot be examined, namely strategic considerations for national- and European-level compliance. The main challenge of using interviews would be to abstract from the complex cases in order to find generalisable conclusions on what aspects drive rule adherence, in particular. Euro area member states are different in many respects, and it would be challenging to find cases to compare with others and to disentangle the competing explanatory variables that this thesis intends to study. Moreover, relying on public officials' or politicians' accounts might underscore the role of reciprocal voting or intentional breach of the rules, as they might refrain from openly admitting that other considerations were more salient than SGP rule compliance at the time.[103]

103 There are three more arguments against interviews with high-level politicians for the purpose of this thesis. Firstly, given that the euro is still rather young, there are not so many data points which are comparable as most part of the history is marked by two crises and several reforms of the rules.

4.3 Operationalisation

4.3.1 Case selection

The experiment is conducted in France, Germany, Greece and Portugal. All euro area member states would of course be of interest, however, due to feasibility, this thesis selects a sample. The choice is made with respect to the explanatory variable 'economic policy ideology'. The original debate about the design of the SGP was led by France and Germany expressing the most prominently opposing views on economic policy, embracing the ordoliberal perspective and a state-interventionist perspective with Keynesian economic thought, respectively (Dyson and Featherstone 1999, Brunnermeier et al. 2016, as well during the crisis, see Blesse et al. 2019).[104] During the economic crisis and the implementation of the financial assistance programmes, Portugal and Greece entered the debate on either side.[105] Portugal took over the German perspective by arguing for economic adjustment notably through structural reforms to facilitate business action and to make the economy more resilient (Lourtie 2012, Magone 2016). Greece supported France by voicing the necessity of stimulus economic policies to counter the crisis' effects and to stabilise economic output (Tsoukalis 2012, Featherstone 2016).

In addition to the 'ideology' hypothesis, the choice reflects two countries of the 'North' (supposed to comply) and two countries of the 'South' (supposed to breach). In terms of size, two of them are large (supposed to breach) and two of them

Secondly, these experiences might shape the interviewee's current perspective, which might even bias their original or general understanding of the rules and economic policy. An interview partner's memory of cases of breaches and compliance, which may date back a significant number of years, might be biased today, due to the development of those cases, how the case was settled, the ensuing events, or more recent discussions in light of the sovereign debt crisis. Moreover, some aspects might be forgotten by that time, or perspectives may have changed so that the case might be perceived differently. Thirdly, interview partners should ideally include relevant senior officials or politicians who were in office or in charge for periods dating back to 1999. It seems unlikely that a representative share of them would agree to take part in the interview.

104 For an analysis of French economic thought during European integration, see Johannsen (1999) and Rosanvallon (1989).

105 Ireland and Spain had also asked for financial support, however, their issues were less perceived as depending on a previously misled economic growth model, but on a bursting housing bubble that fuelled a banking crisis (Gros, 2010, Jones 2015). Cyprus came much later to ask for support when the new EU institutions were already built and the ideational debate had calmed down. The debate for Cyprus rather centred around avoiding a bail-in in the banking sector by the private sector, which later became necessary under the new Banking Resolution and Restructuring Directive (Kruse 2016).

are medium-sized (supposed to comply).[106] However, 'size' coincides with the 'North-South' classification. The experiment seeks to shed light on the different effects at play and to disentangle their role.

The table below summarises the criteria. Germany and France are most similar as regards the 'North-South' hypothesis, supposed to comply, but are most different in the ideology hypothesis, according to which Germany is supposed to comply and France is supposed to breach. There are similar expectations for Portugal and Greece.

Table 4.3-1: Selection on independent variables.

		North-South	
		North (C)	**South (B)**
Economic policy ideology	**Prevention oriented (C)**	Germany	Portugal
	Stimulus oriented (B)	France	Greece

Source: own description. Following the hypotheses drawn from the literature, (C) means 'supposed to comply' and (B) means 'supposed to breach'.

However, it is not clear to what extent these characteristics reflect in the compliance track record of the four countries, and what is due to the business cycle and the general economic conditions. Compliance refers to the deficit rule, which has been the main focus of the SGP, while the debt rule has not yet been used for sanctioning procedures. Germany and France stood together when abrogating the SGP in 2003. Portugal and Greece were the two countries that breached the SGP rules most often before the crisis. The table below presents the compliance track record. While Portugal and Greece were the only countries that (almost) never complied with the SGP before the crisis began in 2008, Portugal quickly implemented its programme requirements, and exited its programme ahead of time in 2014 (Lourtie 2012, Magone 2016). Instead, Greece had tremendous difficulties in implementing the programmes, and even came close to exiting the euro area (again) in 2015 (Featherstone 2016). Afterwards, Greece complied with the programme, but was still under enhanced surveillance until 2022 after having exited the programme in 2018. Germany and France breached in 2003,

106 To ensure representativeness and completeness, the study would have needed to include two small countries. However, this would have gone beyond the limitations of resources for a doctoral thesis.

which led to the reform in 2005, but complied afterwards up until the crisis. Though France never complied with the preventive arm, Germany complied during economically good times. While Germany was heavily hit by the crisis, but found its way out rather soon, France was still struggling with the deficit criterion in 2016. Hence, the question is whether bust times explain breaches and boom times compliance, or whether the reaction to both are (also) driven by the factors named above, namely 'economic policy ideology'.

Table 4.3-2: Reference to the dependent variable.

		Before the crisis	
		Compliance (in boom times)	**Breach**
During & after the crisis	**Compliance**	Germany	Portugal
	Breach	France	Greece

Source: own description.

With the experiment, this thesis can aim to disentangle these aspects. The participants would reveal their economic policy ideology with the choices to be made in the experiment, in particular, with respect to the business cycle developments during the course of the game. The aspect of 'size' is modelled directly in the game using random sampling in one of three groups. With this, the only relevant aspect from a participant's country of origin is their nationality, which is supposed to carry some cultural differences, collective knowledge, or historical imprints on the society.

4.3.2 Participants

Given that the experiment reflects individual choices in a setting that is abstracted from the specific political context, this thesis uses university students as the target group for this experiment. Ideally, the experiment's sample would reflect the general population, which is supposed to drive the government's re-election considerations and, hence, affect policy choices.[107] However, collecting such a sample would be out of the scope of this thesis. Students are a commonly-used target group for experimental studies

107 Literature also shows that at the European level, ministers in the Council are responsive to domestic public opinion (see Hagemann et al. 2017), and that governments based their decisions involved in creating EMU with a view to public support (Dyson and Featherstone, 1999).

(Druckman and Kam 2012), and can be considered as representatives of their cultural and political background (see Cason et al. 2002, Anderies et al. 2011). Moreover, students are a relatively homogenous group as regards their level of education, socio-economic background, income and age, and it is easier to distinguish control variables (such as field of study). For that reason, student samples have the advantage of "availability, replicability, and the limits to control" (Fréchette 2015a: 436). Fréchette (2015a: 435f.) also sees the advantage of having students as a target group instead of the general population or professionals because of their familiarity with participating in experiments. There would be less non-controllable noise stemming from confounding their real-world experiences with the setting and instructions presented in the experiment. In order to avoid misrepresentation and difficulties comparing students across countries, this thesis targets those who live in or around the capital[108] and whose field of study is economics or social sciences.

There are several caveats, though, when using students, namely their age, socio-economic background and field of study, for which one needs to collect data to be able to control for (see also Anderies et al. 2011). Research has found that undergraduate students seem to represent a rather unexperienced and privileged group of people (Ahn et al. 2011). Compared to the average in the general population, their young age makes them rather inexperienced. Moreover, university students tend to come from more privileged backgrounds than the average population. This might be especially the case for some more elite universities in the capitals. Their field of study might bias their perspectives on economic preferences and co-operation. One could also assume that people with specific perspectives rather select one field of study than another. Finally, it seems warranted to control for having lived or studied abroad, i.e., having been affected by other cultural contexts.

All in all, Druckman and Kam (2012) "argue that student subjects are not an inherent problem to experimental research" (p. 41) for external validity, i.e., generalisability. It is only necessary to collect relevant information to control for potentially biasing effects. Fréchette (2015a) presents a large review of studies using different subject pools, and concludes that students do not behave much differently

108 This is to ensure comparability of general living conditions across countries. Otherwise, it would be difficult to compare regions across countries. Moreover, this serves to reduce potential biases stemming from the particular characteristics of living in capital areas as compared to living in rural areas, for instance (see a discussion in Sinclair 2012).

than other subject pools.[109] The crucial aspects are that the game design allows students to meaningfully play the game, and that potential characteristics of students would not interfere with the game's analysis, or would be controlled for in the analysis. Druckman and Kam (2012) summarise it as follows: "any convenience sample poses a problem only when the size of an experimental treatment effect depends on a characteristic on which the convenience sample has virtually no variance" (p. 41). They go on to say: "Researchers then need to elaborate on the direction of the bias: the variation might facilitate the assessment of causation, and/or it might lead to either an overestimation or an underestimation of what would be found in the general population" (p. 53).

As regards replicability of the sample (see for a general discussion King 1994: 26f.), there does not seem to be a risk for the conclusions of the analysis. It is particularly important to collect relevant information on the participants' characteristics to be aware of potential aspects that could bias the analysis' results. Therefore, this thesis' experiment includes a survey (see Section 5.4.1.1). Fréchette (2015a) highlights the particular advantage of using students as target groups, as such target groups are easier to replicate given their rather homogenous characteristics. Replicating the experiment with another target group, for instance, experts or the general population, would be insightful for the validity of the analysis. Another interesting target group would have been finance ministry officials. Even if finance ministry officials do not decide on the final budget, they shape the ministry's economic policy perspective. This would have yielded insights on the perspectives at the policy-making level. However, I did not manage to attract a sufficient number of participants.[110]

109 Students have also been used to compare against experts for studying state behaviour in international conflicts (Hermann and Ozkececi-Taner 2012). Frechette (2015b) reviews several studies, and concludes that for most of them, there is not a substantial difference between students and professionals as subjects. However, it depends on whether professional expertise correlates with the particular game design. In that case, experts outperform students on average (DellaVigna Pope 2016). In contrast to the general population, Bortolotti et al. (2015) find that students seem more susceptible to rules and to the use of enforcement mechanisms for co-operation. This might relate to their rather young age and their particular mindset of learning, while subjects from the general population might rather try to find analogies to their real-world environment, which on average is different than that for students. Belot et al. (2015) compare students and non-students taking part in standard experiments, and find that "experiments using students are likely to overestimate the extent of selfish and rational behaviour in the general population" (p. 26).

110 From the original game, I constructed a simplified version to account for the time constraints of public officials. Between July 2017 and July 2018, I sent email invitations to the relevant State

4.3.3 Sample size

Defining the necessary size of the sample is an important aspect of experiments in order to be able to draw statistically sound conclusions from the results. The necessary sample size is usually defined based on the number of treatment variables and of groups within each treatment. The experiment has two treatments that affect the required size of the sample: Participants are recruited from four countries and they are randomly assigned to one of three different sizes. As both treatments do not have a 'control' group in the standard sense, one way of defining the necessary minimum number of participants is to look at each treatment separately. Building on the only other experimental study in the context of the SGP by Irlenbusch and Sutter (2006) that attracted 42 participants per treatment,[111] for the purpose of this experiment, I would need 40 participants per group for 'size'. Separately, I would need 40 participants per country. As a result, in total, I would need 160 participants.[112]

This coincides with the results of the power analysis to define the required size of the sample for studying the third treatment 'business cycle'. In contrast to the other two, this third treatment allows a standard comparison of treatment, i.e., 'bust' rounds, and control groups, i.e., 'recovery' and 'boom' rounds. To study this treatment, one can perform a power analysis (see Section 6.1.1) that concludes that the sample would need 155 participants to be able to conduct meaningful statistical analyses.

4.3.4 Experimental protocol

The experiment is conducted as a computer-based online game.[113] Participants receive a link and play individually and separately against computer-players. The experiment was designed, programmed[114] and conducted in order to ensure a high level of data

---

Secretaries in the Finance Ministries of the four countries. Three out of four forwarded it to their staff, and 11 officials participated, as well as 10 officials from the EU and national fiscal councils. These are too few cases for an analysis.

111 Irlenbusch and Sutter (2006) use 42 participants per treatment (i.e., voting procedure); in total, their study includes 168 participants. A group consists of 7 participants.

112 However, Irlenbusch and Sutter (2006) have a group and not a single-player set-up. Their observations are groups per treatment, of which they have six. This study's observations are participants per treatment. This means that at least for descriptive statistics, such as those Irlenbusch and Sutter (2006) use, six participants per treatment groups would suffice, i.e., 42 participants in total.

113 The game can be accessed here: http://personal.lse.ac.uk/ciaglia.

114 The programming was done by Martin Starman, a professional programmer. At the time, he was a Master's student of computer science at the University of Applied Sciences in Mannheim,

protection for the participants and of ethics standards, as required by the LSE Ethics Code.[115] Participants were asked to sign an informed consent form (see Appendix A.4.2) and could quit the game at any time without giving a reason.[116] Finally, the programming ensures proper randomisation of allocating people to the treatment 'size'. The study is conducted double-blinded, meaning that neither the participant nor the researcher know what treatment group the former is in (Diekmann 2005: 296f.). Both aspects are prerequisites for good experimental research (see Section 5.4.1.1).

4.3.4.1 Information for Participants

Overall, the formulations are supposed to be as clear and simple as possible in order to facilitate understanding, and to avoid attracting specific groups of participants over others. Participants receive the first information in the invitation via a leaflet, email, or Facebook post (see Appendix A.4.6, A.4.3, and A.4.5). The invitation shows the link and context of the experiment, i.e., being part of a PhD project. It briefly states what the experiment is about, and explains why participation could also be beneficial for the participant. On the game's website, the participant first sees the informed consent form (see Appendix A.4.2). It follows the basic game instructions (see Appendix A.4.4). [117] Further instructions are provided during the game as needed (see Appendix A.5-2, Section 2).[118]

Germany. Martin Starman and I came up with the graphic design concept, while Niklas Zorell, a professional graphic designer, was in charge of the graphic elements. At the time, Zorell was a Master's student of Interactive Mediasystems at the University of Applied Sciences Augsburg, Germany.

115 The programming and conduct of this experiment were approved beforehand by the LSE Ethics Review. The preparation was advised by the LSE Behavioural Research Lab staff, the LSE Library staff and the LSE Data Protection Officer. Participants are never asked to give their name or email address, postal address, or any other information that would allow the experimenter to directly identify them. Instead, participants are provided with unique ID numbers in case they want to delete their participation. See further details in Appendix A.3.1.

116 By not storing the IP address, using cookies, or other similar means of identifying participants, I could not verify whether a person played the game twice. The importance of anonymity, however, predicates the assumption of trust in participants. See details in Appendix A.3.1.

117 For the Finance Ministry version, a link to the basic instructions is included in the footnote of each page and, therewith, also accessible during the game. For the student version, basic instructions were only displayed at the beginning of the game. However, students could open a second page in their browser and revisit the basic game instructions page.

118 The game utilises simple English language and additionally icons to visualise text. The game assumes basic levels of English proficiency and, hence, translating the game into French, Portuguese, Greek or German does not seem necessary.

4.3.4.2 Development and Pre-Test

During the game's development, eleven persons[119] played the game with me beside them to discuss it, and to ensure internal plausibility of the game's design, clarity of the instructions, and the accuracy of interpretations from the game. For the same purpose, I conducted a pre-test with 21 university students in Lisbon (for a detailed list, see Appendix A.4.7.1.1). Moreover, participants did not raise any content-related questions about the game. They only had organisational questions, which I addressed by making the instructions clearer.

4.3.4.3 Recruiting participants

I chose the largest and most well-known universities for economics and social sciences in each capital[120] according to recommendations by my supervisors and other professors in my department.[121] It was not possible to conduct the experiment in collaboration with experimental labs, as not all universities have experimental labs,[122] some are not open to external studies,[123] and several labs required remuneration of participants. However, this experiment was not designed for (success-based) remuneration for reasons of potential bias (see Section 5.4.3.3).[124] Therefore, I opted for a two-step approach: first, I went in person to each university in Lisbon to 'simulate' a laboratory situation.[125] As it proved unnecessary to being around as experimenter, I started to disseminate the link electronically. I chose broad means of dissemination to approximate the usual sample pool of laboratories. In Lisbon, I promoted the experiment through posters, announcements in class, direct contact and

119 Including my supervisors, several PhD colleagues, and friends whose work is not related to research. With Martin, I continuously checked the proper implementation of the design into the programming.

120 See Appendix A.3.7 for a detailed list of contacts. The only university from which I did not get any response was the Technical University in Berlin.

121 I am exceptionally grateful for recommendations by Kevin Featherstone, Paul de Grauwe, Francisco Torres, Waltraud Schelkle, Mareike Kleine, and Vassilis Monastiriotis.

122 For instance, the 'Apple Tree Game' was the first experiment to be conducted on the premises of Universidade Nova de Lisboa, School of Business and Economics (SBE).

123 For instance, at LERNE, the experimental laboratory at Universidade Católica Portuguesa, Economics Department, I could have used their facilities, but would have needed to invite participants myself. The study could not be conducted at the LSE Behavioural Research Lab because they did not run online studies at the time.

124 Even if, paying a lump sum would have been too costly for this thesis. Considering a minimum of a total of 160 participants, of which each receives a minimum lump sum of 10 euro for a study that takes 30 minutes, this would have costed 1,600 euro.

125 See appendix A.3.7.1 for further details on why I started with Lisbon. In total, 18 people played the game under my supervision.

distributing leaflets.[126] From autumn term 2016 until spring,[127] the link was distributed to undergraduate students at all of the named universities in a number of ways: by email, through a learning platform, or a newsletter from their department's administration, professor or teacher[128], and via Facebook posts from student associations and student groups.[129]

Taking all these considerations together, it seems reasonable to conduct the experiment via email, and not in person. The main challenge for online experiments is that the researcher cannot verify whether people understand the game and whether they play it by themselves and alone. The first challenge is a prerequisite of the experiment's success. With the experiences from the pre-test and in Lisbon , however, the risk seems rather low. Moreover, there might even be an advantage to allowing students to play the game in private and not under supervision, as they could then just quit the game without finishing if they do not understand it. This avoids having potentially biased observations in the sample. Secondly, without collecting personal data, I cannot verify whether people played by themselves and alone. However, the risk here is considered negligible because the main focus is on having a coherent strategy, which is the essential requirement for conducting the analyses. If several people played and completed the game, one can assume that they enjoyed it as otherwise they would have quit the game. If they enjoyed it, one can assume that they

126 Firstly, before exams, I announced that the experiment would take place afterwards, and put posters in front of the exam rooms indicating time and place (see appendix A.3.6). Secondly, I distributed leaflets with a short description and the link to the game on campus for people to play it at home. This helped to recruit more students than those who had time to participate at that moment. Moreover, I went up and down the hall in front of the computer room and approached students directly. Finally, as a small remuneration I had candy for the participants. At Universidade Nova de Lisboa, School of Business and Economics (SBE), the announcement was put on a student group Facebook page and on the learning platform for some courses.

127 The link has been circulated electronically between 25 August 2016 and 06 June 2017. With several reminders having been sent out, the conduct spread over the whole academic year. As the experiment is not related to any real-world developments, this long period of time should not be a source of bias.

128 Professors also sent it to their Master and PhD students. Moreover, the invitation has been circulated to one university's experimental laboratory mailing list. A poster has been put up at two campuses since professors were not allowed to circulate external invitations to their students.

129 At the end of Autumn, I realised that participation was not yet sufficient. I sent reminders and I started to advertise the experiment more broadly on Facebook and sent the link to the four national student societies at the LSE. I approached country-wide student organisations such as Young European Federalists via Facebook. I also approached professors at the Hertie School of Governance in Berlin. Professors started to recommend the game to colleagues in other departments, for instance to a geography department or a medical school.

tried to follow a coherent strategy (also see discussion on social desirability in Section 5.4.2).

These multiple ways of distribution have certainly intersected each other, and I could not verify whether one student might have received the invitation several times from different sources. Therefore, the numbers that I report indicate the number of students for each time that the invitation was circulated. I call this measure 'student/times' and I simply add up all these student/times.[130] In total, the link was sent out about 21,881 student/times. Furthermore, the link was posted on Facebook in student organisation groups, and promoted by individual students about 111,942 members/times. The game was started 2,512 times,[131] with 509 participants finishing the game. Participants from countries other than the four target ones and a few other participants are not considered for the analyses.[132] Therefore, the final sample size is 327 participants. With this, the sample is more than twice as large as that considered necessary (Section 4.3.3), allowing this thesis to conduct meaningful statistical analyses.

4.4 Statistical analysis

While the unit of analysis in the real world are country-year cases, the experiment approximates this with participant-rounds. Each participant plays over several rounds. I define the dependent variable for each round as 0 or 1 for compliance or a breach, respectively. In the experimental data, this allocation is straightforward, while in the real world, it is not that clear-cut, and depends on the point in time when it is measured (see Section 2.2). König (2005) highlights the importance of precisely defining the unit of analysis, both in the real world and in the operationalisation of the study.

The final dataset is not biased with regard to the main variables of the analysis, which means that the distribution of cases across treatments is homogenous, and that

130 Because of the broad means of dissemination, not only did students play the game, but also higher academic staff and other people. This is accounted for in the analyses.

131 The total count on the webpage is in fact 2,812 started games. About 300 of them account for the original development of the game and the pre-test stage.

132 There are 97 participants from other countries. Sixty-five participants indicated that they had played the game before, with most of this being attributable to myself and Martin Starman who played during the development and pre-test stage. Nine participants indicated that their English language level was not sufficient enough to fully understand the game. I exclude all these participants from the final dataset.

the distributions are not tilted (see Section 6.1). It includes variables to control for individual characteristics that might affect general behaviour, such as age, gender and field of study. The analyses use mixed models to display the panel structure of the data, and logistic regressions for compliance with the deficit rule (Section 6.2) and with the voting rule (Section 6.3), with the same set of independent variables.

4.5 Summary

Building on existing experimental literature in the field of political science and economics (also see Section 5.1), this thesis develops a novel experimental design to study compliance with the SGP rules. An experiment is especially suited to providing a controlled setting to test the relative role of competing hypotheses. However, experimental evidence cannot *explain* real-world cases. It is simply not observational evidence, and also requires a level of abstraction that necessarily reduces accuracy. Instead, it provides insights into the logic of an SGP-*like* setting, and how particular aspects interact (see Section 6.4). To that end, the hypotheses are translated into expectations towards the experimental results (Section 5.3.4) and external validity is discussed in detail (Section 5.4.1.2). Essential for making meaningful statements from experimental results is internal validity (Sections 5.2, 5.3, and 5.4.1.1). Chapter 6 provides the analyses and discusses the results in light of the real-world case SGP.

5. The ‘cultivation game’

This chapter develops the experimental design, and is organised as follows: First, Section 5.1 reviews experimental literature in the field of the SGP and of common-pool resources and conclude that there is no suitable experimental design that could be adapted for the purpose of this study. Section 5.2 discusses the logic and allegory of the novel ‘cultivation game’ and the specific features that it has to plausibly reflect the real world. It follows a detailed justification of how to draw interpretations from the moves in the game (Section 5.3). As the experiment follows a new structure (common-pool resource that overlaps with private goods) and a new context-rich background for choices, the section also devotes detailed explanations to evaluation using typologies. Section 3.3’s hypotheses translate into specific expectations about the participants’ choices in the game (Section 5.3.4). The chapter closes with a discussion of how this novel design satisfies principles of good experimental research, including concerns regarding validity and social desirability (Section 5.4).

5.1 Literature review on experiments

The following section reviews the literature on experiments, firstly in the field of the SGP and EMU, and secondly, on common-pool resources. There is only one study which investigates the SGP voting rules. While this paper does not study reasons for rule adherence, i.e., other costs and benefits than pecuniary ones, it provides valuable insights into the logic of the experimental design. Secondly, while the standard common-pool resource games provide an excellent abstraction of the underlying logic that is common to all common-pool resource settings, there is no existing experimental design that meaningfully reflects the particular setting of the SGP where the common-pool resource overlaps with private goods. I will draw on these insights when developing the novel experimental design.

5.1.1 Experiments in the field of the SGP

While experimental research has been on the rise for a diverse range of topics in recent years (see Diekmann 2009 and Druckman et al. 2012a for political science, Roth 2015

for economics,[133] Duffy 2015 for macroeconomic research questions, and Palfrey 2015 for research questions in political economy), to the best of my knowledge, I could only find one study in the context of the SGP and EMU (Irlenbusch and Sutter 2006), which will be presented below in more detail. Moreover, there has been one research project, led by Prof. Selten and Prof. von Hagen at the University of Bonn, that seeks to experimentally study the economic impact of a monetary union, but they do not study the SGP (Pope et al. 2006).[134] Moreover, Weisman and Merler (2012) conducted a large-scale simulation of the sovereign debt crisis events following the first bail-out requests. While there is extensive literature that uses game theoretic approaches to study specific aspects of EMU, the models used appear ill-suited to be applied for the purpose of this thesis because they do not account for different valuations of costs and benefits, different characteristics of member states, or they do not model the interplay between national- and European-level compliance explicitly.[135] Likewise, standard games in game theory (see Osborne 2009:11ff.) are not suited as they represent symmetric games.[136]

---

133 The article is part of the Handbook by Kagel and Roth (2015) that focuses on experimental economics in general. The Handbook of Experimental Economics Results by Plott and Smith (2008) covers a range of topics. Specifically, experiments are also used in the field of central banking (see a review by Cornand and Heinemann 2014).

134 Their set-up mirrors a two-country situation which includes numerous actors such as governments, central banks, employers' organisations, trade unions, and firms that also trade. They run two versions, one of which simulates a monetary union, while the other does not. They expect positive effects on the countries' economic development in a monetary union due to reduced complexity and uncertainty for economic decisions.

135 The following studies employ game theory to study the topic: Vöpel (2013) investigates in what elements the institutional design would need to reduce moral hazard effects, and concludes that it needs both a stabilisation mechanism and an exit option. Governatori and Eijffinger (2004) find that member states seek to maximise their deficit as much as possible, and arrive at a certain Nash equilibrium in the co-operation with others. Fourçans and Warin (2007) use a centipede game to show conflicts between member states. Schure and Verdun (2008) study legislative bargaining among member states when implementing the SGP and Hosli (2000) describes bargaining about the creation of EMU. König and Mäder (2014) study member states' implementation with EU law, and respective sanctioning by the European Commission. Game theory has been used to describe the necessary size of sanctions to induce compliance (Ohr and Schmidt 2003, de Haan et al. 2004), the need for fiscal rules when considering savings in the euro area as a common-pool resource (Detken et al. 2004), coordination problems between the European Commission and member states (Barbone and Poniatowski 2013, Wijsman and Crombez 2017), the sovereign debt crisis (Vamvakidis and Woo 2012, Mance et al. 2013, Canofari et al. 2014, Hennessy 2017, Karagiannis and Konstantinidis 2020, Fahrholz and Wójcik 2012, Tsebelis 2016, Gourinchas et al. 2020), the interplay between monetary and fiscal policy authorities (Artis and Winkler 1997, Howarth and Loedel 2004, Dixit and Jensen 2003, Levine and Brociner 1994, Herzog 2005, Beetsma and Jensen 2003), and macroeconomic stabilisation under fiscal policy constraints (Neck et al. 1999, van Aarle et al. 2001, Buti et al. 2001, Cooper and Kempf 2000, Engwerda et al. 2002, Neck and Blueschke 2014, Heinemann 1998).

136 Although each of the games (prisoner's dilemma, the game of chicken, battle of the sexes and the stag hunt game) considers one of the relevant aspects for the SGP (free-riding, riskiness, terms of

Irlenbusch and Sutter (2006)[137] study different voting rules under the SGP. They use a public bad game[138] and test three different voting rules. They find that, compared to equal vote shares, weighted vote shares aggravate breaching. In contrast, withdrawing voting rights from participants who are in an EDP for decisions on other participants' breaches significantly increases overall compliance. Breaching participants cannot 'solidarise' with each other but instead face sanctions by those who comply with the rules. Focusing only on different voting procedures, the experiment is designed parsimoniously and without an illustrative context for the decision-making. In each round, participants receive an endowment, and decide between choice Y (co-operation) and X (free-riding). The latter induces costs on the others depending on the size of the participant (small, medium, large). In case a player chooses X, an SGP like three-stage voting procedure follows. Considering only the payoffs of one period, free-riding is the dominant strategy. Taking the other periods into account as well, the payoff would be maximal if all players co-operated (see further details in Appendix A.5-3).

The strength of this experiment is its simplicity, and this is well-suited to studying different voting procedures. This experiment tells us, under strategic considerations, what kind of voting rule would be most effective at disciplining. However, in their experiment, a breach only depends on financial incentives, and this does not make 'sense' without a context. In the real world, decisions on compliance and breaches are not only determined by these two factors, but mainly by the purpose for which one chooses to breach. It is crucial to understand compliance as based on economic considerations and political strategies. Hence, a context would be needed to understand the meaning and intention of breaches.

In order to develop such an experimental design, one can draw valuable insights from Irlenbusch and Sutter's (2006) experiment. Firstly, in the real world, the

agreement, trustworthiness), none of them can account for the whole picture and hence, would ignore relevant situations and play-based strategies.

137 There is another version of this experiment that also allows communication (Irlenbusch et al. 2003). Participants could freely talk to each other about anything before the voting stage, but were forbidden to reveal their player IDs. They found that communication significantly decreases breaching. This might relate to the successful peer pressure that communication created. However, this might only work when participants are granted anonymity as this prevents participants from using reciprocal voting.

138 In contrast to providing a public good, this type of experiment focuses on avoiding a public bad. In this context, this means avoiding high public deficits.

intentional choice between X and Y may not be as discrete, but might rather happen in a 'grey area' where it might not be clear from the start that a government will end up with X although Y was chosen, or vice-versa. For instance, the experiment does not account for difficult economic situations, and the resulting narrowing down of a government's options to actually stick to Y.

Secondly, designing the SGP as a public bad game is a very elegant way to escape the need to define what the SGP's 'public good'[139] actually is. However, in the real world, member states may not only consider avoiding a public bad, according to the SGP, when deciding about their peers' deficits. Rather, they take their national economy into account. Therefore, it is more realistic to model the SGP and its context as a common-pool resource (monetary stability) that overlaps with private goods (national economies). Then, the contribution to the common-pool resource actually depends on a member state's economic situation and convictions with regard to a potential trade-off between contributing to the common-pool resource, or to their own economy.

Thirdly, the costs and benefits resulting from a player's choices are not necessarily uniform, and do take into account differently perceived consequences from other participant's actions. Although such abstraction seems perfectly fine for the purpose of their experiment, it does not allow the player to form different perceptions of the SGP's consequences for the economy. For instance, for some governments, a low deficit might yield higher 'gains' than a high deficit, thereby coinciding with what the SGP prescribes. Then, the government would 'choose' compliance incidentally, without actually choosing it purposefully. Moreover, when deciding about sanctioning another player, the uniform payments of tokens might not reflect the nuanced considerations at play in the SGP context. Instead, it might be more realistic to have varying numbers of initial endowment according to a player's size, so as to better reflect what is 'at stake' for this player.

139 Moreover, the SGP's good is not a public good (non-exclusive and non-rival) as the use is rival. Nonetheless, common-pool resource games and public good games have long been considered as the same type of game with just a framing difference: while public good games were about contributing, common-pool resource games were about extracting (for a discussion see Apesteguia and Maier-Rigaud 2006). The distinction, however, is a little tricky. For public good games, even though exploitation shares are equal, i.e., non-rival, provision is not equal. Voluntary provision might trigger free-riding, which could also be understood as rivalry.

In conclusion, one could draw from their experiment the idea to include different sizes, a voting procedure that follows several steps and reflects the 'flexible' sanctioning procedure prescribed by the SGP (first opening, several warnings and only finally a fine), to have several rounds and to reveal information after each round. Moreover, there are additional aspects to consider for this thesis' experiment: provide a meaningful context to choices, allow more nuanced choices than just X and Y, not prescribe a specific goal, have the common-pool resource overlap with private goods, and design options so that choices reveal different valuations of costs and benefits.

5.1.2 Standard common-pool resource experiments

This section discusses standard common-pool resource experiments that might help to design the experiment. Common-pool resource experiments seek to find out under which conditions sustainable co-operation emerges for maintaining a common-pool resource. The basic assumption, drawn from a rational choice perspective, is that participants seek to maximise their own benefit and, thereby, would over-exploit a common-pool resource due to the *unrestricted use* available to them. However, due to the *restricted availability* of the good, and the resulting rivalry among its users, this behaviour would soon lead to a 'tragedy of the commons', i.e., over-exploitation and subsequent vanishing of the good. In effect, the originally rational behaviour leads to an irrational outcome, i.e., the vanishing of the good. Therefore, co-operation is necessary for managing provisioning (providing the good) and appropriation (extraction per user).[140]

The literature provides two standard games: the investment game (see Ostrom et al. 1994, Deadman et al. 2000), and the request game (Suleiman and Rapoport 1988, Budescu et al. 1995). In the investment game, participants get an endowment of tokens, which they can either keep or invest in the common-pool. Up to a specific threshold of total invested tokens in the common-pool resource, individual investments in the

140 A large body of experimental literature emerged to study the role of communication, punishment, rewards, uncertainty about the size of the pool, a growing stock of the pool at a natural rate, different sizes of actors, participants playing as a group or individually to show group or individual behaviour, one-shot games or several rounds with or without varying conditions between the rounds or between groups, and a finite or uncertain number of rounds. The section refers to the studies as deemed insightful for this thesis' experiment.

common-pool resource yield higher returns than kept tokens. If that threshold is surpassed, returns on kept tokens are higher than on invested ones.

In the request game, participants request any number of tokens they want from the common-pool resource. However, they do not know the exact size of the pool. They only know that the pool will have a number of tokens between *a* (for instance 10) and *b* (for instance 100). All requests are made at the same time. If all requests are less or equal to the size of the pool, each participant gets their requested number of tokens. However, if the total requests exceed the size of the pool, all participants receive zero tokens.

5.1.2.1 Meaningful zest for tokens

Both standard games have limited applicability to the real-world SGP for two reasons. Firstly, standard common-pool resource settings refer to the exploitation of natural resources for food production, such as fish stocks (see Ostrom et al. 1994, Ahn et al. 2011). There are two important differences to the SGP, which the games cannot capture. On the one hand, a fish stock is a common-pool resource from which fishermen have to extract at least a share, as their existence heavily depends on this job. In experiments, this existential need to extract tokens is reflected in pecuniary rewards for participants that are tied to success in the game. However, there is no such existential need in the real-world SGP context. At least, this need is not uniform and does not equally apply to all governments. Governments do not have to run deficits and 'exploit' an available amount of stable money. On the other hand, governments could 'add' to the common-pool resource by running surpluses (more resources available for other users), and improving the economic conditions (increasing the overall 'amount' of the resource). This is not the case for natural fish stocks.

Secondly, the standard games do not allow participants to develop different valuations of the goods, or of their use for their own purposes. Due to the pecuniary incentives in the standard games, participants strive to gain as much as possible from the common-pool. For instance, studies exist that simulate common-pool resource behaviour using computer algorithms that optimise the highest 'return' (cf. Deadman et al. 2000). However, this is not necessarily the case for the SGP. In a real-world setting, where such decisions are put in a context, participants might attach different

values to the common-pool resource in general, or each additional unit that they might consider extracting given all the other choices they could make. As described in Section 3.1.5, behaviour might differ due to different preferences towards the potential 'trade-off' between monetary stability and economic growth. In the real-world SGP setting, the common-pool resource overlaps with private goods to provide a meaningful context for decisions.

5.1.2.2 Meaningful context

There is one experimental study that uses a context to options, creating a virtual-world setting (Twieg and McCabe 2014).[141] However, the context only provides for studying degrees of exploitation of the common-pool resource without considering the purpose how this good will be used. Similar to standard common-pool resource games, after the game, tokens were converted into money, and this is supposed to tie the participant's interest in governing the common-pool resource back to their real life. Put in more abstract terms, it ties it back to the context where participants would actually make 'use' of the tokens that they extracted from the common-pool resource. With a view to the SGP setting, however, it does not need the participants' real-world to make use of the tokens. Instead, the tokens would be directly used in the game for the private good. It is the very purpose of borrowing to fulfil domestic policy goals.

The literature offers two more ways to reflect different valuations of the common-pool resource even though they are not so nuanced as to also account for non-pecuniary motivations for sanctioning, such as principled sanctioning or reciprocal voting. Firstly, Cason and Gangadharan (2015) use a sanctioning mechanism that accounts for costs for a participant that votes on sanctions. The participant has to pay 1 token to incur a sanction worth 3 tokens. However, this approach is also inadequate for the SGP, as it does not account for incentives not to sanction misbehaviour (for instance because of positive externalities or reciprocal voting). Secondly, the common

141 They study the conditions under which particular kinds of governance emerge. For that purpose, they design an experiment with a common-pool resource, such as a field of berry bushes with berries of different colours. Participants can either choose to agree to treat the common-pool resource as a communal property and prescribe harvesting rules for everyone, or they can assign individual property rights to each participant depending on the colour of the bush and, within each property, a participant can choose their own harvesting rule. They add another regime of a 'costly specialisation' in one of the colours of berries. This is equivalent to paying a fine to be able to exploit a dedicated part of the common-pool resource, which could be considered as strengthening identification with and maintenance of the common-pool resource.

provision of a good that is either ‘good’ (generates a benefit for all) or ‘bad’ (avoids something ‘bad’ for all) could depend on the majority’s preference for it to be either good or bad (Isaac et al. 2018). However, this approach would not allow this thesis to study different nuances of breaches at the European level. Considering a member’s breaches as bad does not imply that the majority would vote in favour of a sanction. Considerations for reciprocal voting might impede this.

5.1.3 Criteria for designing an experiment to study the SGP

Based on the insights from Irlenbusch and Sutter (2006) and standard common-pool resource games, as well as the theoretical considerations in Chapter 3, I define criteria, which an ‘SGP experiment’ has to fulfil to plausibly reflect the real-world logic of the common-pool resource, the institutional setting, the interdependence of actors and the economic context that gives choices about compliance a meaningful background. These are the criteria:

1) The structure should mirror a common-pool resource that overlaps with private goods.
2) The structure should be embedded in an economic context, which allows the participant to reveal economic preferences.
3) The game should offer different options for choice that are suitable to reveal basic economic policy ideology.
4) The game should not prescribe a goal to allow different equally meaningful strategies to evolve. The experiment shall equally allow incentives for either running deficits or surpluses, so as to not predetermine specific economic policy ideologies.
5) The means, i.e., the common and the private good, shall allow the experimental design to create a situation of a possible ‘trade-off’ between maintaining the common-pool resource and advancing private prosperity.
6) The means, however, shall be extractable from the private part, and do not necessarily need to be extracted from the common part.
7) There shall be uncertainty about the actual available amount of the common-pool resources to reflect real-world economic developments.
8) Actions could have both positive and negative externalities on other players.
9) The game should allow co-operation towards maintaining or increasing the common-pool resource.

10) The structure should include SGP-like rules, mirroring both the national and the European level: limiting the extraction of the common-pool resource and voting on sanctioning breaching members.

11) The sanctioning mechanism should be gradual and similar to the SGP, with first a warning and only later a fine.

12) The structure should allow communication between participants and engagement in reciprocal voting for forming mutual non-punishment alliances.

13) The game should reflect different sizes of the member states with regard to economic and political power. Participants shall be assigned randomly to these players to avoid self-selection biases.

14) The game should have several rounds to reflect business cycle developments in order to test the 'business cycle' hypothesis.

15) The experiment should recruit participants from different countries in order to test the 'North-South' hypothesis.

The experiment needs a meaningful context on which decisions will be built. Additionally, the experiment has to be sufficiently abstract from the real world in order to avoid biases due to recent experiences, while preserving the SGP's structure and processes. In particular, Raudla (2010) advocates for a necessary level of complexity for employing the common-pool resource frame towards fiscal institutions. In order to be able to study economic preferences and related co-operation, the setting shall be an economic one with a simplified economy that is to be 'governed' by the participant.[142] Central to economic policy ideology is the possibility of a perceived 'trade-off' between preserving monetary stability, and stimulating economic growth when deciding about public finances and deficits. The game shall be designed in such a way that different economic policy strategies are possible, with not a single one to be valued 'more successful' than another. This is crucial to study economic policy ideology, otherwise the experiment would prescribe one single strategy.[143]

142 However, any personal inclination towards or against taking 'leadership' shall be kept to a minimum; also see Section 5.4.2 on social desirability. Therefore, the game cannot model a politician's duties, but has to find a more nuanced task.

143 While this approach is unusual for economics studies, psychologists usually do not prescribe a so-called 'script', i.e., goal (Hertwig and Ortmann 2001).

5.2 Developing the 'cultivation game'

This section presents the core allegory of the experiment and justifies each design feature. The descriptions are tied back to the real-world case of the SGP. This also includes considerations on internal validity to ensure good experimental research.

5.2.1 Overview

The game is about cultivating an apple orchard together with others. The core allegory is an apple orchard, which symbolises the economy. Nineteen farmers[144] cultivate the field together over several rounds. The field is composed of private trees and shared trees, arranged in a circle around a lake. Each farmer cultivates private trees (private good) and trees that they share with each of their two neighbours (common-pool resource). The trees grow apples, some of which they need for their own purposes. Weather shocks, that symbolise economic shocks ('bust' hypothesis), harm the trees. The farmer can harvest apples (i.e., 'tax the economy') that yield energy points, which can be used, in turn, to harvest more apples and to invest in cultivation items.[145] The farmer can choose between different items that either focus on increasing the growth of new apples, or on protecting the tree against future shocks ('economic policy ideology' hypothesis). In this way, the farmer 'governs' the trees and looks after their prosperity. Moreover, the farmer can co-operate with a neighbour to cultivate shared trees ('co-operation' hypothesis). Harvesting more than half of the apples on shared trees (simplified deficit criterion) corresponds to a breach, and all farmers can

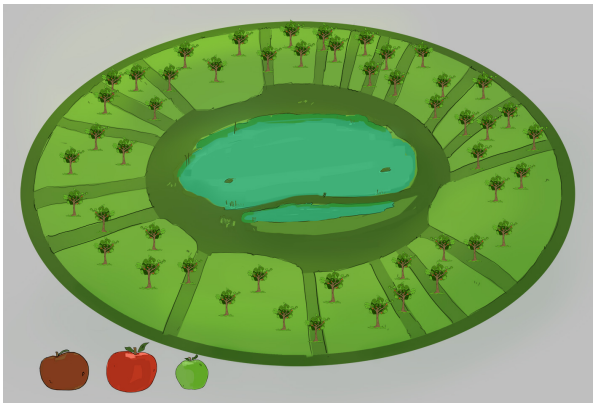


Figure 5.2-1: stylised field; picture: Niklas Zorell.

144 The thesis uses the term 'farmer' to describe the role in the experiment. It uses the term 'participant' to refer to the person who will be playing the role of the farmer when the game is conducted.

145 Additionally, a farmer can spend energy points on 'happiness' points. This option does not draw from any real-world aspect, but is a necessary element in the game to allow excess energy points to be spent without having to let them expire at the end of a round. Otherwise, participants might feel obliged to spend them on harvesting. This would bias results and impede evaluating purposeful harvesting. 'Happiness' points add up over the course of the game, and appears as a 'high score' on the top right corner of the screen. These points do not have any effect on the course of the game.

vote on sanctions. To avoid sanctions, farmers can seal deals with other farmers ('reciprocal voting' hypothesis).

Each round has two stages: A farmer allocates energy points and cultivates the field in Stage 1, while in Stage 2, the farmer can vote to sanction other farmers. Compliance refers to two levels: not to overharvest in Stage 1 and to sanction farmers who have overharvested in Stage 2. This reflects compliance with the SGP at the national (i.e., cultivation) and at the European levels (i.e., co-operation).

In the first round, a farmer receives an endowment of 12 energy points per private tree. Harvesting costs 1 point per apple, and cultivation items cost 6 points each. Participants get assigned randomly to one of three different sizes of farmers, which means that they dispose of either one, two or three own trees ('size' hypothesis).[146] They are invited from four euro-area member states (Greece, Portugal, France and Germany; 'North-South' hypothesis). Participants play individually and separately against 18 'computer' farmers, which follow programmed and a fixed set of responses. The game does not prescribe any particular goal so as not to bias the participants' choices. Instead, they are told to find their own strategy to cultivate the field. There is no 'one best way' to succeed in the game, but cultivation is a matter of what they consider 'best'.

5.2.2 Economic context: the apple orchard

Although at first sight this appears unusually abstract from the SGP, a food-related context seems to be the best fit because for two reasons: firstly, the literature studies easily understandable common-pool resource settings prominently in the agricultural field (see, for instance, Murphy and Cardenas 2004). Secondly, food is a topic that every participant is expected to care about. Thereby, the potential bias occurring from the fact that a participant does not perceive the experiment's topic as sufficiently attractive to fully engage in playing the game can be reduced significantly. The choice for an apple tree is based on three considerations: firstly, the plant should produce fruit that can be consumed – though, to a greater or lesser degree of deliciousness – in three different stages of maturation (green, ripe and windfall; see Section 5.2.4). Secondly,

146 There are 10 small, 5 medium-sized, and 4 large farmers. This corresponds to the number of euro area member states of different sizes according to the number of votes that they have on the Council (see Section 2.1).

the tree's fruit should be a 'standard' fruit in the respective member states to avoid the abovementioned topic bias.[147] For a similar reason, the fruit should trigger as few allergy or dietary concerns as possible.

In the game, there is one apple orchard. This aspect is supposed to give the impression that all farmers are bound together by cultivating this one field. This is similar to the euro area member states, bound together by the monetary union. However, work is divided so that each farmer only looks after their dedicated areas in the field, i.e., private and shared parts.[148] This represents each country's separate economies over which the state has exclusive power to tax and regulate. The shared trees are an abstraction of the 'part' of the country's economy that also affects the economies of other member states through spill-over effects. In aggregate, each farmer's cultivation contributes to the prosperity of the entire field. This gives the impression that a farmer also carries responsibility for the entire field by cultivating the parts they can dispose of. This is important to create motivation for engaging in the sanctioning procedure.

The trees represent the economy and the apples represent money. Apples are consumed to gain energy points, which in turn can be used to harvest more apples and to purchase cultivation items to improve the trees. Apples yield two energy points when they are ripe and one when they are green or windfall fruit. Apples represent the common-pool resource over which there is rivalry among two neighbours that take care of shared tree(s). Provisioning happens through cultivating the tree(s), i.e., improving the basis for monetary stability. The prosperity of the trees determines the possible number of ripe apples that can be harvested.

Figure 5.2-2: apple tree; picture: Niklas Zorell.

147 The fruit type should not change depending on the country as this might introduce a bias. As a result, one could argue that in case the experiment finds differences between countries, they could relate to the different fruit, as the experiment is not exactly the same anymore. Moreover, following Loriot (1976), it is advisable not to use stone fruit (e.g., peach pit) as this might attract stone lice ('petrophaga lorioti'), which would adversely affect fruit supply.

148 In order to make the participant feel immediate closeness to the common-pool resource, each farmer only has two neighbours with whom they share one common part.

Uncertainty exists about the actual number of apples that a farmer can harvest from shared trees exists because both farmers simultaneously choose how many apples to harvest and do not know how many the other farmer would want to harvest. This includes a request game element in the game design.[149] This reflects the economic uncertainty that a government faces when deciding on the budget. Furthermore, the SGP's deficit criterion is calculated as a percentage of the GDP, which also brings in uncertainty about the actual size of the GDP in a given year. Unfavourable harvesting opportunities on one shared part might result in the farmer harvesting more from the part they share with the other neighbour. Thereby, harvesting in one shared part might transmit to the other shared part, and induce overharvesting and conflicts of distribution. This might also lead the respective neighbour to harvest more from their other shared part, and so on. Therefore, all farmers might have an interest in watching the other farmers' harvesting activities on the shared parts, and engage in sanctioning.[150]

With this, the field reflects a setting of a common-pool resource that overlaps with private goods. The private good refers to each farmer's part of the whole field. In a broader sense, all shared trees in the entire field add up to the overall common-pool resource. Each shared part represents the common-pool resource in a narrow sense. Access is non-exclusive to both farmers, but the harvest of apples is rival. The number of trees in shared parts depends on the smaller of the two neighbours reflecting the economic importance of one for the other.[151] This also mirrors the different capacities of each member state, while accounting for their equal potential interest in the common-pool resource and the respective equal responsibility for maintaining it.

The essential element of having the two types of goods overlap is that the means for cultivation, namely the apples, are the same in both the private and the shared parts. With this, it is not essential for the farmer to harvest from shared trees in

149 Taking apples from shared trees is a request and both neighbours' requests get calculated after each round (see programming handbook A.5-2 Section 1.3.5.1.1). If the sum of both requests is larger than the number of available apples, both will get the largest possible fraction of their request. This calculation principle appears to be the fairest possible way to avoid 'valuing' one or the other move as 'better'.

150 These dynamics can be altered by a farmer's size. A big farmer could especially crowd-out harvesting opportunities for small neighbours, whereas a small farmer might not want to 'hurt' a big neighbour so much as this neighbour disposes of more trees to harvest from.

151 While an economically powerful large member state is, in general, important to a small neighbour, the small neighbour will not have a similar impact on this large member state.

order to advance their own trees' prosperity. They could just harvest from their private part. The means to be gained from the common-pool resource come as *additional*. In that regard, at some point, a farmer might want to harvest a great deal more in order to have more energy points with which to purchase multiple items. This mirrors the 'trade-off' between maintaining the common-pool resource (monetary stability) and advancing private prosperity (economic growth). Moreover, since the setting and cultivation items are exactly the same for both the private and the shared parts, the two neighbours could co-operate by investing in their shared part to increase the common-pool resource. With this, overharvesting could also result in positive externalities. Even though this mechanism is not transferable to the real-world case, the effect is comparable.

It is important to note that the overlapping of the common-pool resource with private parts is necessary for tying back the allegory to the 'real world'. In the euro area, each country disposes of the gains of its economy, but as they are bound in a monetary union, a country could, in theoretical terms, borrow on future gains from the other countries' economies by running higher deficits than allowed. In the real world, the member states' borrowing does not equal out, so that, for example, Country A may borrow little, but Country B could borrow the respective amount more. The SGP prescribes borrowing limits, which are equal for all, and disregards any 'equalising' activities such as running surpluses. This is why the shared part of the apple orchard is not constructed as one big tree from which all farmers can harvest.

Finally, an apple orchard and its cultivation allow this thesis to have a much simpler and more accurate design of the experiment than, for instance, trying to sketch a government's economic policy spectrum. One would need to attach specific effects of each policy choice, be it specific investments, the level of social security, expenditures for defence, or introducing special taxes. Deciding which option would have what effect on the growth of the economy and government bond interest rates would be largely deterministic, and would not allow this thesis to study different economic policy ideologies. Moreover, the context of the game would not be abstract enough, and participants might respond in a biased way towards past experiences of their home country. Therefore, the apple orchard provides a simple yet meaningful abstraction of an economic environment.

5.2.3 Participants: farmers show economic preferences and co-operation

The game simulates a simple business setting and the trees could also survive without the farmers intervening. A farmer can, however, harvest and invest and, thereby, 'govern' the trees[152]. By co-operating with their neighbours and voting on sanctions, the farmers can maintain, or even improve, the common-pool resource. Thereby, the experiment is basically about extracting gains from the economy through taxing (harvesting from the trees) and redistributing them in the form of regulations and investments ('cultivation'). In particular, there are several cultivation items that reflect 'stimulus'- and 'prevention'-oriented policies. With this, the participant can express their preferred cultivation strategy, which over the course of the game reflects their economic policy ideology.

Likewise, the computer-player farmers do not follow one, but different strategies so that the participant does not get the impression that there is just one 'right' strategy. All moves are pre-programmed so that the game is comparable across participants. Each computer-farmer harvests, invests, votes and seals deals according to three different strategies each.[153] In case the computer-player needs to respond to the participant, it follows the tit-for-tat strategy, by which it mirrors the participant's behaviour. This is a standard strategy in game theory (Axelrod 1984). The main idea is to present the participant with a large variety of different ways to approach the game and, in particular, not to show any of them as the 'best' strategy.

An essential prerequisite for the experiment to work is to incentivise the participant to consider their actions in light of the common-pool resource, i.e., the cultivation of the entire field. In particular, the experiment has to incentivise the participant to engage in voting. There are two aspects for this. Firstly, computer-players that follow extensive overharvesting are just one farmer away from the participant.[154] This is sufficiently close to the participant to make them aware of possible implications of passing on a shortage of apples. Secondly, having different

152 In order to keep it really simple, the experimental design only includes farmers. In contrast to the real world, there is no European Commission and no European Central Bank and, accordingly, no monetary policies intervening as the game is based on barter. Furthermore, there is no trade possible between farmers, but they can co-operate in a 'purer' way.

153 See programming handbook (Appendix A.5-2), Section 1.3.8.

154 With this, this thesis follows Boosey (2017) who finds that in case free-riders are direct neighbours, there is a loss of co-operation, and this set-up biases the participant's responses.

types of computer-players spread out across the field signals to the participant that their responses to that variety, i.e., voting would be effective. This could trigger the participant to engage in voting. All this is important to allow the participant to react in the way they believe is appropriate.

5.2.4 Tokens: apples as means

Apples yield energy points and, thereby, they are the source of the tokens used in the game. The distinction between apples and energy points is essential for the logic of the experiment for two reasons. Firstly, tokens, i.e., energy points, are invested in actions, of which harvesting is one. With this, there is a conflict of allocation between harvesting and investing, i.e., between 'using' monetary stability and stimulating economic growth. Secondly, in this way, the farmer can transmit tokens from one round to the next.[155]

A tree has three kinds of apples: small, big and fallen apples, corresponding to their stage of maturation, green, ripe or windfall. In each round, a tree grows new small apples. These apples grow big and ripe in the next round, fall off the tree in the round thereafter, and finally become fertile soil in the round after that. Big apples yield double the energy points of small or fallen apples. The different kinds of apples and their growth process symbolise the economy's development over time. This aspect makes the rounds in the game successively build, introducing a forward-looking perspective. This creates a past that determines the present from which the farmer could learn to act in the future. A tree has a natural growth rate[156] which reflects the economy's growth without state intervention. However, the growth rate can be

155 Harvesting small or fallen apples only transmits energy points from one round to the next without actually increasing the farmer's level of energy points. In contrast, harvesting from ripe apples increases the level of energy points. Energy points cannot be stored, and unused energy points will expire at the end of a round. This is a necessary simplification in order to keep the cycle of growing apples concise, and to cause the necessary conflicts of allocation between harvesting and investing, so that choices remain an expression of preference. This makes actions comparable between rounds and participants. Moreover, it is not possible to store apples or energy points because this would mean that apples would be taken away from the economy. In the real world, if a government runs surpluses (i.e., income that it does not spend), it can decide to either cut taxes, or to build reserves, i.e., invest money in financial products. Assuming that a government (or the sellers of the financial products) do not invest outside the euro area, the money stays in the euro area's economy and, hence, the apples remain with the tree. In the game, this corresponds to leaving ripe apples on the tree so that they will fall down in the next round, or leave them on the ground so that they will become fertile soil in the next round.

156 See Appendix A.5.2.1.

affected by cultivation items and exogenous shocks through weather events (see sections below).

The meaning of apples in the real-world example of the SGP is the economy's gains. While the transmission of apples into energy points, of course, does not directly translate from the real world, the effect is comparable. Considering apples as a business' gains, ripe apples correspond to actual gains in the present, while small and fallen apples correspond to likely future gains and reserves, respectively. When fallen apples become fertile soil, gains from previous years are 'reinvested' in the business. A participant can influence the economic growth of a tree, i.e., the number of available apples, with cultivation items. Accordingly, as ripe apples are converted into the highest number of energy points, they correspond to the common-pool resource over which there are conflicts of distribution. With this, ripe apples find their real-world correspondence in 'stable money', which is also conceptualised as a common-pool resource with conflicts of distribution.[157]

There is one specific difference between the shared trees' notion and the euro area's monetary stability as common-pool resource. In the game, it is perfectly clear that if a farmer takes one apple, their neighbour will have exactly that apple less at their disposal. The neighbour will not be able to take that very apple. In the real world, this is part of the co-operation problem. It is not clear how 'much' deficit one government can run until it increases borrowing costs for the others. Additionally, the exact moment that triggers the crowding-out effect is extremely vague. In the game instead, this vagueness is included explicitly with the request game element for harvesting from shared trees. At the same time, the risk of spill-overs arises when a farmer does not find enough apples to harvest from shared trees with one neighbour because of this neighbour's overharvesting. They would revert to their other neighbour for harvesting and, as a consequence, might have to overharvest to satisfy their harvesting needs. With this, the original overharvesting spills over to a third neighbour.

157 There is no inflation in a strict sense. Not having inflation is a necessary simplification in order to prevent judging one particular option in the game as inflation increasing or decreasing. Even though this is one of the main considerations of the real-world SGP, it is not suitable to include it in the game, as this would make it too complex to allow parsimonious interpretations.

5.2.5 SGP-like rules

5.2.5.1 Overharvesting

A request to harvest more than half of the available apples on a shared tree is defined as overharvesting. This corresponds to breaching the SGP's deficit rule.[158] Overharvesting from shared trees is, metaphorically speaking, 'taxing' the other country's economy, which corresponds to crowding out resources.[159] This perspective, however, could be short-sighted given that a farmer could also invest in cultivation items on shared trees, from which also the neighbour would benefit.[160] Even though the SGP rules do not apply in anticipating such possible outcomes, the SGP process allows member states to take this into account when voting on sanctioning breaches.

One could argue that overharvesting is an expression of fairness considerations. However, since a farmer needs apples for their own cultivation purposes, overharvesting is based on an economic need. In particular, it is a reflection about the trade-off between individual purposes and maintaining the common-pool resource. By inducing a crowding-out effect, overharvesting corresponds to a destabilising effect that is comparable to what the real-world SGP seeks to prevent. Moreover, the rules only prescribe a flexible limit as they also take into account the two apples that have to be left on the ground, and the voting rules allow the participant to temporarily trespass the limit similar to the real world. The limit is flexible in order to avoid prescribing what is a 'good' or 'bad' economic policy strategy.[161]

---

158 The 3% deficit rule is simplified to the basic rule that no more than 'half' of the ripe apples on a shared tree should be harvested – keeping in mind that two apples always have to remain on the tree to become fertile soil. Despite the simplification, the numerical limit is comparable – disregarding the fact that, in the real world, the basis is the member state's and not the entire euro area's GDP. Considering all 24 shared trees together, taking half of one tree corresponds to 1/48 which is about 2.1%. Considering all shared tree areas, numbering19, taking one half would correspond to about 2.6%. This is comparable to the 3% deficit rule.

159 In abstract terms, the less other members of the euro area can borrow from the market at a specific price, the more they have to tax their own economy.

160 To reiterate the argument from above, 'overharvesting' does not correspond to equating public revenues and expenditure, i.e., running a deficit as opposed to a surplus or balanced budget, in the real-world SGP. Instead, it corresponds to using more resources (i.e., apples) from the common-pool resource (shared trees) than has been prescribed by the rule.

161 One could have included another option in the game that would have introduced the SGP's debt limit to reflect that, at some point, the debt level might turn unsustainable. This would have introduced the threat of potentially being put in a situation where all farmers would need to bail-out another farmer. This would have introduced the potential non-credibility of the bail-out clause that might have driven strategic consideration for complying with the SGP rules in the past.

5.2.5.2 Sanctioning mechanism

All farmers can vote to sanction any other farmer if they consider that they have not treated shared trees 'well'.[162] This very open formulation is supposed to avoid triggering social desirability, in particular pure rule compliant behaviour. Instead, the intention is to measure purposeful rule adherence based on whether the participant considers the overharvesting rule meaningful for cultivation. The participant can also abstain from voting. This reflects the real world where even though the SGP rule is strict on paper, governments get to decide whether or not to sanction in the Council. The option that the participant can vote on any farmer, even if this farmer has not actually overharvested, is a necessary deviation from the real world where the basis of any Council decision is a respective proposal by the Commission. In the game, this is kept open in order to avoid social desirability or learning, that might make the participant inclined to sanction even though they would not, in fact, want to vote at all.

Sanctioning works according to three stages: The procedure starts with a warning (yellow), a fine (red) is only available for voting in the round after a farmer gets a warning, and the lifting of the procedure (green).[163] Warnings and fines can be repeated infinitely until farmers decide to close the procedure (green). If the farmers decide to impose a fine on a farmer, this farmer has to give apples away.[164] These apples disappear and do not benefit the other farmers in any way, in order to avoid participants voting 'red' just to gain apples.

However, this would have made this experiment even more complex, and the intention has always been to present the simplest experiment.

162 See detailed technical descriptions in Appendix A.5-2 Sections 1.3.5.2 and 1.3.8.2.2.

163 'Warning' mirrors the opening of an EDP and the first 'formal' stages. The fine reflects both the non-interest deposit and the fine in the SGP. It is important not to start with a fine right away (in contrast to other experimental studies with sanctioning mechanisms), as this might bias results significantly (this thesis follows Irlenbusch and Sutter 2006). As a 'warning' gathers several options, farmers can vote 'yellow' as many times as they want. This is similar to the real world, where governments seem reluctant to proceed with the other options ('fine'), and favour repeated 'warnings'. Eventually, as in the SGP, a procedure has to be closed. Having an option for closing the procedure – instead of having the procedure closed 'automatically' in the first round in which a farmer does not overharvest – is necessary in order to avoid strategic patterns of overharvesting.

164 The volume of the fine is 1 ripe apple per 10 ripe apples that were available at the beginning of the round on all private tree(s). This appears relatively small, and might not hurt a farmer that much. However, giving away one or two ripe apples might thwart plans for cultivation. In the real-world SGP, the fine amounts to 0.2% of GDP. Compared to a member state's total budget, this fine is rather small financially. However, given that a fine would additionally have political implications, the level of the fine in this game seems reasonable. Irlenbusch and Sutter (2006) use a much higher fine, which does not seem adequate for the real-world application of this game.

Similar to the SGP, one of the three options needs to gain a qualified majority – without the member state in question. Each farmer's vote share is weighted according to their size, similar to the member states' vote shares in the Council.[165] However, in the game, the participant will always be pivotal. The participant does not know this, of course. For the study of the participant's behaviour, however, it is necessary that their actions do not get constrained by 'learning' through others, prescribing what 'should' have been done.

Programming under what conditions the participant gets sanctioned strikes a delicate balance. On the one hand, the participant has to be sanctioned in the event of a breach, so as to show that their decisions in the game have indeed an effect and that it is not a hollow game. On the other hand, the participant shall not be sanctioned too strictly so as to avoid any trigger for conditioning or social desirability. Therefore, in case they overharvest, the others would voice a warning twice before agreeing on a fine.[166] Moreover, this seeming 'undecidedness' of the other farmers to sanction the participant makes it more difficult for the participant to predict their moves. This reflects real-world situations well. The participant can seal deals with other farmers to avoid sanctions, i.e., engage in reciprocal voting. However, only some of them will be open to deals, depending on how many apples the participant overharvested.

5.2.6 Course of the game

5.2.6.1 Weather reflects the business cycle

The game has 11 rounds.[167] Over the course of the rounds, the field undergoes a stylised business cycle.[168] Organically, there is a natural growth of small apples on a

165 See Appendix A.5.2-1 Section A.5.2.4 for details. The voting rules mirror the SGP rules and the calculation of the qualified majority as they were used since the introduction of the euro.

166 See Programming Handbook (Appendix A.5-2, Section 1. 3.8.2.2).

167 In order to avoid end-of-game biases or response fatigue, the participant is told that the game has 14 rounds, when actually it has only 11. Seemingly infinite repetition is an important aspect in game theory and experiments, in order to avoid strategic biases of behaviour (see Axelrod 1984, Duffy 2015: 12ff., Dal Bó and Fréchette 2018). Firstly, if participants know that the game will end at a certain point in time, they tend to change their behaviour during the end. They try to get as many gains as possible because they do not need to fear possible repercussions anymore. Secondly, from a game theoretical perspective, if a game in which the end moves are crucial is repeated for a finite number of times, the behaviour in the end nodes will be transposed to the very first node (known as 'backward induction'). This is because rational players expect their counterparts to breach at the last node and therefore, will breach one node before, and so on and so forth.

168 See Appendix A.5.2.5

tree, as described above. With a weather shock, a tree experiences a bust as the number of small and ripe apples is cut, and some cultivation items are broken.[169] In Rounds 4, 7 and 10, weather shocks hit the field. Rounds 5, 8 and 11 symbolise the recovery phase, after which follows a boom phase (rounds 6 and 9).[170] A participant's response to bust rounds would be particularly insightful, as it reveals preferences for the most urgent and effective economic policy. The repetition of the three business cycle rounds serves as a double-check of a participant's consistency of actions.

In order to avoid the effects of learning in the game, the first shock is a thunderstorm, the second a drought and the third again a thunderstorm. Weather reports are a crucial part of the game design. In each round, the weather report briefly describes the current weather, and gives a slight outlook without explicitly announcing a shock. Hence, it is up to the participant how seriously to take these into account for their cultivation decisions. This corresponds to economic forecasts and government's evaluations of these for their policy decisions.

5.2.6.2 Cultivation items

A farmer can invest in a number of items to support a tree (see Table 5.2-3).[171] All items have the same 'price', 6 energy points, so as to eliminate potential effects of costs on purchase decisions. Items apply to one tree only.[172] This allows variation and different combinations of items.

Investing in items corresponds to a government running economic policies. As a simplification from the real world for the purpose of this game, there are three groups of items. Firstly, harvesting equipment facilitates harvesting. This item serves as a double-check for efficiency. For instance, it is most efficient to purchase a bucket early

169 The shock only hits the trees (available apples and invested items) and not the farmer directly. The farmer can fully dispose of the energy points that they had harvested in the round before. This aspect is particularly important, as it mirrors real-world scenarios. Whereas a bust hits the economy directly, the government still has some financial leeway due to already collected taxes (or borrowing). This allows the state to step in (to some extent) when the economy is in a bust.

170 As the game has 11 rounds to keep it sufficiently short, round 'boom' will only be repeated twice and not three times like the rounds 'bust' and 'recovery'. Round 'boom' is of lesser importance than the other two rounds for investigating how participants deal with a shock.

171 Some of them are available for purchase from the beginning, and some only become available over the course of the game (so as to reduce complexity for the player). The options are designed so as not to create biases, which could affect the overall results of the experiment. In particular, the different options should not result in exponentially increasing inequality between farmers of different sizes, but instead keep them within overlapping boundaries according to their size.

172 Except the shepherd applies to all trees that P can dispose of.

in the game in order to enjoy the full potential of energy points for harvesting. One could think of a bucket as an investment in efficient tax administration. Secondly, 'stimulus' items increase the natural growth rate. Thirdly, 'prevention' items reduce the impact of shocks on the trees. This corresponds to the two core economic policy approaches in the hypothesis: 'stimulus'- and 'prevention'-oriented policies. Each group has two core items: manure and a beekeeper for stimulus, a water irrigation system and a well for 'prevention' items.[173] Furthermore, both groups each have one item that promises huge effects, but is built on a specific risk assessment. A ladder promises a large number of new small apples, but success is rare.[174] A shepherd would keep sheep away from the field, but it is not certain whether sheep would indeed come to the field and eat all the fallen apples, so that none would remain as fertiliser for the tree.[175] Hence, both items serve as a check with regard to the level of risk aversion.

Table 5.2-3: Cultivation items.

	Round	**Description**	**Effect**
Harvesting equipment:			
Bucket	1	Facilitates harvesting; a bucket increases the number of apples which can be harvested with 1 E.	Harvesting from a tree with a bucket: 1 E yields 2 apples (instead of 1)
Stimulus:			
Manure[176]	1	Improves the conditions under which a tree can grow apples; each item increases the natural growth rate temporarily.	A tree with manure or a beekeeper: NG+3 for the next round A tree with manure & a beekeeper: NG+6 for the next round
Beekeeper	5		
Ladder	5	Conducts a trial to cross-breed blossoms and therewith discover a new breed that boosts the growth of small apples.	A ladder applies to each individual tree and lasts for 3 rounds. There will be no effect; the participant only believes there might be a slight chance it is successful.

173 The items of each group appear sequentially and, thereby, serve as a double-check to verify the consistency of the participant's policy approach.

174 The ladder appears in Round 5 for the first time, while its 'counterpart 'shepherd' only appears in Round 8. This is in order to keep the game going and introduce the same number of items for each of the two core types ('prevention' and 'stimulus'). Whereas the 'prevention' type gets the items 'repair' and 'well', the 'stimulus' type gets 'beekeeper' and 'ladder' at their disposal. Therefore, the shepherd has to come a little later.

175 Sheep eating fallen apples would lead to less fertile ground for the trees. Sheep have no business cycle meaning, but rather correspond to a potential adverse shock that might or might not affect the country's economy. It is upon the participant to assess the risks and act accordingly.

176 Manure and a beekeeper will only be fully beneficial together with a bucket. Otherwise, the item is too costly with respect to the gains. One item costs 6 E and produces three small apples. When they become ripe, they would yield 6 E. However, before, they have to be harvested and this costs 3 E without a bucket. Although these numbers seem harsh, calculations with other numbers have shown that only these numbers provide for a good balance between reducing variation and allowing preferences to be revealed, i.e., keeping resources scarce to induce a conflict of interests.

Prevention:			
Water irrigation system[177]			
- Part 1 - Part 2 - Part 3 - Repair - Well	1 2 3 4 5	Reduces the impact of bad weather by building resilience against shocks and thereby reduces losses. A system will only be operational once three parts are built; it is repaired after a shock and a well is installed against a drought.	In case two fallen apples remained on the ground previously.[178] small apples: 8–1 + (NG so far) + effects of 'stimulus' items ripe apples: –1/3
Shepherd	8	Prevents sheep from eating fallen apples; protects the trees' NG.	A shepherd applies to all of the farmer's trees (own and shared). There will be no effect; the participant only believes there might be a slight chance of sheep coming.

Legend: 'E' means energy points; 'NG' means natural growth rate.

Source: own description.

Whereas 'stimulus' items last only one round and can be purchased each round again, a water irrigation system would last forever in the game and has three consecutive parts before it becomes fully operational. This reflects that, usually, supply-side oriented policies need some time for implementation before their effect unfolds. Growth is supposed to set in at a slower pace, but be more sustainable in the long run. In contrast, demand-side oriented policies are supposed to stimulate growth quickly and over a short time period. Moreover, an incomplete water irrigation system would break after some time if no subsequent part was constructed. This introduces two important aspects: Firstly, this offers the option to show an interest in sustainability, long-term planning and, thereby, sustainable growth in contrast to short-term stimulus. Secondly, this provides the opportunity for mis-investments. If a farmer engages in a project that they cannot in fact afford, or which they are not fully interested in completing, the project will not prove economically successful and will, hence, not increase the tree's strength.

5.2.6.3 Treatments

This experiment has three treatments: economic bust, size, and nationality. The treatment variables are included differently in the experiment. The first treatment

177 The option 'water irrigation system' consists of three consecutive construction phases. Only after the last part has been constructed will the facility start to work. Before that, the single parts are useless.

178 See Table 6 in the programming handbook (Appendix A.5-2).

variable comes as a within-game variation. It simulates economic shocks and defines booms and busts over the course of the game. With this, it is the same for every participant. Secondly, the treatment variable, 'size', is randomly assigned to a participant right before the game starts. The third treatment, 'nationality', comes as a pre-selection for the participants. Only participants from the countries of interest would be invited to play the game, i.e., France, Portugal, Greece and Germany.

5.2.7 Comparability of choices

Tangibly interpreting choices in the game as purposeful expressions of preferences and not as a result of weighing risks is necessary. Therefore, the impact of choices in the game has to be clear: the means-end relationships of policy choices, i.e., cultivation items, are clear and not subject to risk. Even though this is not necessarily the case in the real world, it is important to allow the participant to take decisions purposefully. This ensures that the participant can work out and implement a strategy and does not need to worry about whether their actions would in fact have the desired effect. Therefore, also economic shocks are exogeneous in the game, even though this might not be true for the real world where policy options might facilitate shocks.

Moreover, the experiment is designed in such a way that the allocation of energy points reflects preferences over different choices. Energy points can only be allocated once. For instance, if a participant decides to invest in a beekeeper, simultaneously they decide against investing, for instance, in a water irrigation system. Therefore, the scarcity of energy points shows clear preferences for the investments taken in contrast to those that are not.[179]

5.3 Extracting measures of interest

This section discusses how to extract variables from the many elements in the game. It starts with rule behaviour as the dependent variable (DV); it follows 'economic policy ideology' as the main independent variable (IV) and, finally, treatments and other independent variables. Given the complexity of the possible moves in the game,

179 In order to define these conflicts of allocation, I designed a baseline strategy, which served as a basis for the game's development (see Table A.5-1-2 in the appendix).

the analysis works with types of strategies. A strategy describes all relevant moves in the game by a participant.

5.3.1 Rule compliance

The dependent variable 'rule compliance' is analysed with respect to harvesting behaviour on shared trees, and voting. Overharvesting is defined as requesting more than half of available ripe apples on a shared tree. Compliance with the voting rules is defined as correctly voting for a sanction, as well as against a sanction. For the analyses, it is also possible to construct a typology that combines both Stage 1 and Stage 2 compliance (see Table below from Section 3.3.1).

Table 5.3-1: Rule Compliance Typology.

		Voting (stage 2)		
		Comply	**Breach**	**No voting**
Harvesting (stage 1)	**Comply**	*'respecting'*	*'strategic breach'*	*'indifferent'*
	Breach	*'free-riding'*	*'active breach'*	*'ignorant'*

Source: own description.

The logic of the definition of a Stage 2 breach is comparable to the real-world SGP. The game's variable combines both incorrectly voting against a sanction, i.e., with a view to diminishing the rules, and incorrectly voting for a sanction, i.e., with a view to implementing stricter rules. The motivations appear to be opposites. However, they are similar in that they both seek to interpret and apply the rules according to their preferences, and to impose their preferred way of governing the common field.[180] This idea of discretion is basically what one can observe with the real-world SGP. In the real world, though, governments cannot easily advocate for a stricter implementation of the rules as the rules do not allow this. However, they can voice reform needs in that direction.

180 I also ran the regressions for Stage 2 with dependent variables that only code either of the two as '1'. This does not yield markedly different results.

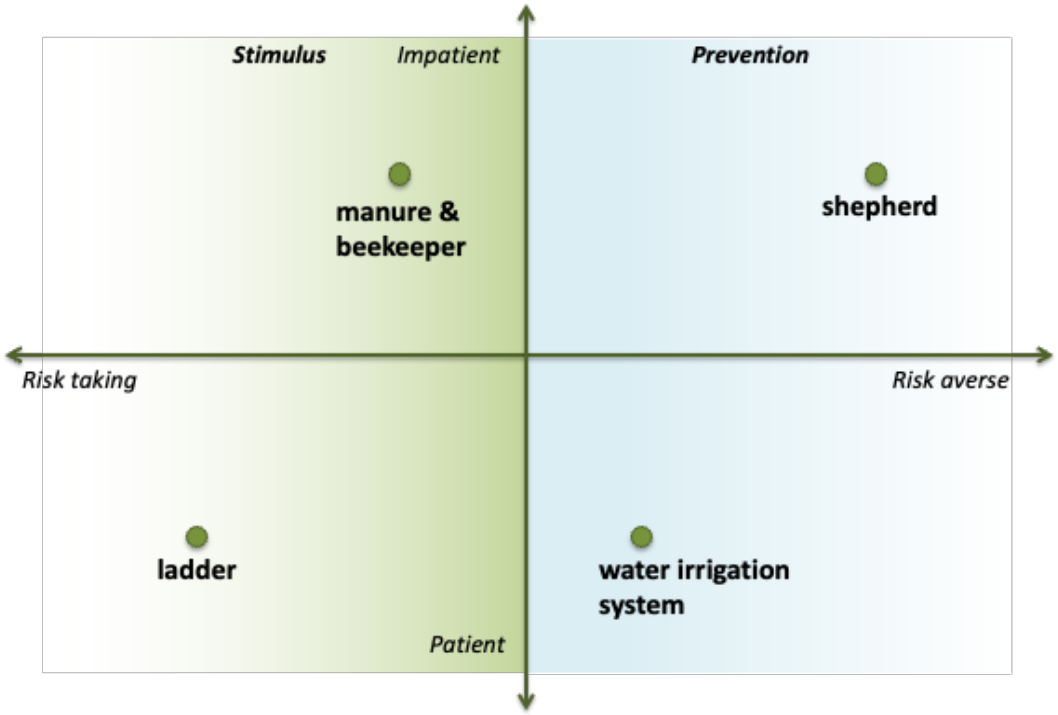
5.3.2 Economic policy ideology

The way of cultivating the field in Stage 1 reflects the economic policy ideology hypothesis. In particular, the investment strategy reveals a preference towards either 'stimulus' or 'prevention' items. Additionally, there are measures for the notions of risk aversion and patience in order to disentangle them from the policy notion of 'stimulus' and 'prevention' items. Moreover, harvesting behaviour reflects preferences for sustainability.

5.3.2.1 Investment strategy

A participant is either assigned to preferring 'stimulus'- or 'prevention'-oriented investment items. This means that a participant either prefers investments that stimulate growth quickly and work as a 'stimulus'-oriented policy, such as manure or a beekeeper, or investments that strengthen the tree so that it becomes more resilient (and loses fewer apples) in busts, and works as 'prevention'-oriented policies, such as a complete water irrigation system. Participants are assigned to 'no type' in case they have not invested in such items, or when a clear preference cannot be observed.

Figure 5.3-2: Investment items and their notions.



Source: own illustration.

In order to disentangle notions of risk aversion and patience from the investment type, the game offers the items of a shepherd and a ladder. The dimension of risk aversion correlates with the distinction between ‘stimulus’ (risk taking) and ‘prevention’ (risk-averse). On the other hand, patience refers to the time horizon that the participant is willing to accept. For instance, if the participant prefers ‘stimulus’-oriented investments, they can show high patience and long-term orientation when purchasing a ladder. The figure below illustrates the relationships.

5.3.2.2 Harvesting strategy

First of all, it is noteworthy that harvesting behaviour is an indicator of efficiency. As ripe apples yield more energy points, a rational participant would always take ripe apples instead of fallen or small apples, as long as ripe apples are available. Furthermore, a participant would rather take fallen than small apples, as small apples – in contrast to fallen apples – would continue to grow in the subsequent rounds.

Accordingly, harvesting shows preferences for sustainability. The participant might eventually face a conflict of interest when they have already harvested almost all the ripe apples on their own trees, and almost half of the ripe apples on shared trees. They need to decide between preferring sustainability or avoiding sanctions. On the one hand, there is the advice to always leave two apples on the tree to ensure that they become fertile soil. Moreover, they might want to avoid harvesting small apples as they would not grow ripe in the next round. On the other hand, there is the risk of getting sanctioned, and this puts at risk co-operation with their neighbours when investing in shared trees. There is no rational best choice, but instead the participant shall express preferences about what they consider more salient.

Therefore, sustainable harvesting reflects in always leaving two apples on the ground and never harvesting a small apple. Leaving two ripe apples on the tree indicates a preference for sustainability, as this supports the natural growth rate. Harvesting small apples shows a tendency for short-sightedness, which stands in contrast to sustainability.

5.3.3 Treatments and strategic interactions

This experiment uses three treatment variables: economic bust, size, and nationality, and allows strategic interactions to drive motivations to comply. Based on the hypothesis, 'bust' rounds correspond to the treatment group over 'recovery', and 'boom' rounds to study the effect on behaviour. Speaking in experimental terms, the experimenter 'manipulates' on this variable to induce behaviour. As regards 'nationality' and 'size', there is no distinction between treatment and control groups, as one could not meaningfully argue for using one group as control group over the other.

'Strategic interaction' has two seemingly opposing notions. On the one hand, it relates to playing strategically as shown with reciprocal voting. On the other hand, it relates to co-operation for cultivating shared trees together with the neighbour both in terms of harvesting and investing. This reflects the participant's engagement with the common-pool resource in a narrow sense. For the first notion, the number of deals that were sealed and adhered to is used, i.e., cases where the participant indeed voted against sanctions towards their partner.

The second notion is reflected in a combination of harvesting and investing as shown in the table below. Here, it only counts if a participant harvests at all from a shared tree. Overharvesting is not considered, in particular. This is important to distinguish because, for instance, a participant might perceive overharvesting as 'legitimate' if they also invested in the shared tree. The co-operation strategy distinguishes between four types: '*co-operative*', '*exploiting*', '*generous*' and '*disregarding*'. Participants will be assigned to one of the types depending on whether they harvest from or invest in shared trees.

Table 5.3-3: Co-operation strategy.

		Investing	
		No	**Yes**
Harvesting	**No**	*'disregarding'*	*'generous'*
	Yes	*'exploiting'*	*'co-operative'*

Source: own description.

5.3.4 Translating hypotheses into expectations

The hypotheses put forward in Section 3.3 are adapted to match the experimental design and to be able to test the experimental results. A comparison table can be found at the end of this section.

5.3.4.1 Null hypothesis

Firstly, the null hypothesis states that all participants would similarly overharvest and vote. Participants would perceive one specific way of playing the game and deciding on one option over another as the most rational way of playing this game. Participants would be considered to have the same preferences, to follow the same goal, and purposefully decide on the same strategy to maximise their utility.

For the specific case of overharvesting and avoiding sanctions by using alternating patterns of breaching and complying, there might be a most rational way to pursue. However, such a strategy might also be subject to the preferred cultivation strategy. The design offers two equally successful investment strategies by focusing either on 'stimulus' or on 'prevention' items.[181] This thesis makes the claim that in economic policies, there is no single best way but instead, economic policy choices build on ideology. Accordingly, participants would have different preferences and goals and would, hence, make different choices to maximise their utility. Consequently, this thesis aims to provide some empirical evidence and explanation for seemingly irrational choices.

5.3.4.2 Economic policy ideology

Section 3.3 discusses in detail the level of abstraction of the *economic policy ideology hypothesis*, which investigates 'ideology' by focusing only on 'stimulus'- and 'prevention'-oriented economic policies. Additionally, there are two hypotheses on personal traits that might intervene with measuring 'ideology'. These are preferences for 'risk aversion' and 'sustainability'. Given the thick economic context of the experiment, whose allegory has been justified above in detail to sufficiently reflect an

181 The mixture of both might not be equally successful as it might cost more energy points to follow a little of both strategies as the water irrigation system only reveals its full potential once all three parts are built.

SGP-like setting, the 'ideology' hypothesis can be translated into similar expectations for the experimental results. Accordingly, participants that prefer 'prevention'-oriented items over 'stimulus'-oriented items are more likely to endorse the idea of restricting budgets and follow the rules. Participants are expected to reflect 'ideology' when deciding on investment options. If investments are consistent over the course of the game, one can conclude this to reflect an investment strategy and accordingly, a 'stimulus'- or 'prevention'-oriented type.

Similarly, one can expect that risk-averse and 'sustainability'-oriented participants are more likely to follow the rules during both stages. Firstly, as regards the *risk aversion hypothesis*, as outlined above in Section 5.3.2.1, investments in shepherds reflect risk-averse and investments in ladders risk-taking preferences. As intended and justified in Section 3.2.4, these traits coincide with a notion of patience in order to ensure measuring risk aversion and not, by accident, patience. One can assume that the notion of risk aversion is more salient in an economic context than the notion of patience. Therefore, a shepherd reflects short-term and a ladder long-term orientation. Assuming that risk aversion could affect compliance is also supported by the experimental literature. In co-operative experiments, studies show risk aversion to drive expectations towards what others might do (see Morales et al. 2004). With respect to this thesis' experiment, this would refer to avoiding getting sanctioned, or avoiding breaching the voting rule.

Likewise as regards the *sustainability hypothesis*, 'sustainable' orientation towards cultivating the field reflects preferences for avoiding overharvesting and enforcing similar behaviour on the other farmers by using the voting rules. Such orientation makes it more likely that the rules are endorsed and followed by principle because the purpose seems to be shared and understood as the 'right' policy to cultivate and govern the field. This reflects a similar logic to the hypothesis for the real-world states.

5.3.4.3 Economic necessity and political capacity

Three treatment variables test for the prevalent explanations in the literature: economic shocks, size and nationality. Drawing from the hypotheses set out in Section 3.3, they are rather straightforward to adapt to this experiment.

Applying the *business cycle hypothesis* is rather straightforward as regards overharvesting.[182] However, the expectation on sanctioning behaviour might play out differently. As formulated in the hypothesis, a participant would grant the same flexibility in applying the rules in bust times as they would hope for themselves. They would not vote for sanctions if they also do not comply with the overharvesting rule. However, it could as well be the case that even though a participant overharvests in bust rounds, they would not want others to do the same in order to ensure high maintenance of the common-pool resource. In that regard, the awareness of the consequences for the common-pool resource might be felt more immediately than in the real world. One needs to consider this for the interpretation of results.

While the literature's *size hypothesis* builds on considerations about economic and political power[183] and related leeway for strategic voting, the experiment's size hypothesis only reflects economic power in terms of number of private trees. While also in the experiment, political power, i.e., the number of votes, depends on size, the political use of size in terms of strategic voting is reflected in another variable in the experiment, namely the number of sealed deals. This facilitates investigating only the role of the economic starting conditions with the 'size' hypothesis, while leaving the actual 'use' of these starting conditions to the reciprocal voting hypothesis (see below).

Due to the fact that a farmer of each size has, in relative terms, the exact same economic needs and opportunities (as these are defined per each individual tree) and the same number of energy points as initial endowment, one would not expect 'size' to have any effect on rule compliance in addition to its 'use' reflected in the reciprocal voting hypothesis. Nonetheless, a large farmer disposes of more leeway to engage in targeted cultivation, for instance, by investing all energy points into one tree. Moreover, the shared trees reflect the mutual economic importance of one for the other. This serves as a proxy for their interconnectedness due to, in real-world terms, their trade relationship. This artificial starting point cannot, of course, be found in the

182 One could argue that while in the real world, a breach is measured also using the structural balance and, hence, based on discretionary policies rather than on automatic stabilisers, in the cultivation game, there is only the discretionary element. However, if one assumes that in an economic downturn also automatic stabilisers could be subject to policy changes, one can conclude that any economic policy response to an economic downturn could be considered discretionary, i.e., purposeful.

183 This relates to the size of a country's population and economy.

real-world SGP context. However, it helps to disentangle the role of size in terms of political and economic power in the experiment.

The literature's *'North-South' hypothesis* refers to a number of economic and political arguments, and concludes that governments in countries located in the north would be more fiscally prudent than their counterparts in the south. As discussed in Section 2.3, overall, this reflects economic or political capacity to accept and follow budget limits. While the experiment cannot mirror the thick capacity issues, it seeks to control for a number of the economic issues. This leaves cultural aspects, collective experiences, for instance, with state institutions, and economic policy traditions to the experiment's treatment variable 'nationality'. In that regard, the real-world hypothesis can be transferred to the experiment. As argued in Section 4.3.2, students can be considered representative of their cultural background. Additionally, 'nationality' could also reflect other cultural aspects, which might affect co-operative behaviour (see experimental findings in Cason et al. 2002, Prediger et al. 2010, Gehrig et al. 2019, van Klingeren 2020). One can control for this to some extent by using variables from the survey. However, one cannot control for aspects of which one is unaware. One has to keep this in mind when discussing the results.

5.3.4.4 Strategic interactions

Applying the *co-operation hypothesis* and the *reciprocal voting hypothesis* is rather straightforward. Accordingly, engaging in interactions with other farmers could have diverging effects on rule compliance. Co-operation on shared trees, i.e., maintaining the common-pool resource, is contrasted with forming non-punishment alliances. Investing in shared trees is supposed to reflect that a participant cares about the common-pool resource and, therefore, is less likely to breach at both stages[184].

184 Experimental literature supports this as they find that those who co-operate are more willing to sanction than others even in case it is costly (Fehr and Gächter 2000). Moreover, co-operation entails considerations on reciprocity, which is called 'conditional co-operation' (see for a discussion Fischbacher Gächter 2010, Hartig et al. 2015, Spiller et al. 2016, Boosey 2017, see also Diekmann 2009: 209ff.). Fischbacher et al. (2001) find that co-operative behaviour for provisioning a good depends on and increases in parallel to expectations of how much other participants co-operate. This can even lead to higher co-operation than expected, which is a common theme in public good experiments (Spiller et al. 2016). In contrast to public good games, common-pool resource experiments usually provide similar levels of co-operation than expected. Andreoni (1995) finds that the difference relates to the framing of co-operation as providing something good or avoiding something bad. The prospect of doing something 'good' increases co-operation. This could drive investments on shared trees.

However, this might also make the participant believe to have a higher 'stake' in a shared tree than their neighbour due to which they could 'legitimately' also harvest more from it. This logic might be felt more immediate in the game than in the real world and this needs to be accounted for in the analysis.

In contrast, participants who engage in reciprocal voting try to avoid getting sanctioned for breaching the overharvesting rule and, thereby, show a clear disregard of the rules in that moment. In the real world, this aspect might be less obvious than in the experiment, as finance ministers could try to find more objective arguments not to vote for sanctions. As discussed in the literature review (Section 2.3), it seems to be the case that such influence already happens earlier, before a member state is actually put up for voting on a sanction.

5.3.4.5 Rule adherence

Applying the *rule adherence hypothesis* is rather straightforward and assumes a coherence between compliance at both stages' rules. Rule behaviour at one stage is considered an independent variable for rule behaviour at the other stage. Participants who often overharvest do not seem to care about the common-pool resource or about ensuring sufficient apples on shared trees. Therefore, they would probably also not vote for sanctioning other farmers who overharvest. However, as discussed in Section 3.3.1, it could also be the other way around, that overharvesting reflects 'free-riding' with the expectation that other members should comply. In turn, voting correctly is an expression of caring about the common-pool resource. Therefore, participants who vote correctly are probably less likely to overharvest, and vice versa. This is also supported by experimental literature that finds that those who co-operate are also willing to sanction even in cases where it is costly, or there are no material benefits linked (Fehr and Gächter 2000),[185] Altruistic or pro-social punishment, i.e., where a co-operator punishes a breaching participant, emerges in lab experiments even in large anonymous groups (Boyd et al. 2003). This shows that, in social environments, rational behaviour might also depend on other considerations that are relevant to the specific game design, such as social norms, fairness, or other-regarding preference (see

185 Fehr and Fischbacher (2004) find that even unaffected third-party players do punish non-co-operative behaviour, moreover, even in case it is costly for them. This can be interpreted as humans following deep intrinsic norms of justice.

overview in Cooper and Kagel 2015).[186] This shows that the combination 'comply-comply' is a plausible strategy.[187]

As pointed out in the research design chapter (Chapter 4), it is difficult to compare voting in the experiment with the real-world setting, as there are no official voting records in the Council. Therefore, one might overestimate the role of voting for sanctions, as in the real world, finance ministers might be more reluctant to voice outright opposition to another member state's deficit financing.

5.3.5 Summary

As a comparison, the table below relates the hypotheses as set out in Section 3.3 into expectations towards the experiment's empirical results. The core purpose of the experiment is to find out under which conditions and characteristics participants comply with or breach the harvesting rule and the sanctioning mechanism.

Table 5.3-4: Comparison of hypotheses and expectations.

	Real-world	**Experiment**
Economic policy ideology and personal traits		
Economic policy ideology hypothesis (EH)	Governments that prefer economic policies targeted at stimulus as opposed to prevention are more likely to run higher deficits and are more reluctant to sanction breaching governments.	Participants who rather invest in beekeepers and manure than in complete water irrigation systems are more likely to overharvest and to vote incorrectly.
Risk aversion hypothesis (RiH)	Governments that prefer a risk-taking long-term perspective, as opposed to a risk-averse short-term one, are more likely to run higher deficits and are more reluctant to sanction breaching governments.	Participants who rather invest in ladders than in shepherds are more likely to overharvest and to vote incorrectly.
Sustainability hypothesis (SuH)	Governments that prefer a short-term perspective, as opposed to a long-term one, are more likely to run higher deficits and are more reluctant to sanction breaching governments.	Participants who always leave two apples per tree on the ground and never harvest small apples are less likely to overharvest and to vote incorrectly.

186 Mussweiler and Ockenfels (2012) also find that participants strive to co-operate with participants that they perceive similar to themselves.

187 In contrast, this thesis does not expect to see so-called 'anti-social punishment', i.e., where a breaching participant punishes a co-operator. While in group-based experiments such anti-social punishment could impede the evolution of pro-social punishment (Powers et al. 2012), there does not seem to be such risk for this experiment as it has a single-player set-up. For a discussion of reasons see Bryson et al. (2014).

Economic necessity and political capacity		
Business cycle hypothesis (BH)	Governments rather run higher deficits and are more reluctant to sanction breaching governments in bust than in boom times.	Participants are more likely to overharvest and to vote incorrectly in bust rounds (Rounds 4, 7 and 10) than in recovery (Rounds 5, 8 and 11) or boom rounds (Rounds 6 and 9).
Size hypothesis (SH)	Governments of small and large member states are more likely to run higher deficits than those of medium-sized ones. Governments of large member states are more reluctant to sanction breaching governments than those of medium-sized or small member states.	Participants playing as small or large farmers are more likely to overharvest than medium-sized ones. Participants playing as large farmers are more likely to vote incorrectly than medium-sized or small ones.
North-South hypothesis (NSH)	Governments of member states that are geographically located in the 'South' are more likely to run higher deficits and are more reluctant to sanction breaching governments than those from the 'North'.	Participants from member states that are geographically located in the 'South', i.e., Greece and Portugal, are more likely to overharvest and to vote incorrectly than those from the 'North', i.e., France and Germany.
Strategic interactions		
Co-operation hypothesis (CH)	Common-pool resource literature: Those who contribute to the maintenance and improvement of the common-pool resource are also less likely to breach the common governance rules.	Participants who invest in shared trees are less likely to overharvest and to vote incorrectly.
Reciprocal voting hypothesis (RVH)	Governments that engage in alliances to avoid sanctions are more likely to run higher deficits and be more reluctant to sanction breaching governments.	The more deals a participant seals over the course of the game, the more likely they are to overharvest and to vote incorrectly.
Rule adherence		
Rule adherence hypothesis (RuH)	Governments that run higher deficits are less likely to sanction breaching governments and vice versa.	Participants that overharvest are more likely to vote incorrectly, and vice versa.

Source: own description.

Due to the experimental design, there are a couple of further variables that could be extracted from the data, such as the number of happiness points or investments in a bucket. These are, however, not expected to have a theoretically grounded relationship in rule compliance. Rather, they build on practical considerations that arise from the experimental design. Therefore, this study does not formulate hypotheses for them.

5.4 Good Experimental Research

This chapter closes with a discussion on how the proposed experiment satisfies requirements for good experimental research (see McGraw 1998, McDermott 2012, Kittel et al. 2012). The simplification of the experimental design comes at the expense

of external validity, while allowing high internal validity of the game. The following section describes how to ensure that the experiment allows the researcher to measure the variables that are intended to be measured (internal validity, Section 5.4.1.1), discusses the limitations to generalise from the results of the experiment to the real-world case that it seeks to simulate (section 5.4.1.2), and argues that the risk of biased results stemming from social desirability is low (Section 5.4.2).

5.4.1 Validity

5.4.1.1 Internal validity

Internal validity ensures measuring what is intended to be measured (see King 1994: 25), i.e., that options and moves in the game can indeed be interpreted as intended by design, and that participants understand them in a similar way. As described in the previous sections, internal validity is intended to be high as options have a clear impact on the field, for cultivation and for co-operation with other farmers. There are two further aspects that seek to ensure high internal validity: The course of the game has to be identical for all participants to be comparable, and, secondly, the sample shall be unbiased with regard to the treatment and control groups.

Firstly, in order to ensure comparability across participants, the course of the game is the same for all participants. Each round is exactly equal. The only element that changes depending on the participant's actions are the computer-player responses, which however also follow the same strategy (tit-for-tat) and are 'reset' in Rounds 4, 7, and 10 where they are given a fresh start and are once again offered co-operation (except for voting on sanctions). Moreover, in order to 'reset' the game and ensure similar starting positions in Rounds 4, 7 and 10, all cultivation items break during a weather shock.

One important aspect of good experimental research cannot be satisfied due to practicability reasons, which, though, is not expected to risk biasing the results. Usually, in laboratory experiments, participants are not allowed to speak to each other or other people as only their individual and unbiased intuitive playing during the game shall be subject to the analysis. However, as explained in Section 4.3.4.4, it was not possible to conduct the experiment in laboratories. Instead, the experiment was

conducted as an online game that participants could play from wherever and whenever they wanted. Nonetheless, I do not expect this to have a significant biasing impact as would be the case for simpler experiments with few choices or for surveys where participants could be tempted to rush through and randomly select answers. In contrast, the particular setting of this experimental design is expected to attract people who would be willing to participate and might, therefore, wholly engage in it (also see Section 5.4.2 below on social desirability). Similarly, while non-engagement in online survey experiments is a known problem (Harden et al. 2019), this does not seem relevant to this experiment given that participants are free to choose whether or not to participate and there is no explicit or implicit obligation for them to continue playing. They could quit at any time. I assume that they have done so and their answers would be deleted and not bias the analyses.

Secondly, in order to ensure an unbiased sample composition of treatment and control groups, randomisation and collecting information on the sample are key (Druckman et al. 2012b: 17f.). Ideally, the experiment has to be conducted double-blind (Diekmann 2005: 296f.). This means that neither the participant nor the experimenter know in which group a participant is. This thesis' experiment satisfies these conditions for all three treatments.[188] 'Size' is allocated randomly. Participants are recruited anonymously based on their nationality, are unwitting subjects regarding the selection process, and the experimenter is not present during the experiment to possibly intervene and bias results. Thirdly, the business cycle treatment appears within the game in an identical way for all participants. In particular, the shock is not explicitly announced. While the first two treatments are between-subject treatments, where a participant is allocated to one group only, the third one is a within-subject treatment, where the participant experiences all parts of that treatment, i.e., bust and non-bust periods (Druckman et al. 2012b: 18, Niederle 2015: 121f.). In order to avoid usual learning effects for within-subject treatments, the design uses different types of weather events (Druckman et al. 2012b: 18).

188 Moreover, participants are recruited from large samples of students via electronic means, which grants anonymity to the participants. Therefore, participants do not feel an obligation to participate when they are addressed.

Moreover, it is important to understand potential overrepresentation in the sample composition. [189] Therefore, a survey collects relevant information about the participants' characteristics, such as age, gender,[190] field of study, and other potentially relevant information (see discussion in Anderies et al. 2011).[191] The participant is also asked what the goal of the game was for them (see discussion in Anderies et al. 2011).[192] This serves to understand the participant's objective during the game as either focused on having many apples, or ensuring the good condition of the trees, reflecting a 'stimulus' or 'prevention' orientation.[193] The survey also helps us to understand potential overrepresentation due to a self-selection bias, which is a common risk for experimental results (see discussion on social desirability in Section 5.4.2). Even though this study has a clear target group (see Section 4.3.2) and tries to approach them most effectively (see Section 4.3.4.3), there might be a risk of self-selection. Self-selection typically occurs for participants that are interested in the experiment's topic and do have time and resources to do the experiment. Another source of self-selection could lie with the complexity of the game. Some participants might find it too difficult and drop out. One could argue that this resembles self-selection on the grounds of cognitive or intellectual capacities. Even if this was the case, I could not think of any plausible expectations of this correlating with any of the variables of interest. Therefore, while one can assume self-selection to have taken

189 In particular, this information is crucial because this study does not have similar information on the population from which the sample is drawn, or on the population for which the sample shall be representative (see discussion below, in Section 4.3.2 and in Section 6.1). As a result, one could not say to what extent overrepresentation could actually be a misrepresentation.

190 While gender is a usual control variable, it is of particular importance for this study because Niederle (2015) reports that gender could conflict with risk and co-operation attitudes.

191 The game starts with a short survey on age, gender, nationality, field of study, and main occupation to collect the most relevant information. After the game, the participant is thanked for their participation and asked to fill out a further information survey (such as whether they: have lived abroad, live in a capital city, are proficient in the English language, position themselves on the political left-right spectrum, and their self-stated goal of the game, see Appendix A.5-2.2 points 1b, 62, and 63). For the sake of ethical considerations, a participant does not need to answer all questions, and there is also the option "prefer not to say".

192 According to their answer, they get simplified high scores of their achievements during the game in the form of a text congratulating them on their farming achievements. The high scores evaluation is just for fun for the participant. The respective texts can be found in Appendix A.5-2.2, point 64.

193 There is a second survey that seeks to shed light on potential overrepresentation of general attitudes and convictions in the dataset that relate to the game's topics, such as co-operation and economic perspectives (see discussion in Anderies et al. 2011). Being too extensive, this data was not used for the purpose of this thesis. The questions refer to the general level of trust, preference for an either liberal or interventionist government economic policy, attitudes that are relevant for the game (on the level of risk-taking, patience, a long-term or flexible oriented way of planning), whether co-operation is perceived as leading to better outcomes for themselves, whether laws shall be followed, having an option to borrow apples, and the relevance of leisure time.

place, it is not expected to pose a problem to the analysis of the experiment. In order to ensure this, respective control variables will be included in the analyses.

5.4.1.2 External validity

Four aspects are relevant for external validity: firstly, it hinges on the level of abstraction and accuracy that an experimental design uses to replicate the real-world case. Secondly, the role that participants play in the game and the type of participants shall be comparable to the topics and conflicts of decisions that decision-makers in the real-world case face, whose behaviour they seek to simulate. Thirdly, external validity requires that one can properly measure preferences and tangibly interpret them for the real-world case. Finally, ensuring replicability of an experiment is crucial for the generalisability of the results.

Firstly, an experimental design needs to strike a balance between accuracy and abstraction for the particular purpose of the research question, and clearly be aware of limitations of generalisability (see Niederle 2015, Lohmann 2001[194]). McGraw (1998) points to the necessity of conceptually clarifying the purpose of an experiment with regard to its design and capacity to explain a phenomenon. Abstract experiments could well show a specific cause-effect relationship, but they are very limited when it comes to explaining exactly why the link occurs, and under which conditions this holds[195] true. Moreover, the risk of underspecified models is to draw generalisations that in fact are not appropriate (see Bardsley 2010). On the other hand, detailed experiments might be difficult to interpret and to apply to other cases with slightly different conditions. The allegory of this thesis' experiment seeks to strike a balance between accuracy and simplicity. The context adequately mirrors the topic of the study. As explained in detail above, even though the game mirrors private business decisions, a participant reveals economic preferences towards one particular cultivation strategy over another. Moreover, being bound together with other farmers in the common-pool resource

194 Lohmann (2001) points to caution when abstracting from a real-world phenomenon to an abstract model. By trying to capture the essentials, the researcher might miss that subtleties can be essential, too.

195 Experimenters from the economics field argue that simplicity in fact ensures a pure relationship between cause and action, and that a context risks biasing this relationship, while, in psychology they argue that it is exactly a context that provides actual meaning to decisions (Dickson 2012).

approximates the perceived necessity of co-operating with others, as is the case in the SGP context.

Even though the game is built on the neoclassical understanding of the SGP that one energy point can only be spent once, and hence ignores the idea of a fiscal multiplier, it also allows Keynesian perspectives to reveal through specific cultivation items (for instance a 'ladder' represents a bet on high future gains[196]) and harvesting opportunities (three different kinds of apples, leaving two apples on the ground, and harvesting from either private or shared parts). In particular, this satisfies the Keynesian idea of public spending having a multiplier effect. In this experiment, the multiplier comes from investing. Moreover, as there is no goal prescribed, the goal of the game can be freely interpreted. Finally, economic shocks come exogenously in the game (through weather events), and are not triggered by the participant's performance, also reflecting Keynesian perceptions. Only the effects of shocks on the participant's field depend on how the field has been cultivated beforehand.

Secondly, limitations arise from the fact that participants are not governments, but this thesis argues that the conflicts they face are comparable. In the experiment, there is no re-election threat and related intentions to please voters.[197] Instead, the experiment seeks to approximate similar existential concerns in a more implicit way. The participants also face existential costs, such as losing apples in a shock, losing items in a shock, or getting sanctioned. Moreover, deciding between options, and in particular investment items, is restricted by a 'budget constraint' (i.e., not having unlimited energy points). Moreover, the trees can be considered as an economy that needs to be governed by the participant similar to the workings of a government. Likewise, the impact of shocks follows a similar logic to that in the real world. As a result, the farmer is responsible for the prosperity of the field, which can be considered similar to governments. Still, when interpreting the results for the real-world case, one has to keep the limitations in mind that arise from having students and not politicians as participants and from using experimental evidence to inform a real-world case. Section 6.5 will elaborate on this in detail.

196 In particular, a ladder reflects the intention to create high value from comparably small investment.
197 This aspect was not included in the experimental design in order to minimise social desirability concerns arising from a participant's possible reluctance to leadership (see discussion below).

Thirdly, as is typical for experiments, this experiment collects information on behaviour and assumes that it is an expression of beliefs, convictions or principles (see Woon 2012). The assumption is that choosing one option over another shows preferences for a particular option, and that the combination of several options into a strategy would show the conviction that this strategy is 'better' suited than another for the purpose of achieving the participant's goal.[198] In particular, Woon (2012) notes that "subjects must understand the rules of the game but should not be told or given suggestions about how to behave" (p. 59) in order to ensure that participant inferences can be made. This experiment allows different strategies to be equally meaningful and does not prescribe one strategy as economically most 'successful'. Moreover, it puts the purpose of strategic interactions about a common-pool resource into the context of advancing a private good. With this, the experiment studies *principles* that I consider comparable to economic policy ideology, in particular a general orientation of either 'stimulus' or 'prevention' orientation. As discussed in Chapter 3, this thesis assumes that these principles also build the foundations for macro-economic policy strategies.

Finally, King (1994: 26f.) and McGraw (1998) point to the importance of replicability for external validity of an experiment. Replicating a specific experimental design with another sample, at a different point in time, or with another variation in the game, could provide meaningful insights into the condition under which the experiment shows generalisability, i.e., external validity.[199] Based on the survey, this experiment could be conducted with and compared to different samples. It also facilitates replicability by providing an identical course of the game and pre-programmed fixed responses of computer-players. Instead of having a group of human participants that play the game, individual participants play against computer-players. This ensures measuring individual behaviour, which is directly comparable to other participants. Moreover, the survey ensures that relevant information about the participants is collected, on which a proper comparison is based.

198 Woon (2012) recommends i.e., that "subjects must understand the rules of the game but should not be told or given suggestions about how to behave" (p. 59) in order to ensure that inferences can be made.

199 In fact, McGraw (1998) suggests that especially in political science where cases are often very specific, it would be beneficial to engage in sets of experiments and meta-analyses in order to understand a specific logic that is similar to a number of cases.

In conclusion, even though the experimental design is inspired by the SGP context and constructed so as to mirror it most accurately, it cannot explain real-world cases as, for that purpose, it is too abstract. For instance, it is not possible to infer the extensive process of deciding a public budget from individual participants cultivating an apple orchard. Moreover, the experiment neither includes the ECB, the financial market, nor the European Commission, which can be considered main actors in the setting (see Section 2.2.2). Therefore, in order to keep external validity most parsimonious, this study is supposed to yield insights into what general factors drive compliance with an SGP-like rule. In particular, the intention to design an SGP-*like* rule puts the participant in a situation in which they perceive similar conflicts of interest.

5.4.2 Social Desirability

Experiments in general risk inducing participants to consider the social appropriateness or desirability of their choices in the experiment (see also Diekmann 2005: 382ff.). This risks biasing the results. Social desirability could relate to the experimental design, its promotion or its conduct (for the latter, see Diekmann 2005: 518ff.). Moreover, the experimenter risks projecting their perspectives on the topic into the experimental design, for instance, via offering only options that they perceive possible or plausible (see also Diekmann 2005: 40ff.). In order to reduce such biases, I conducted test rounds and discussed the game design with a diverse audience of people (see Section 4.3.4.2). Therefore, I feel confident about having sufficiently addressed potential risks and having made aware of remaining limitations.

To start with, as regards the experiment's design, there are number of elements in this experiment that seek to reduce potential social desirability biases: (1) a vague definition of the goal of the game, (2) a vague definition of the overharvesting rule and the voting rule, (3) no success-based remuneration, (4) to create a realistic yet fictional atmosphere during the game, and (5) not to use deception. Firstly, the participants are not given any specific goal of the game. While this approach is unusual for economics studies, psychologists usually do not prescribe a so-called 'script', i.e., goal (Hertwig and Ortmann 2001). The core of the game is to see what the participant defines as their goal and their 'best' strategy for playing the game. Hence, prescribing a goal would impede the participants to reveal their preferences.

Secondly, in order to avoid learning and conditioning, the overharvesting rule and the voting rule are simple, but also entail discretion. As said, there is a 'flexible' limit to overharvesting by leaving two apples on the tree or not, and through the possibility of harvesting small and windfall apples instead of ripe apples, as those are not part of the sanctioning procedure. Sanctioning by other computer-farmers is lagged so as to avoid effects of conditioning.[200] Moreover, a number of computer-players sufficient to form a blocking minority[201] agree to seal deals in case the participant only slightly harvests more apples than allowed. Similarly for voting, the participant shall freely decide whether to vote for sanctions, or not at all. While being explicit about what the 'basic rules' of the game are, alleviating the strictness of the rules also allows for individual 'divergences' according to what the participant believes is for the best.[202]

Thirdly, participants might behave differently in a fictional setting such as in computer games rather than in the real world because their actions would not have any consequences for their lives. The usual way to tie the participant's 'experimental' to their real life is to pay the participant according to their (successful) moves in the game (Smith and Walker 2008),[203] which is not possible for this experiment. A success-based remuneration, that increases or decreases according to the moves in the game, is neither suitable nor possible for this experiment, as there is no specific single strategy that is more or less 'successful' than another one. Similarly, a lump-sum remuneration might also risk putting participants in a mindset of competitive winning – and potentially undermine the intrinsic motivations to play. One could criticise that this might lead to a self-selection bias by which only specific people would agree to

200 Moreover, the pre-programmed and fixed responses of the other 18 computer-players follow each of the three different types of behaviour for both harvesting and voting. The types simulate laid-back, moderate and aggressive behaviour, and are distributed across the field so as to show the participant a large variety of different combinations. The main purpose is to avoid participants being conditioned or learning, and, therefore, the game does not prescribe any one approach as being the 'best' one.

201 See Appendix A.5-2, Section A.5.2.4.

202 Moreover, Sutter et al. (2010) show that endogenously chosen enforcement mechanisms find higher compliance rates than the same exogenously given mechanisms. Therefore, the experimental design seeks to alleviate the impression that the rule was given exogenously.

203 Economics studies usually pay participants according to their success in the game. In psychological studies that usually do not prescribe a specific goal, it is rather common to pay a lump-sum (Hertwig and Ortmann 2001, Dickson 2012). There are also experimental studies on economic behaviour not using pecuniary incentives (Tsikas 2021). Gintis (2018) also finds it particularly interesting to study co-operation that evolves without a financial incentive as it could better reflect intrinsic motivation.

playing the game. However, a self-selection bias might also occur in the event of a lump-sum being paid, as it might attract people who would only play for the sake of earning the money and not for the sake of engaging truthfully in the game.[204]

One could argue that disregarding what strategy a participant chooses to play, it is most probably a coherent strategy, i.e., a strategy that this participant perceives to be plausible and meaningful for pursuing a specific goal. Moreover, it is likely to assume that a participant would pick a strategy that they are particularly fond of, which might reflect their true preferences. It is unlikely that a participant plays randomly, and does not follow any particular strategy at all. This provides little fun, and such participants might drop out rather than continuing the game to the end. All in all, I do not see a plausible link between this type of self-selection and the results of the game, and in particular with the dependent and main independent variables. Such random strategies are most likely insignificant in the quantitative analysis as they would not affect the variables of interest but rather centre around the null hypotheses.

Fourthly, participants might behave differently in the game when they know that they are playing with computer-farmers instead of actual people. They might be less willing to co-operate, or play less 'socially' when it is a pre-programmed algorithm that is making choices instead of another person. Therefore, the game seeks to create a more realistic atmosphere. For instance, there is a brief 'waiting page' after each stage saying: "Waiting for the other players" accompanied by the sound of biting into an apple. In fact, a positive aspect of having participants play against the computer could be to reduce social desirability risks. Velez et al. (2008) show that, on average, participants in experiments seek to "balance[…] self-interest with a strong preference for conformity" (p. 1). Likewise, experimental research usually finds more co-operation than is expected by the theory (Bru et al. 2003, Anderies et al. 2011).

Fifthly, this experiment does not need to use deception as the apple orchard allegory already provides a meaningful context. Therefore, there are no risks of social desirability stemming from unclear elements of deception. The use of deception is contested in experimental research (see Dickson 2012). Economics studies oppose any

204 Furthermore, the LSE Ethics code recommends not providing financial incentives in order to avoid potential biases. Finally, even though it simply would have been fair to pay participants a lump-sum for spending their time and effort on this experiment, this research project did not have enough financial resources to do that.

misinformation or omission of information because they argue that this risks biasing the effect of treatment and control by an unmeasurable factor. If participants knew that deception would be used, they might form all sorts of assumptions about what element in the experiment is deceptive, and try to understand what the researcher is trying to hide. In contrast, in psychology, deception is used to induce exactly the 'right' mindset that they want their participants to be in. The idea is to frame elements in the game in the most realistic possible so as not to have any unmeasurable intermediate effects. Participants might form assumptions about why the researcher approximates that particular situation instead of giving the impression of being in that specific situation. Therefore, this experiment does not use deception, but instead creates an allegory, a virtual playing field, within which the elements fully make sense. In this way, this experiment does not need any deception. Participants are told that "the game is about economic preferences and co-operation" (see informed consent form, Appendix A.4.1.1[205]). This is correct. However, the instructions omit saying that the experiment is inspired by and will be related to the SGP.[206] This is to avoid biases due to recent experiences during the crisis years.[207]

As regards social desirability biases stemming from promotion, the information that is given to the participants has a potentially critical effect on social desirability. Therefore, the invitation to participate (see Section 4.3.4.3) was targeted to the fun part of playing only. This is supposed to strengthen the intuitive and intrinsic motivation to participate and to put participants in the mood to express themselves freely. Moreover, the risk that people do not report information in the survey correctly seems negligible given that there is no indication in the experiment that it might be 'socially

205 The wording has been approved by the LSE ethics review.

206 One could also consider telling participants this as they might probably recognise the SGP rules in Stage 2, anyway. In fact, one could argue that there might not actually be a worrisome risk to social desirability bias at all. The game is constructed in such a way that many strategies in the game could be perceived as 'socially appropriate'. As long as a participant is not sure about what behaviour is most likely 'socially desired', there might not actually be a social desirability bias at all. In contrast to studies about racism, for example, where it is clear which behaviour is perceived as 'appropriate', economic preferences are subject to discussion and conviction. Likewise, the participant might find it difficult to say what a 'desirable' co-operation on shared trees would be as this could refer to different harvesting, investing, and voting strategies, which could all be considered fair or co-operative. For instance, co-operation might relate to a strict implementation of the rules, or a more flexible interpretation where harvesting more is balanced with also investing more. Nonetheless, telling participants about the SGP link carries a potential source of bias, which could not be controlled for and, therefore, I chose not to mention the topic at all.

207 Hermann and Ozkececi-Taner (2012) show that a participant's prior experience of a real-world situation might bias their actions in the experiment.

desired' to give one specific answer rather than another. Additionally, there is the option 'prefer not to say', which also forestalls false reporting.

Finally, the conduct of the experiment has to be as neutral as possible in order to avoid participants projecting ideas about social desirability. A good way of doing this is not to be present as an experimenter, as he or she might become the participants' reference for social desirability. This experiment was conducted online with invitations sent electronically.

In conclusion, it will not be possible to prevent people from behaving differently in any experiment than they would do in their real lives. The 'social desirability' bias is a common risk in experiments, and cannot be fully excluded here. This has to be kept in mind when interpreting the results.

5.5 Summary

The experiment simulates the SGP in a simplified and abstract game about cultivating an apple orchard together with others. The game is based on a common-pool resource that overlaps with private goods, and allows the participant to reveal styles of cultivation on the one hand and styles of co-operation and reciprocal voting on the other, thereby mirroring an SGP-like setting and sanctioning procedure. When dealing with scarce resources that are shared with others, participants reveal economic preferences, strategic interactions and a specific style of co-operation. Playing over several rounds verifies whether behaviour remains stable across externally induced boom and bust cycles – or varies depending on the business cycle. Participants are randomly assigned to one of three different sizes (small, medium, large) and are university students in the capital cities of four member states (France, Portugal, Germany and Greece).

This design is singular in experiments so far, allowing this thesis to study in particular the role of economic preferences and co-operation for compliance with an SGP-like rule. The 'cultivation game' builds upon the literature, but is different from standard experiments in several regards. Firstly, it does not model a few choices only, but allows more complex strategies to be revealed. Furthermore, the decisions on compliance are given a thematic context. Finally, the game reflects the overlapping of private goods and a common-pool resource.

While internal validity is supposed to be high, external validity is limited due to the level of abstraction from the real-world setting. Despite the fact that the experiment seeks to simulate the many facets of the SGP setting most accurately, the design remains too abstract to *explain* compliance with the SGP rules. Instead, the experiment allows the researcher to study how participants behave towards an SGP-like framework depending on different conditions (business cycle, size) and their different perspectives (cultivation strategies, nationality). For this reason, the experiment is designed to mirror an SGP-*like* rule that puts the participant in a situation in which they perceive similar conflicts of interests, costs and benefits.

As usual for experiments, this one measures preferences that are assumed to be an expression of beliefs. Internal validity, as well as the conduct of the experiment as a single-player, double-blinded randomised game with an identical course of the game seeks to ensure the plausibility of this assumption. Even though this experiment's design offers room for a large variety of options and playing strategies, the risk of social desirability biasing the results seems rather low. It is essential that the design, promotion and conduct of the experiment provide as few triggers for biased moves in the game as possible. The forementioned aspects are supposed to minimise the risk.

6. Experimental evidence

In the following, I present the empirical material, analyses and results. In the first section, I describe how the dataset and the variables are constructed (6.1). Section 6.2 presents the analyses for breaches at Stage 1, i.e., 'overharvesting', and Section 6.3 the analyses for breaches at Stage 2, i.e., 'voting incorrectly'. 'Voting incorrectly' refers to falsely voting for or against a sanction, as opposed to abstaining or voting correctly for a sanction. Section 6.4 summarises the findings.

6.1 Description of the dataset and the variables

The following section is organised as follows. The first section describes the dependent variables and descriptive statistics (6.2.1). Secondly, the dataset is presented and it can be used for the analyses as it can be considered unbiased with regard to the treatments, or such bias can be addressed (6.2.2). The third section presents the independent variables, and concludes that they can be used for the analyses as they do not show significant collinearity (6.2.3). The final section checks whether the game works as intended by design (6.2.4).

6.1.1 Dependent variables and power analysis

The dependent variables are 'overharvesting' and 'voting incorrectly'. 'Overharvesting' is defined as having overharvested on any of the shared trees (1) or not (0) per round.[208] For compliance with the voting rule, this study constructs a variable that takes the value 'comply' if the participant voted correctly for a sanction (yellow or red), or against a sanction (green) on all computer-players. If at least one vote is incorrect, the variable is coded as 'breach'. In case a participant did not vote at all, the variable is coded 'indifferent'. For using it as a dependent variable, 'voting incorrectly' is coded '1' and the other two values are coded '0'.

During the game, there were 82 participants that overharvested at least once on a shared tree. This corresponds to 25.1% of all 327 participants. The majority of 45

208 To recap, 'overharvesting' is defined as having requested more than half of the available ripe apples on a shared tree.

participants overharvested once only, while 37 participants overharvested more than once (11.3%). Participants overharvested in 154 out of 2,616 rounds (6%).

Table 6.1-1: Number of rounds in which participants overharvested.

	Number of rounds							
	0	1	2	3	4	5	6	Total (>0)
Number of participants	245	45	19	8	5	3	2	82
	74,9%	13,8%	5,8%	2,4%	1,5%	0,9%	0,6%	25,1%

Source: own calculation.

As regards the voting rules, 233 out of 327 participants (71.3%) breached at least once from the voting rules (see figure below). Ninety-four participants voted correctly (37 participants), or abstained (57 participants) in all rounds. Incorrect voting happened in 907 out of 2,616 participant-rounds (34.6%, see figure in Appendix A.6-1-1). This shows that voting rule compliance is much less followed than harvesting rule compliance.

Table 6.1-2: Number of rounds in which participants voted incorrectly.

	Number of rounds									
	0	1	2	3	4	5	6	7	8	Total (>0)
Number of participants	94	30	37	38	56	20	18	15	19	233
	28.7%	9.2%	11.3%	11.6%	17.1%	6.1%	5.5%	4.6%	5.8%	71.3%

Source: own calculation. ‘Comply’ and ‘indifferent’ are both considered under ‘0’.

Plotting overharvesting per round shows clearly how smoothly overharvesting follows the business cycle hypothesis (see figures in Appendix A.6-1-2). Voting shows a similar picture (see figures in Appendix A.6-1-4) – except the peaks are not ‘bust’ rounds, but ‘recovery’ rounds. Compliance is highest in boom rounds and similarly, low in bust and recovery rounds, where there is also a declining trend over the course of the game. This could reflect some fatigue effect, however the rather stable numbers for ‘abstain’ indicate that this is not the case. Rather, given the inverse development of ‘breach’, it shows that it might become difficult to adhere to the rules over time.

To define whether the dataset is large enough to analyse both variables for breaches, one can compute the experiment’s power (see Appendix A.6-1.1.2). Power is defined as “the probability of detecting an effect when it exists” (Statistical Consulting Group 2021a). Calculating power with R (DataCamp 2020, Kabacoff

2017, Statistical Consulting Group 2021a) shows that the dataset is large enough for the analysis with regard to both 'overharvesting' and 'voting incorrectly'. It requires 1,055 participant-rounds and this dataset has 2,616. As regards 'voting incorrectly', the analysis requires 1,235 participant-rounds. As a result, the analyses need 155 participants, while the dataset has 327 participants.

6.1.2 Dataset

The dataset[209] includes 327 participants and eight rounds,[210] so that in total there are 2,616 observations. These are not independent as the data is organised in a panel structure, and this study accounts for this by using a mixed model design. As discussed in Section 5.3, this study constructs a number of analytical variables (see Appendix A.6-1 for a detailed description). Some variables enter the analysis as per round variables and some as aggregate variables. This means that for some, it is important in which round they take what value, while some represent general behaviour which only solidifies across all rounds.

The table below shows summary statistics of all variables. The first column depicts the variable as dependent, treatment, hypotheses-related and control variables, according to which this section is organised. The second column gives the name of the variable, the third column gives the values (for character variables), or descriptive statistics (for numerical variables), the fourth column provides the numbers per values (for per round variables, this is the number of rounds across all participants, and for aggregate variables, this is the number of participants), and the final column shows a graphical distribution of values based on column four.

209 I excluded data from participants who stated that they did not understand the English language well (9 participants), and I kept data from those who stated that they understood it in a medium way (21 participants). There is another dataset which includes 97 participants from a number of more countries, but this will not be used here.

210 The first three rounds are not taken into account but considered as introductory rounds where the participants familiarise themselves with the game. This approach has proven to be useful in the pre-tests.

Table 6.1-3: Summary statistics of all variables.

	Variable	Statistics / Values	Frequencies	Graph
Dependent variables	**Overharvesting**	Min : 0	0 : 2462 (94.1%)	IIIIIIIIIIIIIIIII
	[numeric]	Mean : 0.1	1 : 154 (5.9%)	I
		Max : 1		
	VotingIncorrect	Min : 0	0 : 1709 (65.3%)	IIIIIIIIIIII
	[numeric]	Mean : 0.3	1 : 907 (34.7%)	IIIIII
		Max : 1		
	VotingCorrect	Min : 0	0 : 1700 (65.0%)	IIIIIIIIIIII
	[numeric]	Mean : 0.4	1 : 916 (35.0%)	IIIIIII
		Max : 1		
Treatment variables: economic necessity and political importance	**BusinessCycle**	1. boom	654 (25.0%)	IIIII
	[character]	2. bust	981 (37.5%)	IIIIIII
		3. recovery	981 (37.5%)	IIIIIII
	Size	1. large	97 (29.7%)	IIIII
	[character]	2. medium	118 (36.1%)	IIIIIII
		3. small	112 (34.3%)	IIIIII
	Nationality	1. France	68 (20.8%)	IIII
	[character]	2. Germany	108 (33.0%)	IIIIII
		3. Greece	86 (26.3%)	IIIII
		4. Portugal	65 (19.9%)	III
Economic policy ideology	**InvestmentType**	1. a_none	96 (29.4%)	IIIII
	[character]	2. prevention	112 (34.3%)	IIIIII
		3. stimulus	119 (36.4%)	IIIIIII
	InvestmentTypeR	1. a_none	1699 (64.9%)	IIIIIIIIIIII
	[character]	2. prevention	561 (21.4%)	IIII
		3. stimulus	356 (13.6%)	II
	RiskAversion	1. a_none	90 (27.5%)	IIIII
	[character]	2. risk-averse	187 (57.2%)	IIIIIIIIII
		3. risk-taking	50 (15.3%)	III
	Sustainable	1. 0	257 (78.6%)	IIIIIIIIIIIIII
	[character]	02. Jan	70 (21.4%)	IIII

Strategic interactions	**Co-operation** [character]	1. a_exploiting 2. co-operative	166 (50.8%) 161 (49.2%)	IIIIIIIIII IIIIIIIII
	ReciprocalVoting [numeric]	Mean (sd) : 0.1 (0.3) min < med < max: 0 < 0 < 1 IQR (CV) : 0 (2.4)	16 distinct values	: : : : : .
	ReciprocalVotingR [numeric]	Mean (sd) : 0.1 (0.2) min < med < max: 0 < 0 < 1 IQR (CV) : 0 (3.9)	12 distinct values	: : : : : .
Rule adherence	**OverharvestingT** [numeric]	Min : 0 Mean : 0.3 Max : 1	0 : 245 (74.9%) 1 : 82 (25.1%)	IIIIIIIIIIIII IIIII
	VotingT [character]	1. comply 2. breach 3. indifferent	38 (11.6%) 231 (70.6%) 58 (17.7%)	II IIIIIIIIIIIII III
	Voting [character]	1. comply 2. breach 3. indifferent	916 (35.0%) 907 (34.7%) 793 (30.3%)	IIIIIII IIIIII IIIIII
Control variables	**HappinessPoints** [numeric]	Mean (sd) : 3.1 (4.7) min < med < max: 0 < 1.2 < 24.3 IQR (CV) : 3.3 (1.5)	83 distinct values	: : : : : : .
	Bucket [numeric]	Mean (sd) : 0.4 (0.5) min < med < max: 0 < 0.3 < 2.7 IQR (CV) : 0.7 (1.2)	28 distinct values	: : : : : : : : : . .
	InvestmentRate [numeric]	Mean (sd) : 0.4 (0.2) min < med < max: 0 < 0.4 < 0.7 IQR (CV) : 0.2 (0.4)	64 distinct values	: : : : : : : . . : : : : : : : : : : .

International	1. 0		159 (48.6%)	IIIIIIIII
[character]		02. Jan	168 (51.4%)	IIIIIIIIII
Economics	1. 0		78 (23.9%)	IIII
[character]	2. Economics		153 (46.8%)	IIIIIIIII
	3. SocSci		96 (29.4%)	IIIII
PolSpec	Mean (sd) : 1.6 (1.2)		0 : 63 (19.3%)	III
[numeric]	min < med < max:		1 : 108 (33.0%)	IIIIII
	0 < 1 < 4		2 : 69 (21.1%)	IIII
	IQR (CV) : 2 (0.7)		3 : 59 (18.0%)	III
			4 : 28 (8.6%)	I
Trees	1. 0		138 (42.2%)	IIIIIIII
[character]	2. Apples		56 (17.1%)	III
	3. Trees		133 (40.7%)	IIIIIIII
Male	Mean (sd) : 0.4 (0.5)		0.00 : 198 (60.6%)	IIIIIIIIIII
[numeric]	min < med < max:		0.50 : 16 (4.9%)	
	0 < 0 < 1		1.00 : 113 (34.6%)	IIIIII
	IQR (CV) : 1 (1.3)			
Age	Mean (sd) : 24.8 (5.8)		28 distinct values	:
[numeric]	min < med < max:			:
	18 < 23 < 56			: .
	IQR (CV) : 6 (0.2)			: : :
				: : : . .
Capital	1. 0		42 (12.8%)	II
[character]		02. Jan	285 (87.2%)	IIIIIIIIIIIIIII

Source: own calculation. The variables refer to their commonly used level as either per round (‘R’) or the total at aggregate level (‘T’). The other variable receives the respective suffix 'R' or 'T'. The variable ‘size’ is not ordinal but categorical, as one cannot assume that small, medium-sized, and large farmers affect compliance in a linear way. The variable ‘business cycle’ is also categorical for the same reason.

The dataset is balanced with regard to the treatment variable size (see Appendix A.6-1.2.1). As argued in Chapter 5, due to the game design, there are fewer rounds for 'boom' than for 'bust' and 'recovery'. The number of participants per nationality are similar for Portugal (65 participants) and France (68), while Greece (86) and Germany (108) are higher. Given the different numbers of participants per category, besides the conceptual perspective, it also seems reasonable from a statistical perspective to include the two treatment variables in the analyses to control for the bias in the dataset.

All in all, participation reflects well the desired sample criteria for nationality, field of study, and degree (see A.6-1.2.1). In order to account for characteristics of the participants, the analyses include the usual control variables for age, gender, field of study, and the self-positioning on the left-right political spectrum (see Section 5.4.1.1). Age ranges from 18 to 60 years, while the vast majority are in their early 20s. The female participants outnumber the males by almost two to one. [211] As regards the field of study or profession, as intended, most of the participants are from social sciences and economics. However, the dispersion is quite different across nationalities. The vast majority of participants are university students. However, this effect is already captured through the 'age' variable and therefore, the analyses do not include a variable for the main occupation. Many participants considered themselves to be rather in the middle or on the right side of the classic political spectrum. Additionally, I include control variables that may have an effect on a person's tendency to co-operate given personal experiences. 'International' is a dummy variable and describes participants that have lived abroad, or are currently not living in their country of origin. 'Capital' is a dummy variable, and describes participants who currently live in or around the capital. 'Goal' describes the self-stated goal and takes the values 'apples', 'trees' or 'other'.

6.1.3 Independent variables

This section presents the independent variables. Appendix A.6-1.3 provides details on how the variables are built and checked for plausibility. The variables do not significantly correlate with the other variables so that they can be used in the analyses.

211 There are only 16 participants that did not want to state their gender. I coded them as 0.5.

The explanatory variable of main interest is the 'investment type', which reflects the 'economic policy ideology' hypothesis (see Appendix A.6-1.3.1). The variable is calculated as a comparison: those who invest more energy points over the course of the game in beekeepers and manure than in water irrigation parts are defined as 'stimulus' type and vice versa.[212] As a third category, there are those participants that do not invest in any of the options, or similarly in both ('no type'). 'Risk aversion' takes the value 'risk-averse' or 'risk taking' (see Appendix A.6-1.3.2). It is constructed in such a way that purchases of shepherds, supposed to show risk aversion, are compared to purchases of ladders, supposed to show risk-taking behaviour.[213] 'Sustainable harvesting' is a dummy variable (see Appendix A.6-1.3.3) and is defined as always leaving two apples on the ground on each owned tree, and never harvesting a small apple from both owned and shared trees. The variable 'co-operation' takes either the value 'exploiting' or 'co-operative', reflecting cultivation behaviour on shared trees (see Appendix A.6-1.3.4). The variable 'reciprocal voting' shows the number of deals complied with as a share of all sealed deals per round (see Appendix A.6-1.3.5).[214] The variable is coded as a range between 0 and 1, where 1 indicates that a participant complied with all deals per round.

For all these variables, the analyses use information on the aggregate level because the main interest is in principled behaviour that manifests over the course of the game. For 'sustainable', for instance, it is not particularly interesting whether a participant played sustainably on their own trees in a specific round, but whether they did so in general.[215] This can be understood as principled behaviour, which forms conviction and beliefs. Moreover, there are some conceptual reasons for using the aggregate variable.[216]

---

212 All investment items are summed up across all trees, items and rounds. The variable only considers water irrigation system items when the participant managed to build at least once a fully-functional system (i.e., when they purchased part 3). Due to a programming error, the beekeeper only appears in Round 6 instead of Round 5. One cannot know how this would have changed investments for the participant. However, one can assume this to have a minor effect given that a participant who favours 'stimulus'-oriented investments could have made such investments nonetheless by purchasing 'manure'.

213 Again, the variable simply compares the total number of investments in these items.

214 This means that based on a deal, a participant voted green for another farmer.

215 We checked, and the per round variable also shows a negative effect on overharvesting. If it had shown a positive effect, it would have indicated that participants who harvest sustainably on their own trees need to harvest more on shared trees so as not to 'hurt' their own trees.

216 For 'risk aversion' for instance, one needs to do this because the respective items only became available later in the game and, therefore, the per round coefficient would be distorted by the early

The table below shows the correlation of all variables. The correlation between some variables is statistically significant. However, for the main variables of interest, the table does not show significant collinearity. Therefore, only those variables that do not show significant collinearity are used in the final models.

rounds' '0'. The analyses include 'co-operation' as aggregate variable because the per round variable's reference value ('exploiting') correlates with overharvesting, given the construction of this typology.

Table 6.1-4: Correlations between variables.

	OverharvestingT	Overharvesting	VotingT	Voting	Size	Nationality	BusinessCycle	InvestmentType	RiskAversion	Sustainable	Co-operation	ReciprocalVoting	ReciprocalVotingR	International	Economics	PolSpec	Trees	Male	Age
OverharvestingT																			
Overharvesting	0.43 ***																		
VotingT	0.03	0.03																	
Voting	0	0.02	0.46 ***																
Size	0.05	0.03	0	0.04															
Nationality	0.02	0.01	0	0.09 ***	-0.04														
BusinessCycle	0	0.02	0	0.13 ***	0	0													
InvestmentType	0.05	-0.01	-0.13 ***	-0.12 ***	-0.09 ***	-0.05	0												
RiskAversion	-0.08 *	0.01	0.02	-0.04	-0.2 ***	0.01	0	0.1 ***											
Sustainable	-0.11 ***	-0.06	0	-0.06	-0.01	-0.19 ***	0	0.08 **	0.06										
Co-operation	0.11 ***	0.04	-0.08 **	-0.07*	-0.16 ***	-0.06	0	0.25 ***	0.07 *	0.1 ***									
ReciprocalVoting	0.22 ***	0.11 ***	-0.04	-0.04	0.14 ***	-0.04	0	0.15 ***	0.01	0.05	0.04								
ReciprocalVotingR	0.16 ***	0.13 ***	-0.03	0	0.08 *	-0.04	0.01	0.1 ***	0.02	0.01	0.06	0.62 ***							
International	-0.03	0	-0.07 *	-0.14 ***	-0.05	-0.29 ***	0	0.04	0.02	0.12 ***	0.02	0.14 ***	0.09 **						
Economics	-0.03	-0.04	-0.06	-0.04	0.07 *	-0.1 ***	0	0.09** *	0.06	-0.04	0.02	0	0.01	0.13 ***					
PolSpec	0.03	0.02	0.01	0.03	-0.01	0.23 ***	0	-0.05	0.03	-0.13 ***	-0.05	-0.03	-0.02	-0.14 ***	-0.13 ***				
Trees	0.02	-0.02	-0.15 ***	-0.15 ***	-0.02	-0.07 *	0	0.2 ***	0.04	0.17 ***	0.1 ***	0.04	-0.02	0.02	0.05	0.01			
Male	0.04	0.03	-0.05	-0.03	-0.02	0.09 **	0	0.03	-0.08 *	-0.05	-0.06	-0.02	-0.05	-0.06	-0.04	0.06	-0.04		
Age	-0.09 **	-0.06	-0.02	-0.04	0.04	-0.08 *	0	-0.07 *	-0.04	0.1 ***	-0.02	-0.11 ***	-0.1 ***	0.17 ***	-0.09 **	0.02	-0.01	0.05	
Capital	-0.07 *	-0.07 *	0.02	0.01	-0.05	0.07 *	0	0.01	0.03	-0.02	0.12 ***	0	0	-0.12 ***	0.05	-0.07	-0.06	-0.03	-0.23 ***

Source: own calculation, significance levels: * p<0.1; ** p<0.05; *** p<0.01. Dark blue marks values above 0.3 and below -0.3, blue marks values between (-)0.3 and (-1)0.15, light blue marks values below/above (-)0.15, and white marks all values without statistical significance.

6.1.4 Does the game work as intended by design?

Figure 6-1-5 shows the number of energy points that a participant has at the beginning of a round across all eleven rounds for all participants, coloured according to their size. Each line represents one participant, where yellow shows large, orange shows medium-sized and black shows small farmers. This graph looks as expected, and indicates that the game works as intended. All three boom and bust cycles are reflected well in the number of energy points.[217] Almost all lines move up and down at the same time. Moreover, the three boom and bust cycles have similar amplitudes, showing that they can be used as experiment-internal double checks for behaviour within one business cycle. This also shows that there is no systematic 'response fatigue' towards the end of the game, but instead that the participants' engagement remains stable across all rounds.

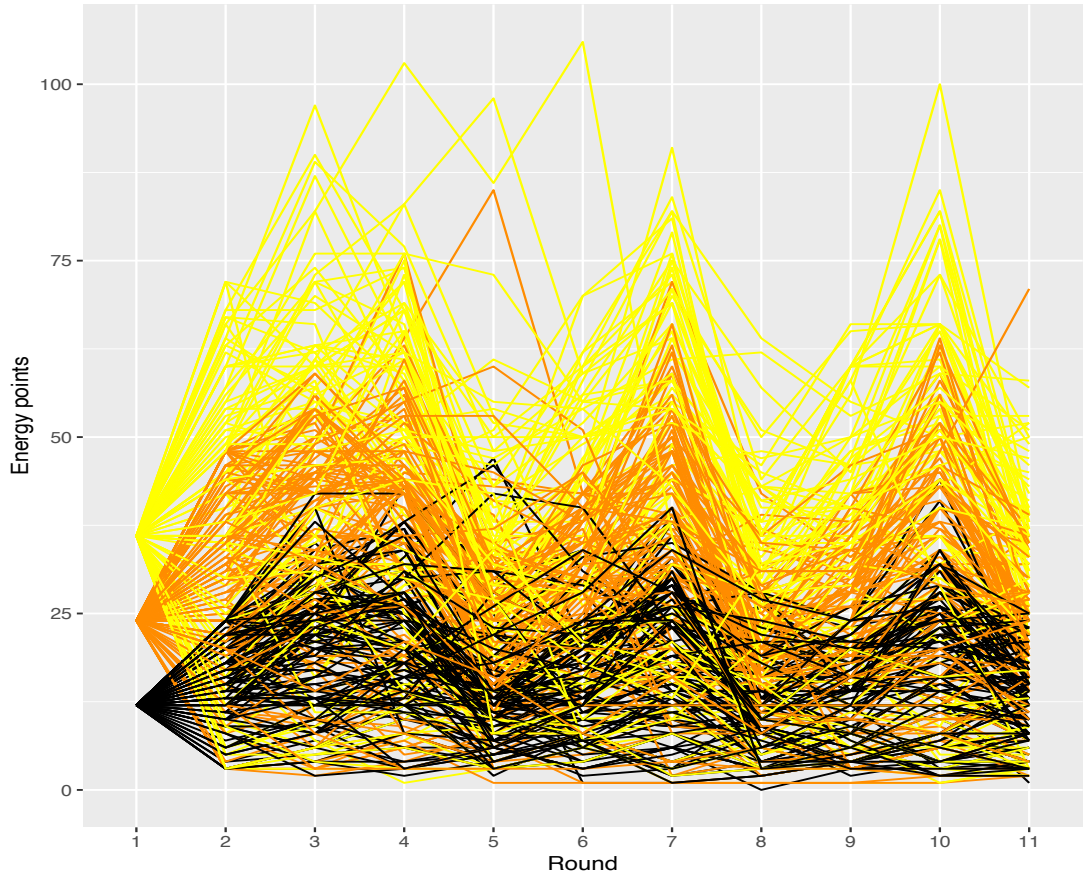
Furthermore, the three groups of sizes are roughly 20 energy points apart from the next higher or lower one. This shows that while growth seems limited as intended by design,[218] shrinkage can happen to participants of all sizes of farmers. Small farmers have difficulties becoming 'rich' and reaching levels of energy points similar to the medium-sized farmers, and medium-sized farmers experience the same limits in comparison to large farmers. On the other hand, low levels of energy points can equally be found for small, medium-sized and large farmers.

In conclusion, developments are not determined by a participant's size, but are up to a participant's behaviour. Similarly, the number of available ripe apples at the beginning of a round reflects the business cycle well, and shows that the game works (see Appendix A.6-1.4).

217 This is reflected with a lag of one round as intended by design. The participant disposes of energy points in the round after they collect the relevant apples.

218 Levels of energy points shall not grow in an unlimited fashion as this would result in making conflicts between investment decisions obsolete. Such conflicts are intended so as to study the participant's decisions for one option and against another.

Figure 6.1-5: Number of energy points at the beginning of a round – per participant.



Source: own illustration. Yellow lines show large, orange lines medium-sized, and black lines small participants. The graph includes all eleven rounds for a better illustration.

6.2 Compliance at Stage 1: analysis of overharvesting

This section uses regression analysis to better understand overharvesting. The next section will focus on Stage 2 compliance and voting incorrectly. The section is organised as follows: it first explains the methodological background (6.2.1), then it presents the analysis and discusses the empirical results (6.2.2). The summary compares the results to the hypotheses (6.2.3).

6.2.1 Statistical prerequisites

Given that the dependent variable 'overharvesting' is dichotomous, one can use logistic regression models. The data is structured as a panel with the participant's ID number as the so-called 'group level' (I follow Winter 2014 and 2016).[219] This means

219 For further details on how to perform the analysis and implement it in R, see Lohmann (2007), Krause and Urban (2013), and Levy (2017).

that the analyses consider each individual as having characteristics which are not able to be fully grasp with the variables that are measured in the experiment. One accounts for such characteristics by assigning each individual a separate intercept and considering each round as a separate observation within the 'group' of each individual. This is called a mixed model[220] with both fixed and random effects, i.e., the usual independent variables and, additionally, a random intercept.[221]

This carries the assumption that each observation per individuum is independent from the other. It is obvious that rounds are not fully independent from each other given the sequential nature of the game. However, one can assume that overharvesting does not depend on the previous or the next round, but rather on the structure of specific rounds following weather events. This is accounted for in the analysis with the 'business cycle' variable. The only thing that links rounds is when a participant harvests small apples in one round that cannot grow ripe in the next, and thereby might provoke overharvesting as the only option to harvest the desired number of ripe apples. The analyses measure this with the variable 'sustainable'.

I compute a logistic regressions model using the statistical software R.[222] In the following tables, the regression coefficients and the standard errors are presented. One cannot easily interpret the bare coefficients because they are on the logit scale. In order to show more meaningful results, the study provides predicted probabilities using graphs (following King et al. 2000). This indicates the predicted probability that the dependent variable is 1 given each value of the independent variable, while keeping all other independent variables at their mean (for numeric variables) or reference value (for categorical variables, see appendix A.6-2.1.3 for a table).

220 Mixed models or mixed-effects models are also referred to as multilevel or hierarchical models.

221 In social sciences, there is no homogenous definition of the concept of fixed and random effects as regards conceptual considerations of model choice. Therefore, this thesis refers to the term 'random intercept' to be most precise, and to refer to methodology and not only concept.

222 Statistics were done using R version 4.1.1 (R Core Team 2021), RStudio (RStudio Team 2022), a number of packages (Alboukadel 2019, Arnold 2021, Bates et al. 2015, Bolker and Su 2011, Chang 2019, Comtois 2021, Dowl and Srinivasan 2021, Fox 2003, Fox and Weisberg 2019, Hartig 2021, Hlavac 2018, Lander 2021, Long 2021, Lüdecke 2018, Lüdecke et al. 2021, Makowski 2018a, Nieuwenhuis et al. 2021, Sarkar 2008, Sievert 2020, Wickham 2016 and 2021, Wickham et al. 2019, Wickham et al. 2021, Wilke 2021, Zeileis and Hothorn 2002), and based on inspiration from examples of coding scripts (Bartlein 2020, Bolker 2012, DWR447 *no year*, Holtz 2018, Kabacoff 2020, Lüdecke 2021, Makowski 2018b, Oller 2017, StackExchange 2018 and 2021, Wickham et al. *no year*). The model formula is inspired by Statistical Consulting Group (2021b).

Ideally and conceptually, it seems interesting to include a four-way interaction between 'business cycle', 'size', 'nationality', and 'investment type'.[223] Descriptive statistics indicate that some combinations of these four variables have more cases of overharvesting than others (see Appendix A.6-1.3). The table below shows a 4-way combination for cases of overharvesting. For instance, among all the Portuguese participants, there are eight cases for medium-sized 'prevention'-oriented participants that overharvest in bust rounds. In absolute terms, most cases happen in bust rounds, for small participants and for Greek participants of 'stimulus' type. However, the number of observations is too small as the table below shows. Therefore, the analyses take three routes: 1) full model with interaction between 'size' and 'nationality' (Model 6); 2) full model with additional interaction between 'business cycle' and 'investment type' (Models 61 and 62); 3) regression model using 'overharvesting' at aggregate level, without taking the business cycle into account (Model 7).

Table 6.2-1: 4-way combination – number of cases of overharvesting.

		Boom			Bust			Recovery			Total
		small	medium	large	small	medium	large	small	medium	large	
France	a_none	0	1	0	0	2	2	1	2	0	8
	Stimulus	0	0	1	4	1	2	1	0	2	11
	Prevention	0	0	0	2	2	0	1	0	0	5
Germany	a_none	2	0	1	8	0	1	6	0	0	18
	Stimulus	0	1	0	5	6	1	2	1	1	17
	Prevention	1	1	3	5	3	0	2	2	6	**23**
Greece	a_none	1	0	2	0	1	2	0	0	2	8
	Stimulus	1	0	2	7	5	6	8	2	4	**35**
	Prevention	0	0	0	0	0	0	0	1	1	2
Portugal	a_none	2	0	0	2	1	0	1	0	0	6
	Stimulus	0	1	0	1	3	3	1	0	1	10
	Prevention	0	0	0	1	8	0	1	1	0	11
Total		7	4	9	**35**	**32**	17	**24**	9	17	154

Source: own description, yellow indicates combinations with at least five cases. Recall that cases are not evenly spread across all values of a variable. For instance, there are more German than Portuguese participants. Therefore, one cannot simply compare the absolute numbers.

223 I could also calculate predicted probabilities for specific combinations of variables. This would add up the probabilities. However, this is not equivalent to analysing interactions. For this, the model needs to estimate the joint change of variables and not the individual ones.

6.2.2 Analyses of results

In the following, the results are discussed alongside the hypotheses. The section starts with the treatment variables, followed by economic policy ideology, strategic interaction, rule adherence, and finally the control variables to complete the picture.

The table below shows the regression models. The appendix shows the sequential building of the models as suggested by Hox (2010, cited from Krause and Urban 2013) by adding explanatory variables one after the other (A.6-2.1.1), and discusses the goodness-of-fit of the models (A.6-2.1.2). Model 5 represents the full model with all variables and without any interaction. It serves as a comparison to the following models that study the abovementioned options on 'size', 'nationality', 'business cycle' and 'investment type'. Model 6 follows option 1 named above, and includes the interaction of 'nationality' and 'size'. Model 61 represents option 2.[224] Model 62 uses the per round variable for the investment type to reflect what items participants purchase in bust rounds. However, the variable does not show a significant effect. One can conclude that it probably does not matter exactly what a participant purchases in a specific round, but rather their behaviour that manifests throughout the game. Section 6.2.2.2.2 discusses Option 3, and builds models using overharvesting at aggregate level as dependent variable. To discuss all other variables' effects, Model 61 is used because conceptually, it fits best what the analyses seek to explore, and it also has an appropriate goodness-of-fit (see Appendix A.6-2.1.2).

224 The intercept is not anymore statistically significant. Looking at the size of the coefficient and the standard error, one can see, however, that the intercept still has a significant magnitude. The standard error is very close to the one in Model 6, and the coefficient is marginally smaller but still near the one in Model 6. This is probably due to the interaction added as it includes 'BusinessCycle', which has a large impact on overharvesting, and the absolute number of cases of overharvesting within combinations are reduced.

Table 6.2-2: Regression analyses for overharvesting per round.

	Dependent variable: Overharvesting perR			
	(5) (full)	(6) (option 1)	(61) (option 2)	(62) (option2R)
Sizemedium	0.0981	0.0663	0.0213	0.1226
	(0.4026)	(0.9543)	(0.9475)	(0.9381)
Sizesmall	0.1383	0.0826	0.0723	0.2728
	(0.4160)	(0.9412)	(0.9331)	(0.9228)
NationalityGermany	0.7054	0.5661	0.4928	0.5292
	(0.4605)	(0.9001)	(0.8940)	(0.8881)
NationalityGreece	0.2973	0.7793	0.7145	0.8716
	(0.4845)	(0.8754)	(0.8695)	(0.8599)
NationalityPortugal	0.3655	-0.5611	-0.4969	-0.3074
	(0.5095)	(1.0519)	(1.0356)	(1.0193)
BusinessCyclebust	1.3785***	1.3728***	0.5050	1.4813***
	(0.2856)	(0.2851)	(0.4894)	(0.3477)
BusinessCyclerecovery	0.6471**	0.6420**	-0.1528	0.4440
	(0.2976)	(0.2972)	(0.5223)	(0.3569)
InvestmentTypeprevention	-0.1777	-0.1429	-0.8027	
	(0.4491)	(0.4457)	(0.7207)	
InvestmentTypestimulus	0.7193*	0.7991*	-0.5063	
	(0.4227)	(0.4211)	(0.6964)	
InvestmentTypeRprevention				-0.6771
				(0.8149)
InvestmentTypeRstimulus				0.1862
				(0.6702)
RiskAversionrisk-averse	-0.4752	-0.4574	-0.4589	-0.4614
	(0.3628)	(0.3611)	(0.3601)	(0.3571)
RiskAversionrisk-taking	-0.1185	-0.0585	-0.0464	0.1489
	(0.4802)	(0.4866)	(0.4845)	(0.4778)
Sustainable1	-1.1572**	-1.0936**	-1.0778**	-1.0105**
	(0.4551)	(0.4559)	(0.4534)	(0.4472)
Cooperationco-operative	0.5439	0.5859*	0.5741*	0.7408**
	(0.3433)	(0.3407)	(0.3392)	(0.3244)
ReciprocalVoting	1.4703***	1.4010***	1.3703***	1.2081***
	(0.4491)	(0.4466)	(0.4446)	(0.4332)
VotingT_rcomply	-0.9578	-1.1087*	-1.0499*	-0.9368
	(0.6292)	(0.6329)	(0.6288)	(0.6198)
VotingT_rbreach	-0.2508	-0.3851	-0.3554	-0.2779
	(0.4324)	(0.4319)	(0.4310)	(0.4256)
Age	-0.0622**	-0.0564*	-0.0593*	-0.0635**
	(0.0305)	(0.0304)	(0.0305)	(0.0308)
Capital1	-1.1765***	-1.2784***	-1.2911***	-1.2316***
	(0.4452)	(0.4441)	(0.4422)	(0.4395)
Sizemedium:NationalityGermany		-0.1850	-0.0901	-0.1701
		(1.1696)	(1.1635)	(1.1563)
Sizesmall:NationalityGermany		0.4264	0.4465	0.1992
		(1.1411)	(1.1339)	(1.1224)
Sizemedium:NationalityGreece		-0.8597	-0.7974	-0.8223
		(1.1884)	(1.1829)	(1.1736)

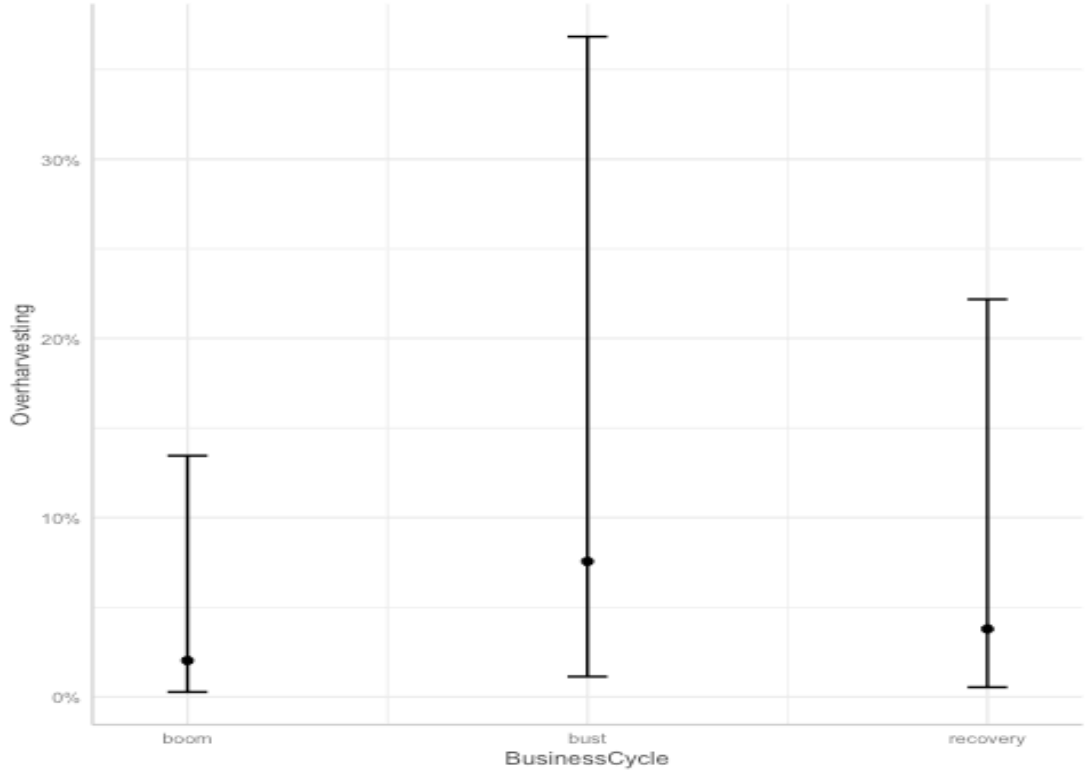
Sizesmall:NationalityGreece		-0.7385	-0.6555	-0.8179
		(1.1756)	(1.1670)	(1.1558)
Sizemedium:NationalityPortugal		1.5557	1.4569	1.1528
		(1.3183)	(1.3037)	(1.2877)
Sizesmall:NationalityPortugal		0.7452	0.5767	0.2060
		(1.3687)	(1.3567)	(1.3374)
BusinessCyclebust:InvestmentTypeprevention			0.7145	
			(0.7256)	
BusinessCyclerecovery:InvestmentTypeprevention			0.9215	
			(0.7615)	
BusinessCyclebust:InvestmentTypestimulus			1.6046**	
			(0.6914)	
BusinessCyclerecovery:InvestmentTypestimulus			1.3215*	
			(0.7281)	
BusinessCyclebust:InvestmentTypeRprevention				0.1033
				(0.8748)
BusinessCyclerecovery:InvestmentTypeRprevention				1.0670
				(0.8957)
BusinessCyclebust:InvestmentTypeRstimulus				-0.6247
				(0.7614)
BusinessCyclerecovery:InvestmentTypeRstimulus				0.0935
				(0.8580)
Constant	-2.7428**	-2.6780**	-1.8584	-2.3826*
	(1.1242)	(1.2711)	(1.2922)	(1.2642)
Observations	2,616	2,616	2,616	2,616
Log Likelihood	-481.9314	-478.8342	-475.6958	-479.9051
Akaike Inf. Crit.	1,003.8630	1,009.6680	1,011.3920	1,019.8100
Bayesian Inf. Crit.	1,121.2510	1,162.2730	1,187.4740	1,195.8920
Note:			*p<0.1; **p<0.05; ***p<0.01	

Source: own calculation.

6.2.2.1 Economic necessity and political importance

The hypotheses state that participants are more likely to overharvest in bust rounds ('business cycle hypothesis'), if they are assigned to small or large farmers ('size hypothesis') and come from Greece or Portugal ('North-South hypothesis'). In all models, neither 'size' nor 'nationality' have a statistically significant effect. Instead, 'bust' has the expected positive effect on overharvesting, and is statistically significant in all models. Moreover, in all but Model 62, also 'recovery' has a positive effect on 'overharvesting'.

Figure 6.2-3: Predicted probabilities depending on 'business cycle'.



Source: own calculation based on Model 6. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The solid lines indicate the 95% confidence interval.

The figure above illustrates the variable's effects by computing predicted probabilities that overharvesting equals 1 for different levels of the respective independent variable.[225] Despite the fact that the 95%-confidence bounds overlap, even though the variable is statistically significant, the different dots indicate that the predicted

225 The R code is inspired by Lüdecke (2021).

probability for overharvesting in bust rounds is about six percentage points higher than in boom rounds. All else equal, i.e., all other variables kept at their reference value, the likelihood that a participant overharvests in 'bust' rounds is 8%.

6.2.2.2 Economic policy ideology

6.2.2.2.1 Investment typology and the role of the 'business cycle'

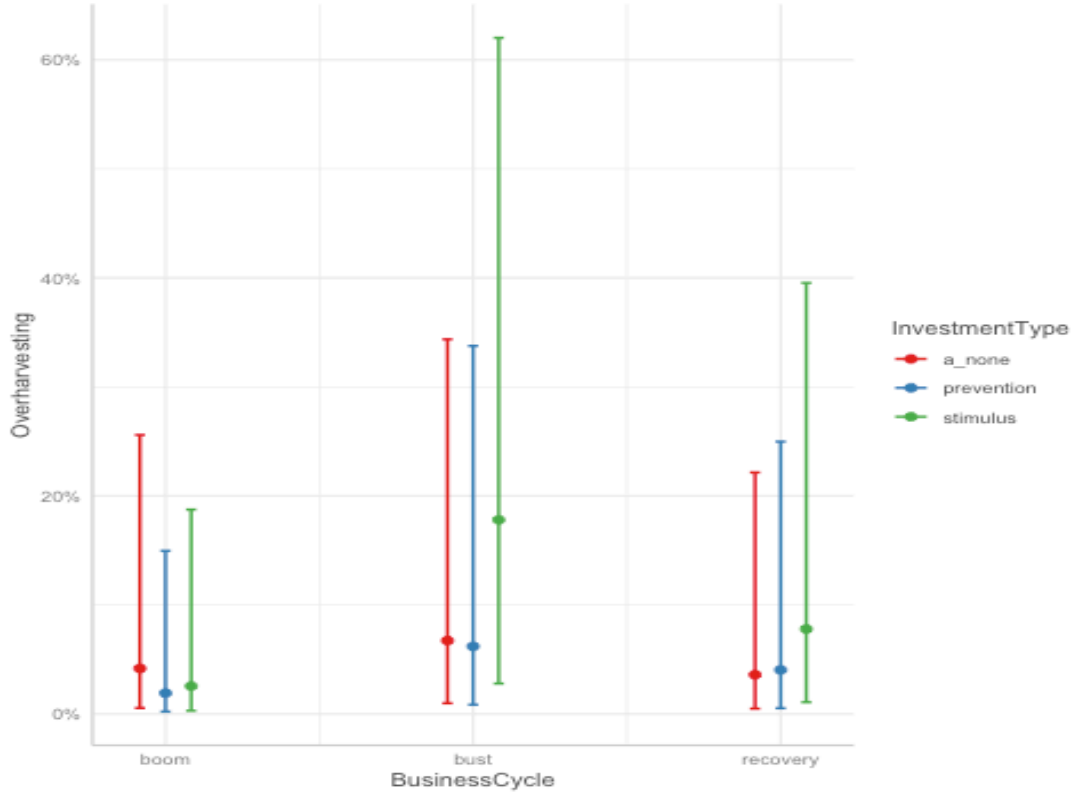
The 'economic policy ideology' hypothesis states that participants who rather invest in beekeepers and manure, supposed to reflect 'stimulus'-oriented cultivation, than in complete water irrigation systems, supposed to reflect 'prevention'-oriented cultivation, are more likely to overharvest. In all models, the 'stimulus'-oriented investment type shows the expected effect and is statistically significant. In particular, Model 61 neatly shows that the previously explanatory power of 'business cycle' is driven by the 'stimulus' investment type. The interaction between both is significant while the effect of 'business cycle' alone is not significant anymore. This holds true for both 'bust' and 'recovery' rounds. Instead, Model 62 does not show any significant effects of the 'stimulus' type. Instead, there is an indication that 'prevention'-oriented participants might be more likely to overharvest in 'recovery' rounds. The effect has a significant magnitude, but is not statistically significant. In all other models, the 'prevention'-oriented type is not significant, meaning that such participants are similarly likely to overharvest than 'no type'-participants, which serve as a baseline category for this variable.

The figure below reflects the predicted probabilities for the interaction between the business cycle and the investment type.[226] It shows the large impact of 'stimulus' in 'bust' rounds. While the previous graph showed that 'bust' alone has a probability of 'only' 8% to result in overharvesting, together with 'stimulus', the probability is 18%. In contrast, the probability for 'prevention' and 'no type' ranges between 6% and 7%. The impact of 'stimulus' is still prevalent in 'recovery' rounds, though to a lesser extent, while the probability to overharvest in boom rounds ranges between 1% and 4%, and that there is little difference between the investment types. The magnitude of the effect of the 'stimulus' type is large compared to the magnitude of the other

[226] It is important to keep in mind that predicted probabilities use French participants as baseline for 'nationality' and large participants for 'size'.

variables. Only ‘reciprocal voting’ comes close to this magnitude, as will be shown below.

Figure 6.2-4: Predicted probabilities depending on ‘investment type’ and ‘business cycle’.



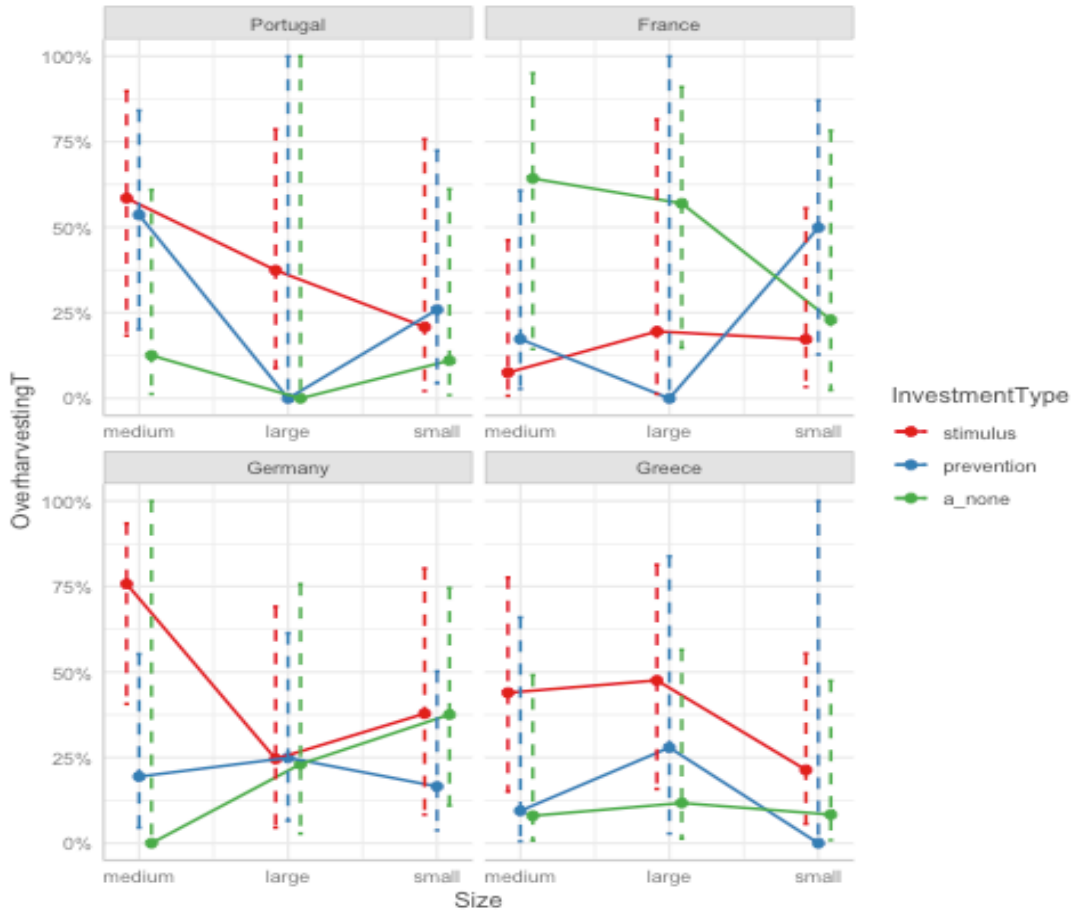
Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variables on the x-axis and marked in colours. The solid lines indicate the 95% confidence interval. ‘a_none’ represents those rounds in which participants have not invested at all or similarly in both ‘prevention’ and ‘stimulus’ items.

6.2.2.2.2 Role of ‘size’ and ‘nationality’

Following option 3 named above, the study runs a series of models that have as dependent variable whether a participant ever overharvested during the game (see Appendix A.6-2.1.4). Since the business cycle hypothesis is supported, one can check the likelihood to overharvest with the aggregate models using the same independent variables as in Models 5 to 62. First of all, it is noteworthy that all other independent variables have similar effects to those in Model 61. This shows that one can indeed compare this model to the ones above. Model 73 includes the three-way interaction, and there is one statistically significant combination: a positive effect for ‘stimulus’-type, large Greek participants. There is also a slightly significant magnitude

(comparing the coefficient with the standard error) for 'no type' small German participants, but it is not statistically significant. However, these results have to be considered with caution. A three-way interaction produces a large number of combinations, i.e., dummy variables, which reduce degrees of freedom in the model. Therefore, it might happen that a dummy becomes significant if it exactly matches a few cases.

Figure 6.2-5: Predicted probabilities depending on 'size', 'nationality' and 'investment type' – aggregate model.



Source: own calculation based on Model 73. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variables on the x-axis and marked in colours. The dashed lines indicate the 95% confidence interval. 'a_none' represents those participants that have not invested at all or similarly in both 'prevention' and 'stimulus' items. As one can see with the dots at zero and confidence bounds spreading across the entire range, there are no cases for the following combinations: large Portuguese 'prevention' and 'no type', large French 'prevention', medium-sized German 'no type' and small Greek 'prevention' type.

Therefore, the figure above shall be understood only as indicative to illustrate that the 'investment type' might play out differently across 'nationality' and 'size'. For ease of reading, the figure includes connecting lines between the dots according to the

investment type. They do not have any statistical meaning. For all countries, one can see that the 'stimulus' type (red line) ranks highest or among the highest probabilities, except for French participants, where the 'no type' ranks highest. The highest probability to overharvest is for 'stimulus'-type medium-sized German participants, at around 75%. This means that three out of four such participants overharvest at least once during the game. This is significant. At the other end, 'stimulus'-type medium-sized French and 'no type' Greeks of all sizes range below 10%.

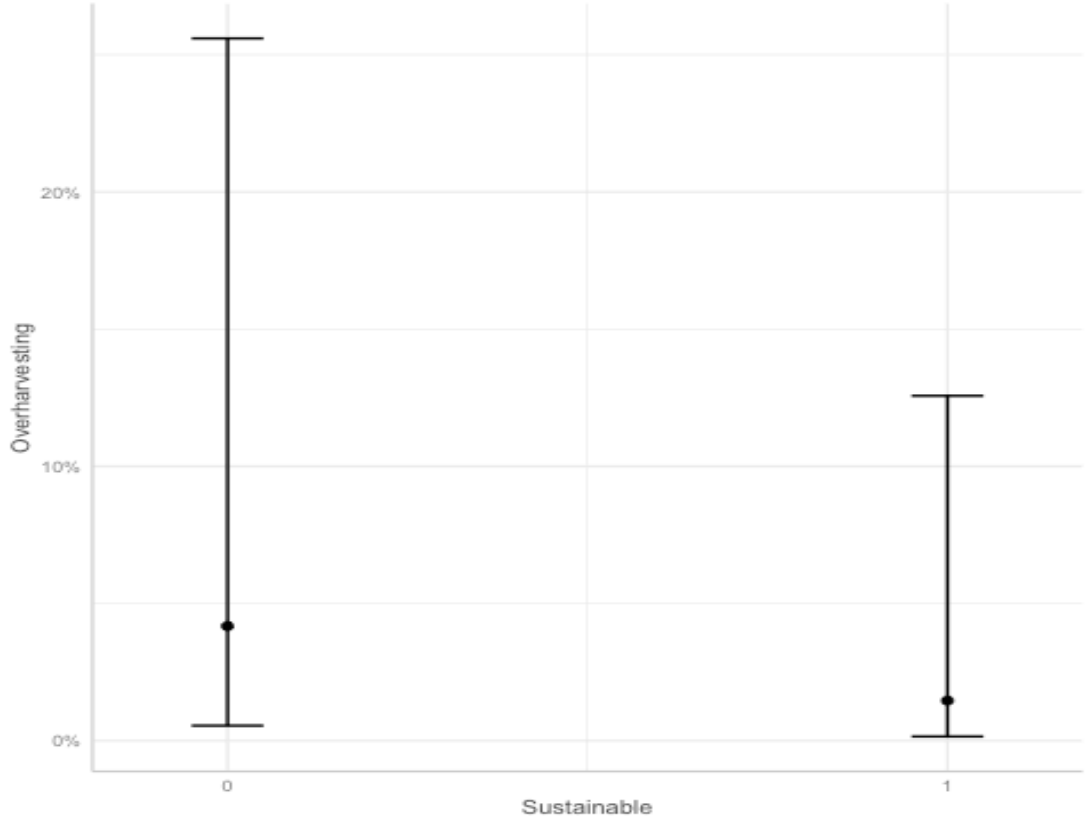
Additionally, I run a linear model (Model 73_lin) that has as dependent variable the number of rounds in which a participant overharvested during the game. The analysis (see Appendix A.6-2.1.4) shows five statistically significant combinations, which support the dots in the figure above. 'Stimulus'-type medium-sized German, Greek and Portuguese participants (all except French) overharvest in two rounds more often than on average, which the intercept indicates with about one round. In contrast, 'stimulus'-type medium-sized French and 'no type' medium-sized Greek participants do not overharvest on average. This gives an indication that one might also see more granular effects for the logit regressions above if the dataset was larger. All in all, the linear model supports the prevalence of the 'stimulus'-oriented type for overharvesting.

6.2.2.2.3 Risk aversion and sustainability

In order to disentangle notions of risk aversion and sustainability from economic policy ideology, there are two more hypotheses. The 'risk aversion hypothesis' states that participants who rather invest in ladders, supposed to reflect 'risk-taking'-oriented cultivation, than in shepherds, supposed to reflect 'risk-averse'-oriented cultivation, are more likely to overharvest. The 'sustainability hypothesis' states that participants who always leave two apples per tree on the ground and never harvest small apples, supposed to reflect 'sustainability'-oriented cultivation, are less likely to overharvest. 'Risk aversion' does not show a statistically significant effect in all models. The magnitude of the coefficient for the 'risk-averse' type has a small significance (comparing the coefficient with the standard error), and the effect is negative as expected, but it is not statistically significant. [227]

227 In the aggregate models 7ff. (see appendix A.6.-2.1.4), the effect of 'risk-averse' becomes statistically significant, also the effect of 'risk taking' in the final model 73 though with an unexpected negative direction. This is an indication that the effect might indeed exist and would

Figure 6.2-6: Predicted probabilities depending on 'sustainable'.



Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The solid lines indicate the 95% confidence interval.

In contrast, in all models, 'sustainable' shows a strong negative effect on overharvesting as expected. The magnitude is large and the effect is statistically significant. The figure below shows the magnitude of the effect as predicted probabilities. Participants who harvest sustainably on own trees are less likely to overharvest by three percentage points compared to the other participants.[228] For those participants, the principle to harvest sustainably in general, i.e., also on shared trees, is hence more prominent in the data than the alternative scenario, where such

be more prominent in a larger dataset. This also indicates that this variable rather measures risk aversion than patience as expected. According to the hypothesis, one can expect risk-averse participants not to breach the harvesting rules. In contrast, one would expect impatient participants to overharvest to have sufficient energy points and not needing to wait long for that.

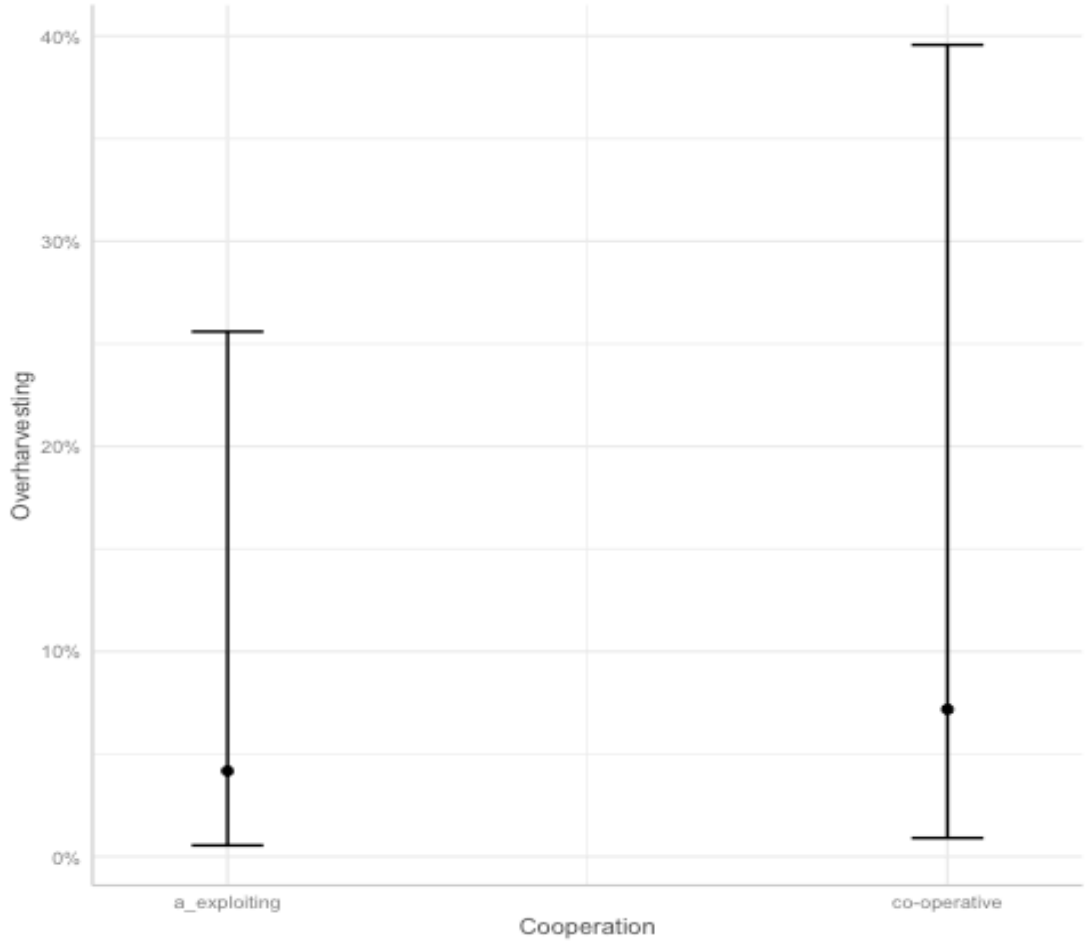
228 The confidence bounds for when 'sustainable' equals 1 are more concentrated. This could indicate principled behaviour as participants who play 'sustainably' are also more likely to refuse overharvesting. In turn, other participants who play 'sustainably' might overharvest occasionally, which reflects in the rather large variation shown by the large confidence bounds for when the variable takes '0'. This might be due to the definition of the variable being rather strict: participants that harvested unsustainably only once during the game would also classify as '0'.

participants would use overharvesting on shared trees as an option to harvest more in order to avoid unsustainable harvesting on own trees.

6.2.2.3 Strategic interactions

The hypotheses state that participants who invest in shared trees are less likely to overharvest ('co-operation hypothesis') and that the more deals that a participant seals over the course of the game, the more likely they are to overharvest ('reciprocal voting hypothesis'). The 'co-operation' variable shows surprising results that are opposite to expectations. In all models, the effect is positive and statistically significant (except for Model 5). Those who harvest any kind of apple from any shared tree throughout the game, and do not invest once in shared trees ('exploiting' type) are less likely to overharvest than those who also invest in shared trees ('co-operative' type). The explanation could be that these participants might consider it their 'right' to overharvest since they also invested in the trees and, thereby, contributed to the maintenance and improvement of the common-pool resource. The figure below illustrates the effect, and shows that 'co-operative' type participants are three percentage points more likely to overharvest than 'exploiting'-type participants.

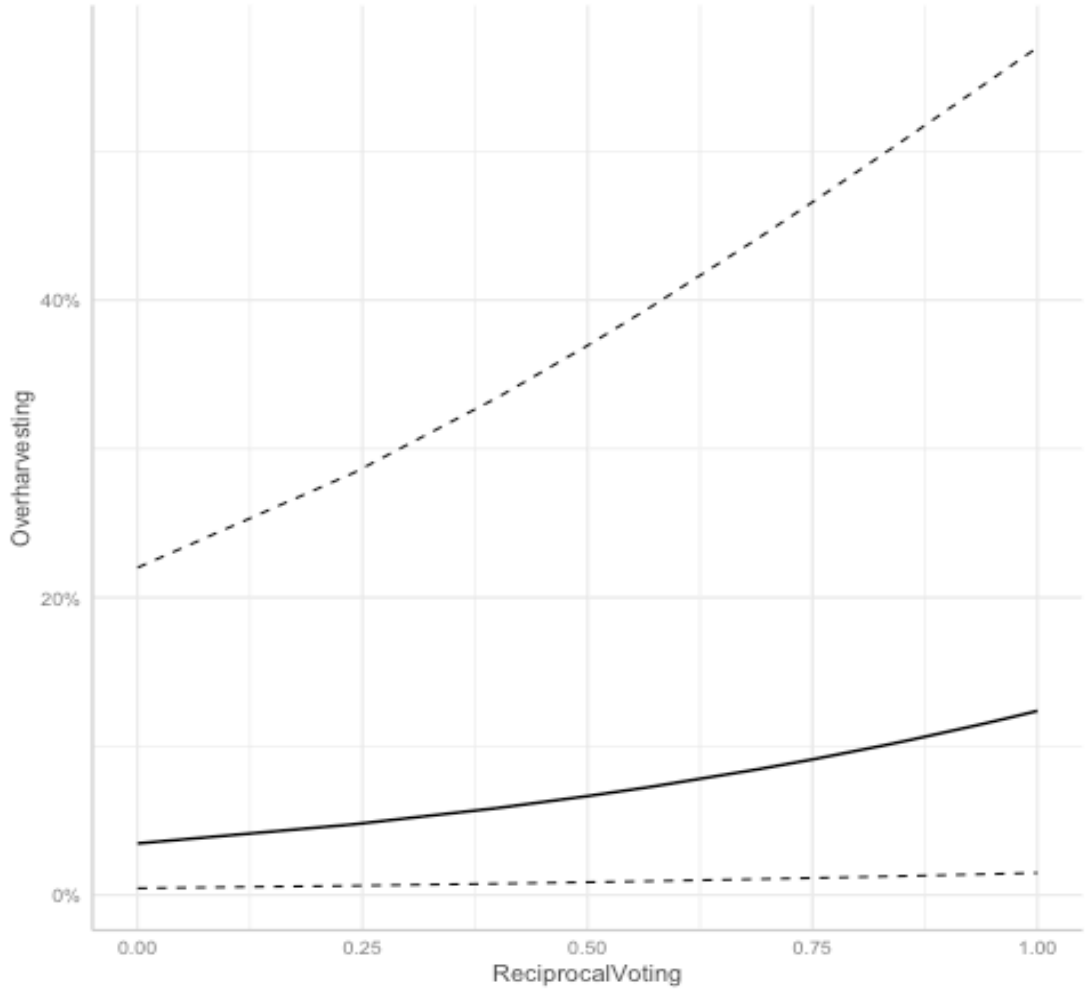
Figure 6.2-7: Predicted probabilities depending on 'co-operation type'.



Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The solid lines indicate the 95% confidence interval.

In all models, the effect of 'reciprocal voting' is positive as expected and statistically significant. The figure below illustrates this. Having sealed and always complied to deals over the course of the game, i.e., when the variable takes the value '1', increases the probability of overharvesting by almost ten percentage points.

Figure 6.2-8: Predicted probabilities depending on 'reciprocal voting'.



Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The dashed lines indicate the 95% confidence interval.

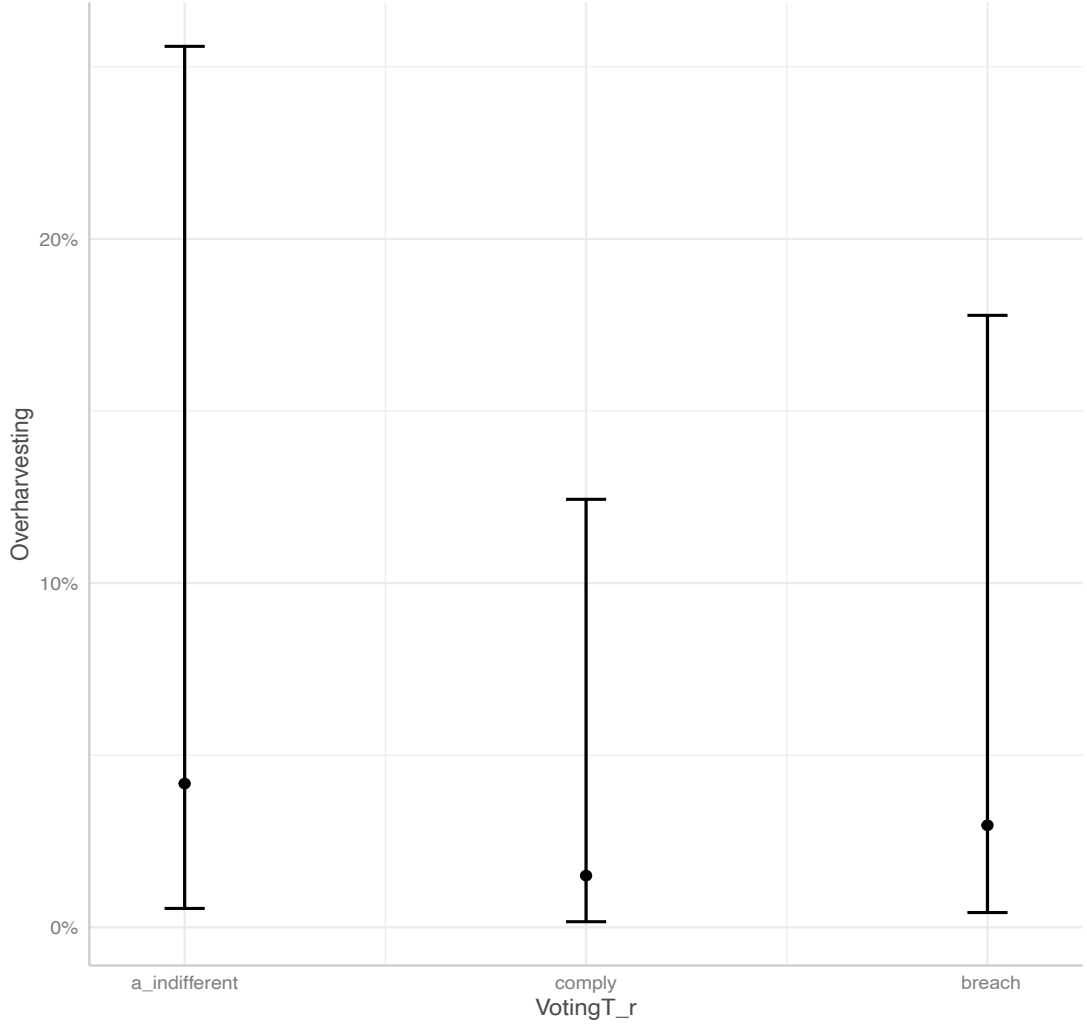
6.2.2.4 Rule adherence

The 'rule adherence' hypothesis states that participants that vote correctly are less likely to overharvest. In Models 6 and 61, compliance with the voting rules shows the expected negative and statistically significant effect on overharvesting.[229] The figure below shows that the likelihood to overharvest is only 2%. The effect for 'breach' is not significantly different from those who are indifferent towards the voting rules, which is the baseline category of this variable. The figure below, however, indicates that breaching might also have a negative effect on overharvesting compared to the

229 In Models 5 and 62, the coefficient is not statistically significant, but the magnitude is significant and similar to the other models.

'indifferent' type. All in all, participants who do not vote at all are more likely to overharvest than those who do.

Figure 6.2-9: Predicted probabilities depending on 'voting'.



Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The solid lines indicate the 95% confidence interval.

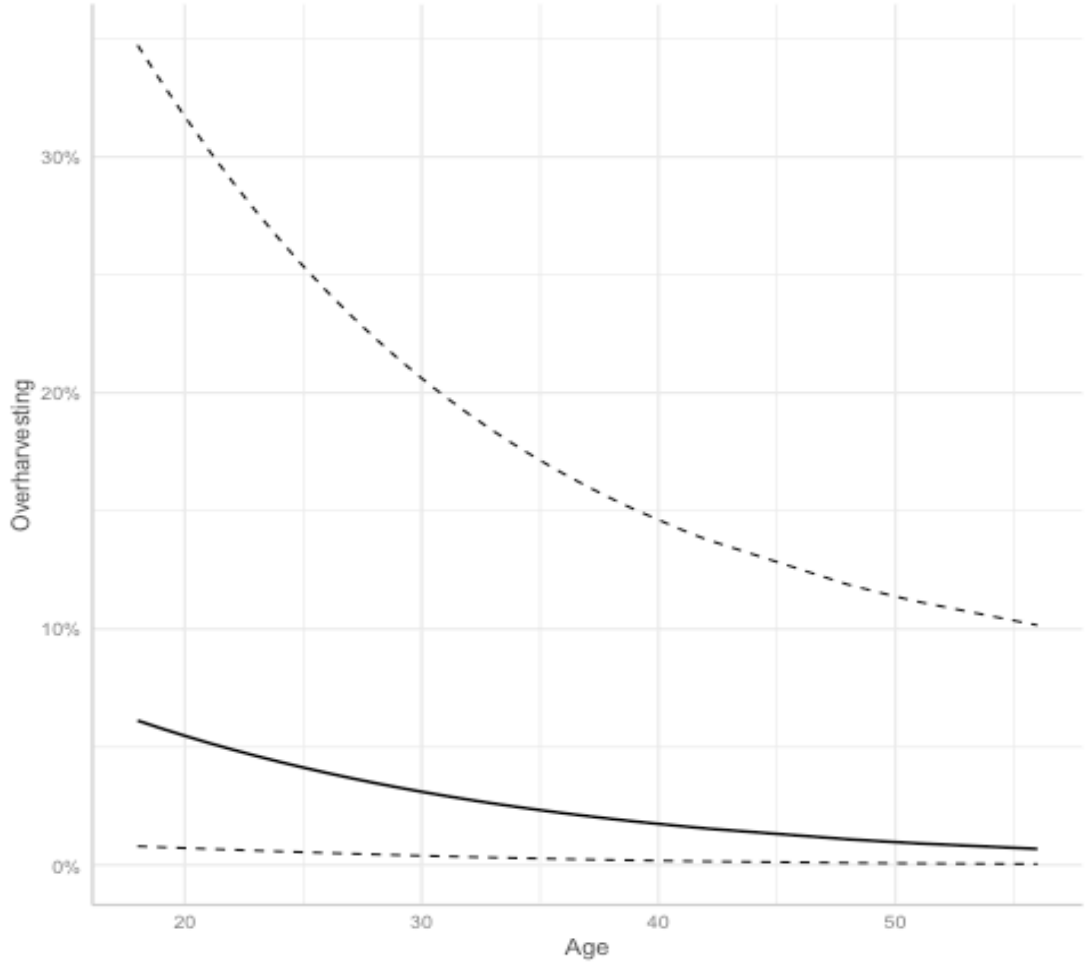
6.2.2.5 Control variables

Of all suggested control variables, only 'age' and 'capital'[230] are statistically significant in all models (see Appendix A.6-2.1.1). They also have very similar magnitudes across all models, which shows that there does not seem to be an

[230] As argued above, specific variables from the survey are considered in the analyses to control for a potential bias in the sample stemming from the recruitment of participants. While bachelor and master students from capitals were the target group to ensure comparability across countries, also other participants from other areas in the country took part. The sample was not restricted to include only the target group and, therefore, I had to include respective variables in the analyses to control for the composition.

interference with the other variables. This shows a typical result for experiments, where life experience stemming from age and living in a large city might affect co-operation in general. The effect of 'capital' could also reflect non-experience with agriculture for participants living in the capital, which might make them more prone to focusing more on rule compliance than on cultivating the field. However, this is only speculation.[231] The figure below shows a decline of about five percentage points in the probability to overharvest from young to older participants.

Figure 6.2-10: Predicted probabilities depending on 'age'.

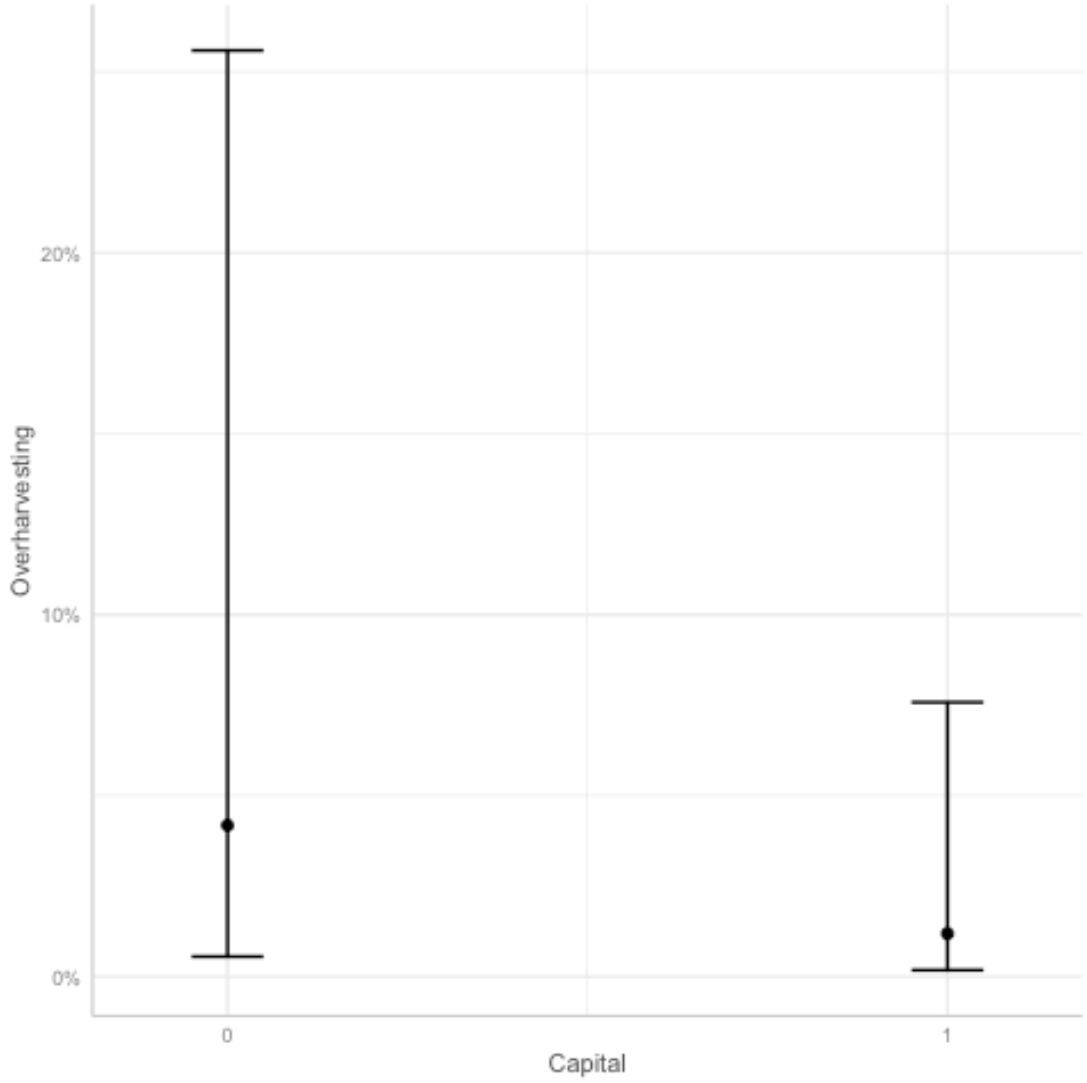


Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The dashed lines indicate the 95% confidence interval.

231 This could be validated by replicating the experiment with a respective sample, as the measure for not living in the capital is rather broad and can entail many more backgrounds than just a rural one. Such participants could also have lived in other cities before.

The effect of ‘capital’ is also remarkably large. Despite the fact that the lines overlap, the confidence bounds for those who do live in the capital are rather small, and behave rather similarly in contrast to those who do not live in the capital. The explanation could be that participants who live in the capital are rather experienced with co-operation and used to rule compliance. However, this is only speculation, and it is not relevant for the analysis.

Figure 6.2-11: Predicted probabilities depending on ‘capital’.

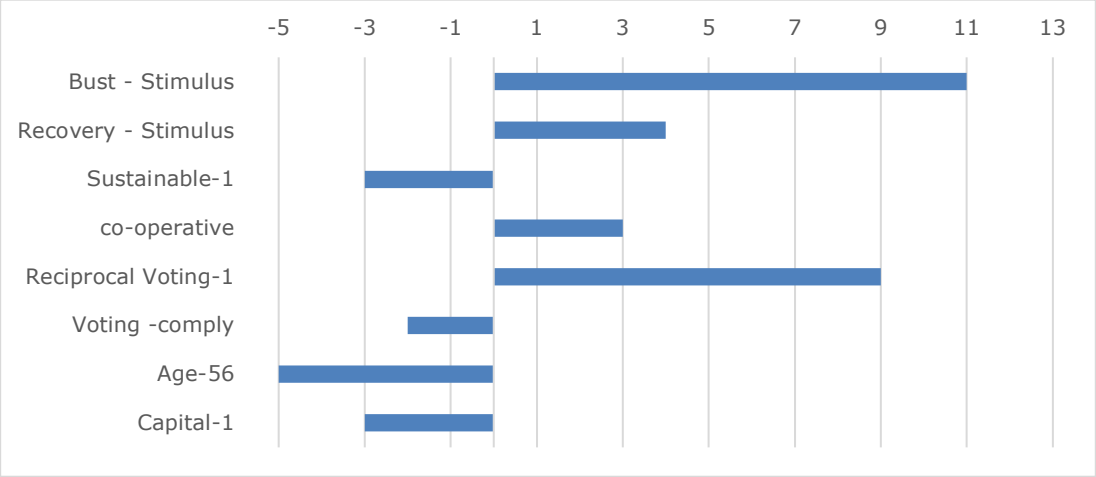


Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The solid lines indicate the 95% confidence interval.

6.2.3 Summary

Two tables help to summarise the results. The first table provides an idea about the size of the effects. The second table summarises the size, direction and statistical significance of the effects, and compares them to the hypotheses.

Figure 6.2-12: Predicted probabilities of statistically significant coefficients.



Source: own calculation based on Model 61; percentage points of predicted probabilities as difference towards baseline category; for interaction of 'business cycle' and 'investment type', the baseline category is 'investment type: none'. For ease of reading, no confidence intervals are shown.

The table above shows the predicted probabilities of all statistically significant variables. Depending on a variable's value, and in comparison to its baseline value, this shows the probability that 'overharvesting' equals 1. The largest probability is related to 'stimulus'-type participants in 'bust' rounds (11pp). These participants are eleven percentage points more likely than 'no type'-participants in 'boom' rounds, which represent the baseline combination. The results also show that 'stimulus'-type participants in 'recovery' rounds are more likely to overharvest by four percentage points. In comparison to 'no type' participants, 'prevention'-participants are similarly likely to overharvest. This supports the 'economic policy ideology' hypotheses as 'stimulus'-type participants are more likely to overharvest than 'prevention'-type participants. The results also support the role of 'bust' as suggested in the 'business cycle' hypothesis and, additionally, show that also in 'recovery' rounds, overharvesting is more likely for 'stimulus'-type participants.

Secondly, participants who engage in reciprocal voting are nine percentage points more likely to overharvest than those who never did during the game. This

corresponds to the hypothesis that participants who engage in reciprocal voting to circumvent getting sanctioned, also overharvest more often.

Thirdly, it follows ‘age’ with a negative effect of five percentage points, and ‘capital’ with three percentage points. This shows a common result for experiments, where life experience stemming from age and living in a large city might affect co-operation in general.

Likewise, the predicted probabilities of ‘sustainable’ and ‘co-operative’-participants are three percentage points large. In line with the hypothesis, ‘sustainable’ participants are less likely to overharvest than the others. Instead, the ‘co-operative’ type is more likely to overharvest than the ‘exploiting’ type, which goes against the hypothesis. This is a surprising result, the explanation of which could be that these participants might consider it their ‘right’ to overharvest since they also invested in the trees and, thereby, contributed to the common-pool resource.

Finally, compliance with the voting rules reduces the probability of overharvesting by two percentage points compared to ‘indifferent’. Participants that breach the voting rules are not significantly differ in their likelihood to overharvest than participants that are ‘indifferent’ towards the voting rules. This is in line with the hypothesis, even though the size of the effect is surprisingly small compared to the other variables. This indicates that voting rule compliance might not be as decisive for overharvesting rule compliance as was expected.

The table below sums up the results. In conclusion, one can see that most explanatory power lies with the ‘business cycle’, the ‘stimulus’-type reflecting the ‘economic policy ideology’ hypothesis, and ‘reciprocal voting’. In particular, the results support the idea that the role of the investment type is mediated by the business cycle. Compliance with the voting rules and ‘sustainable’ show a rather small effect, which is an interesting result, especially for the former. A surprising result is the rejection of the ‘co-operation’ hypothesis. Moreover, the results do not support the hypotheses on ‘size’ and ‘nationality’. However, the analyses indicate that there might be different patterns across both, also with regard to the ‘investment’-type. This would require further research with, for instance, larger or different samples.

Table 6.2-13: Empirical findings and hypotheses – overharvesting.

	Size of predicted probabilities	Direction of effect	Statistically significant?	Hypothesis supported?	Hypothesis
Economic policy ideology / cultivation strategy					
Investment type	2% (no type – baseline) 2% (prevention) 4% (stimulus)	none to +2pp	yes (only stimulus)	**Hypothesis supported**	**Economic policy ideology hypothesis (EH):** Participants who rather invest in beekeepers and manure than in complete water irrigation systems are more likely to overharvest.
Risk aversion	4% (no type – baseline) 3% (risk-averse) 4% (risk-taking)	-1pp to none	no	**Hypothesis not supported**	**Risk aversion hypothesis (RiH):** Participants who rather invest in ladders than in shepherds are more likely to overharvest.
Sustainable harvesting	4% (0 – baseline) 1% (1)	-3pp	yes	**Hypothesis supported**	**Sustainability hypothesis (SuH):** Participants who always leave two apples per tree on the ground and never harvest small apples are less likely to overharvest.
Economic necessity and political importance					
Business cycle	2% (Boom – baseline) 8% (Bust) 4% (Recovery)	+2pp to +5pp	yes	**Hypothesis supported**	**Business cycle hypothesis (BH):** Participants are more likely to over-harvest in bust rounds (Rounds 4, 7 and 10) than in recovery (Rounds 5, 8 and 11) or boom rounds (Rounds 6 and 9).
Size	4% for all three values	none	no	**Hypothesis not supported**	**Size hypothesis (SH):** Participants playing as small or large farmers are more likely to overharvest than medium-sized ones.

Nationality	4% (France – baseline) 7% (Germany) 8% (Greece) 3% (Portugal)	-1pp to +4pp	no	**Hypothesis not supported**	**North-South hypothesis (NSH):** Participants from member states that are geographically located in the 'South', i.e., Greece and Portugal, are more likely to overharvest than those from the 'North', i.e., France and Germany.
		Strategic interactions			
Co-operation	4% (exploiting – baseline) 7% (co-operative)	+3pp	yes	**Hypothesis rejected**	**Co-operation hypothesis (CH):** Participants who invest in shared trees are less likely to overharvest.
Reciprocal voting	3% (0 – baseline) 12% (1)	+9pp	yes	**Hypothesis supported**	**Reciprocal voting hypothesis (RVH):** The more deals a participant seals over the course of the game, the more likely they are to overharvest.
		Rule adherence			
Voting rule	4% (indifferent – baseline) 2% (comply) 3% (breach)	-2pp to -1pp	yes (only comply)	**Hypothesis supported**	**Rule adherence hypothesis (RuH):** Participants that vote correctly are less likely to overharvest.
		Control variables			
Age	6% (18 – baseline) 1% (56)	-5pp	yes	---	---
Capital	4% (0 – baseline) 1% (1)	-3pp	yes	---	---

Source: own description. The first column shows the predicted probability assigned to a variable's value that 'overharvesting' equals 1, all else equal. The second column shows the probability change associated with a unit or a value change within one variable compared to its baseline value. The third column refers to Model 61, and for variables which are part of an interaction, i.e., 'investment type' and 'business cycle', to Model 6 so as to account for their pure effect.

6.3 Compliance at Stage 2: analysis of voting behaviour

Similar to the previous section, this section uses regression analysis to better understand voting incorrectly, i.e., breaches from the voting rule. The section is organised as follows: it first explains the methodological background (6.3.1), it then presents the analysis and discusses the empirical results (6.3.2). The summary compares the results to the hypotheses (6.3.3).

6.3.1 Statistical prerequisites

Given that the dependent variable has three values, 'breaching', 'compliance' and 'indifferent' for not having voted at all, one needs to run two separate analyses, with the 'indifferent' cases counting towards 'compliance' in order to analyse 'breaching' and vice versa.[232] Below, the table only present the results for the former. For the latter, there are similar results,[233] but the models are less successful in correctly predicting cases (75% compared to 80% of the 'breaching' models). The analyses use the same type of regressions as for the analysis of 'overharvesting'. This carries the assumption that each observation per individuum is independent from the other. One can assume that 'voting incorrectly' does not depend on the previous or the next round, but rather on the structure of specific rounds following weather events. Participants might perceive a path dependency stemming from having formed non-punishment alliances in previous rounds. However, there is no such obligation by design.

In contrast to the Stage 1 analyses, the table below does not provide indication that a four-way interaction might be insightful, even though conceptually, it would have been interesting. The table indicates that the majority of cases spread within each variable, such as for 'business cycle', 'size' and 'investment type', while 'nationality' does not seem to show particular patterns. Therefore, the analyses only use the full model (Model 5) and the full model with an interaction between 'business cycle' and 'investment type' (Models 61 and 62).

232 We cannot restrict the dataset to exclude 'indifferent' cases. This would have led to a different number of observations per participant ID, which impedes panel structure analyses.

233 Some variables show as more pronounced which is likely due to the fact that for them, the dependent variable's value 'indifferent' lies closer to 'breach' than to 'comply'. For instance, the analyses find a larger effect for 'risk-averse', which is consistent with the hypothesis.

Table 6.3-1: 4-way combination – number of cases for voting incorrectly.

		Boom			Bust			Recovery			Total
		small	medium	large	small	medium	large	small	medium	large	
France	**a_none**	1	1	4	3	4	8	9	2	9	41
	Stimulus	3	2	2	8	10	8	11	8	8	60
	Prevention	2	4	2	8	24	11	6	28	13	98
Germany	**a_none**	4	5	2	12	17	8	14	18	10	90
	Stimulus	2	1	2	7	11	17	11	14	13	78
	Prevention	5	7	4	18	20	26	20	20	26	**146**
Greece	**a_none**	6	3	2	13	8	11	15	11	11	80
	Stimulus	4	6	9	13	21	17	10	19	22	**121**
	Prevention	1	0	2	4	4	7	6	7	7	38
Portugal	**a_none**	0	3	2	2	5	6	1	6	6	31
	Stimulus	2	0	4	5	7	11	5	9	13	56
	Prevention	0	1	1	5	17	12	6	15	11	68
Total		30	33	36	98	**148**	**142**	**114**	**157**	**149**	907

Source: own description, yellow marks indicate combinations with at least five cases. Recall that cases are not evenly spread across all values of a variable. For instance, there are more Germans than Portuguese in the game. Therefore, absolute numbers cannot be compared.

6.3.2 Analyses of results

In the following, the results are discussed alongside the hypotheses. The section starts with the treatment variables, followed by economic policy ideology, strategic interaction, and rule adherence.

The table below shows the regression models. The appendix provides a sequential building of the models similar to the Stage 1 analysis (A.6-2.2.1) and for checking the goodness-of-fit (A.6-2.2.2). Model 5 represents the full model without an interaction. Model 61 includes the interaction of 'business cycle' and 'investment type'. Model 62 uses the per round variable for the investment type to reflect what items participants purchase in bust rounds. For discussing all variables' effects, Model 61 is used because conceptually, it fits best what the analyses seek to explore, and it also has a good goodness-of-fit (see Appendix A.6-2.2.2).

Table 6.3-2: Regression analyses for voting incorrectly per round.

	Dependent variable: Voting incorrectly perR		
	(5) (full)	(61) (interaction)	(62) (interactionR)
Sizemedium	-0.5292*	-0.5283*	-0.5310*
	(0.3075)	(0.3119)	(0.3099)
Sizesmall	-1.6208***	-1.6331***	-1.6484***
	(0.3363)	(0.3411)	(0.3391)
NationalityGermany	0.0269	0.0315	0.0524
	(0.3470)	(0.3522)	(0.3492)
NationalityGreece	-0.0101	-0.0161	-0.0486
	(0.3758)	(0.3810)	(0.3732)
NationalityPortugal	-0.6147	-0.6404	-0.6213
	(0.3978)	(0.4038)	(0.4006)
BusinessCyclebust	2.1352***	1.5273***	1.7262***
	(0.1738)	(0.3062)	(0.2154)
BusinessCyclerecovery	2.3653***	1.9206***	2.2277***
	(0.1757)	(0.3080)	(0.2068)
InvestmentTypeprevention	0.1030	-1.0772**	
	(0.3395)	(0.4909)	
InvestmentTypestimulus	-0.1518	-0.3075	
	(0.3334)	(0.4615)	
InvestmentTypeRprevention			-0.7563*
			(0.3980)
InvestmentTypeRstimulus			-0.5171
			(0.4645)
RiskAversionrisk-averse	-0.1819	-0.1855	-0.1793
	(0.2986)	(0.3027)	(0.2996)
RiskAversionrisk-taking	-0.0026	0.0026	-0.0401
	(0.4037)	(0.4094)	(0.4040)
Sustainable1	-0.0561	-0.0692	-0.0636
	(0.3115)	(0.3161)	(0.3136)
Cooperationco-operative	0.1772	0.1807	0.1360
	(0.2686)	(0.2724)	(0.2577)
ReciprocalVoting	2.6197***	2.6706***	2.7048***
	(0.3864)	(0.3929)	(0.3858)
OverharvestingTA	0.2556*	0.2535*	0.2434*
	(0.1309)	(0.1330)	(0.1313)
BusinessCyclebust:InvestmentTypeprevention		1.5135***	
		(0.4431)	
BusinessCyclerecovery:InvestmentTypeprevention		1.3128***	
		(0.4443)	
BusinessCyclebust:InvestmentTypestimulus		0.3127	
		(0.4035)	
BusinessCyclerecovery:InvestmentTypestimulus		0.0712	
		(0.4045)	
BusinessCyclebust:InvestmentTypeRprevention			1.2197***
			(0.4411)
BusinessCyclerecovery:InvestmentTypeRprevention			0.4349
			(0.4431)

BusinessCyclebust:InvestmentTypeRstimulus			0.8704*
			(0.5191)
BusinessCyclerecovery:InvestmentTypeRstimulus			0.3216
			(0.5667)
Constant	-2.4535***	-2.0151***	-2.2279***
	(0.5029)	(0.5428)	(0.4886)
Observations	2,616	2,616	2,616
Log Likelihood	-1,260.2610	-1,252.5630	-1,254.4790
Akaike Inf. Crit.	2,554.5220	2,547.1250	2,550.9580
Bayesian Inf. Crit.	2,654.3020	2,670.3830	2,674.2160
Note:			*p<0.1; **p<0.05; ***p<0.01

Source: own calculation.

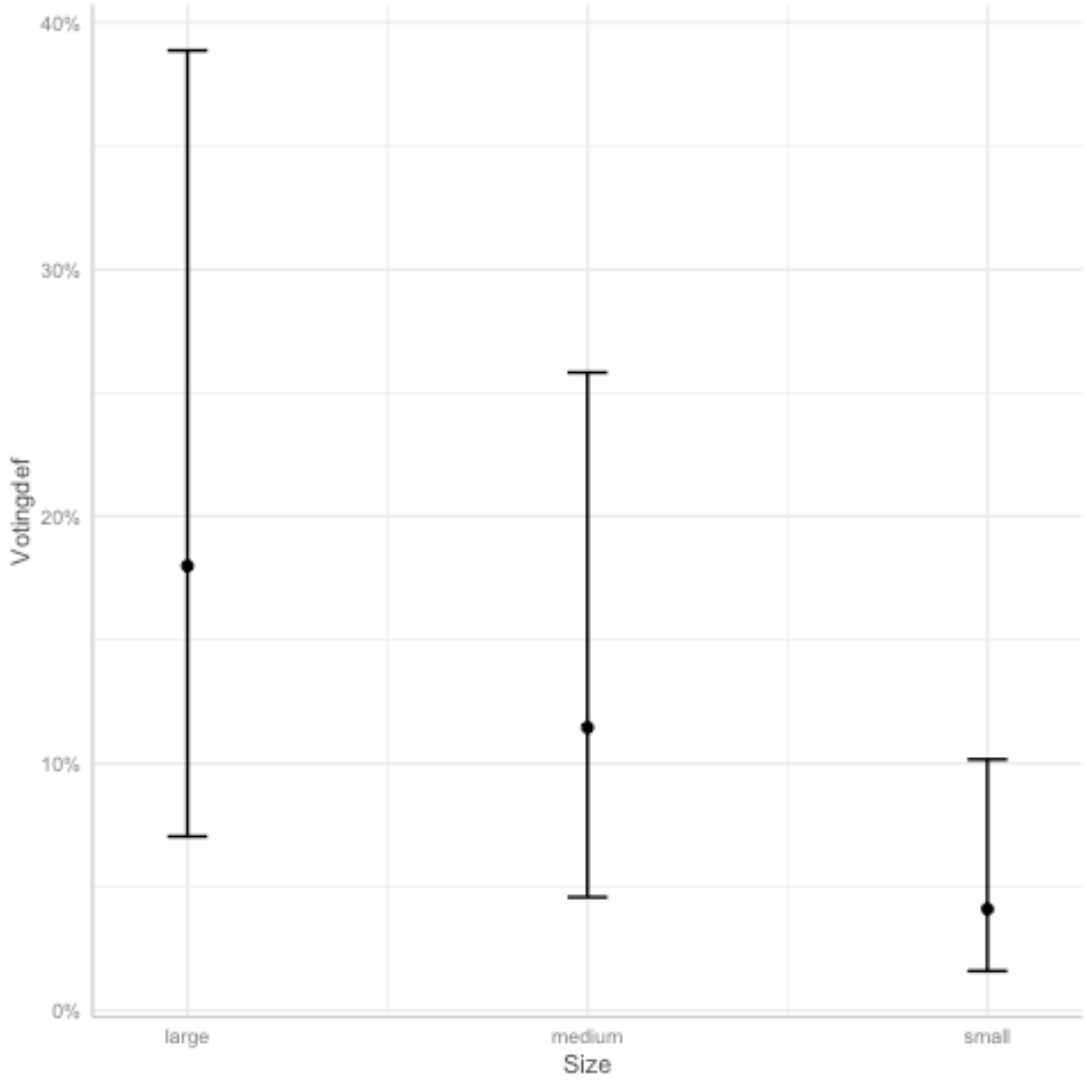
6.3.2.1 Economic necessity and political importance

The hypotheses state that participants are more likely to vote incorrectly in bust rounds ('business cycle hypothesis'), in cases where they are assigned to large farmers ('size hypothesis'), or come from Greece or Portugal ('North-South hypothesis'). In all models, 'nationality' does not have a statistically significant effect. Instead, the 'size' hypothesis sees strong support: medium-sized and small participants are significantly less likely to vote incorrectly. Moreover, the results support the 'business cycle' hypothesis: 'bust' has the expected positive effect on overharvesting, and is statistically significant in all models. Additionally, participants are also more likely to vote incorrectly in 'recovery' rounds as compared to 'boom' rounds.

The figure below illustrates the variable's effects by computing predicted probabilities that 'voting incorrectly' equals 1 for different levels of the respective independent variable, while keeping constant all other variables at their mean (for numerical variables), or reference value (for categorical variables, see Appendix A.6-2.2.3 for a table).[234] Despite the fact that the 95%-confidence bounds overlap even though the variable is statistically significant, the different dots indicate that the predicted probability for 'voting incorrectly' is 14 percentage points less for small, and 7 for medium-sized participants compared to large participants. All else equal, a large participant is per se 18% likely to vote incorrectly. This might well reflect the political power that small participants perceive themselves able to exert. Accordingly, they might refrain from voting incorrectly, but instead try to push for rule compliance.

234 The R code is inspired by Lüdecke (2021).

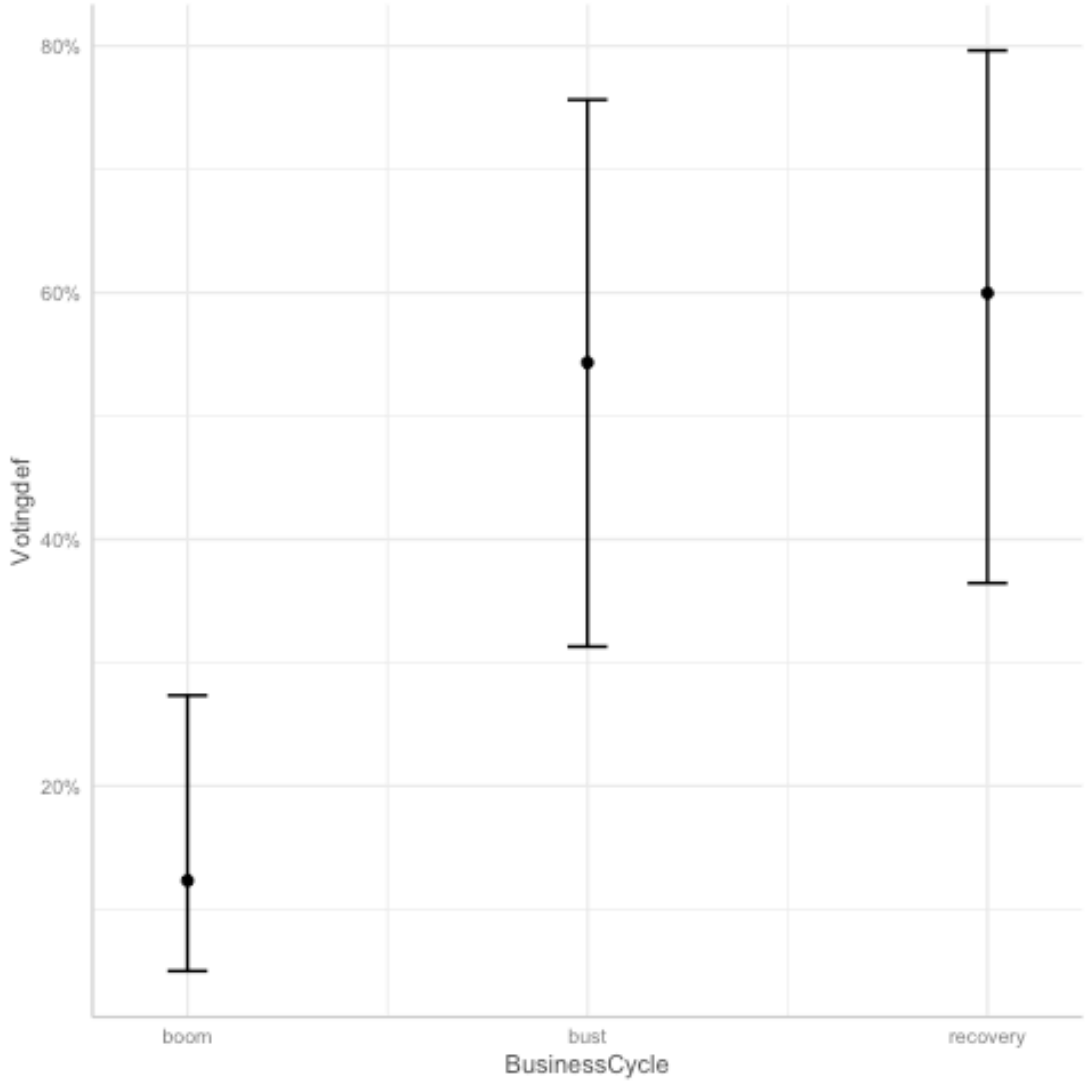
Figure 6.3-3: Predicted probabilities depending on 'size'.



Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The solid lines indicate the 95% confidence interval.

The effect of the 'business cycle' is remarkably high. The likelihood of voting incorrectly in 'bust' rounds is about 42 percentage points higher than in 'boom' rounds. In 'recovery' rounds, the likelihood of 'incorrect voting' is highest, and stands at 60% all else equal. This coincides with the experiment's design where punishment starts with a warning before participants can vote for a fine and, thereby, is lagged by one round. Nonetheless, it is not self-evident that participants would vote even more often incorrectly on a fine than on a warning. 'Boom' rounds range only at 12%. In contrast to the analysis of Stage 1 compliance, one can see that breaches are most likely in 'recovery' rounds, not in 'bust' rounds.

Figure 6.3-4: Predicted probabilities depending on 'business cycle'.



Source: own calculation based on Model 5. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The solid lines indicate the 95% confidence interval.

6.3.2.2 Economic policy ideology

6.3.2.2.1 Investment typology and the role of the business cycle

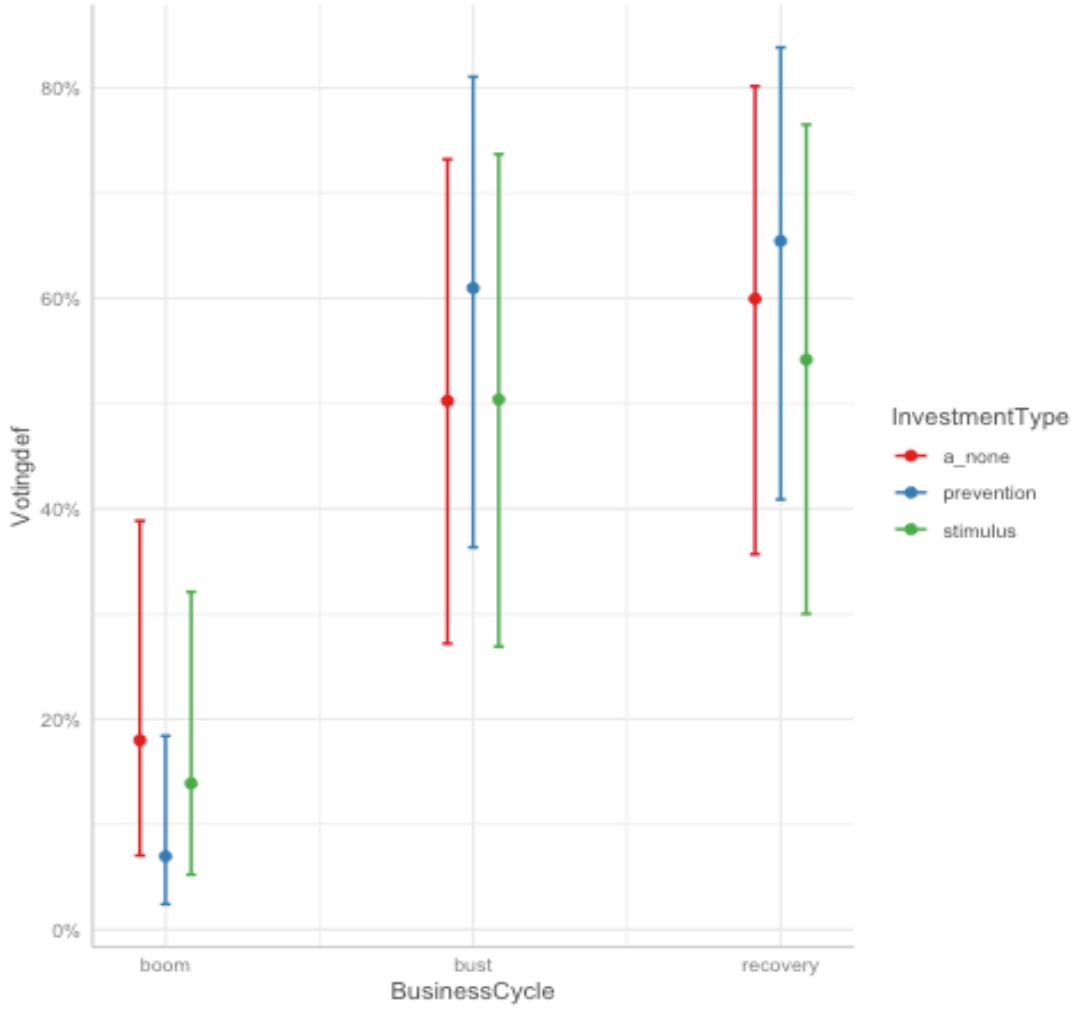
The 'economic policy ideology' hypothesis states that participants who rather invest in beekeepers and manure, supposed to reflect 'stimulus'-oriented cultivation, than in complete water irrigation systems, supposed to reflect 'prevention'-oriented cultivation, are more likely to vote incorrectly. In contrast to the hypothesis, Model 61 shows that it is not 'stimulus'-type participants but 'prevention'-type participants that are more likely to vote incorrectly in 'bust' and 'recovery' rounds. In normal times, though, they are significantly less likely to vote incorrectly, which supports the

hypothesis. The ‘stimulus’-type does not have a statistically significant effect as it is not significantly different from the ‘no type’, which serves as a baseline. The study runs the same analyses with ‘voting correctly’ as dependent variable, and this additionally shows that both ‘prevention’- and ‘stimulus’ type are significantly more likely to vote correctly in general (Model 6) and in ‘boom’ rounds than ‘no type’ participants. Given that the baseline is ‘no type’, one cannot conclude that ‘prevention’-type is more or less likely to comply than ‘stimulus’-type.

In Model 62, the interaction reflects the item that participants purchase in specific rounds. The results show that, irrespective of the investment choice, it is ‘bust’ that makes voting incorrectly more likely: all three coefficients are positive and statistically significant, including the plain coefficient on ‘bust’ reflecting the combination with the ‘no type’. However, the model also shows that ‘voting incorrectly’ in ‘recovery’ rounds is driven by participants that do not invest in either (or similarly in both) in such rounds. Participants that invest in ‘stimulus’ or ‘prevention’ items in that round are not more or less likely to vote incorrectly. The coefficients are insignificant. Instead of supporting or rejecting the hypothesis, this shows that engaging in either strategy makes voting rule compliance more likely, compared to ‘no type’ participants who do not invest at all or similarly in both sets of items.

The figure below shows the predicted probabilities for the interaction between the business cycle and the investment type. It shows the large impact of ‘prevention’ in ‘bust’ and ‘recovery’ rounds: the predicted probability ranges around 61% and 65%, respectively. Likewise, the ‘prevention’-type ranges least in boom times, with 7% in comparison to the ‘stimulus’-type at 14%. While ‘stimulus’-type participants are more likely to vote incorrectly in ‘boom’ rounds, ‘prevention’-type participants are more likely in ‘bust’ and ‘recovery’ rounds. The former supports the hypothesis; the latter rejects it.

Figure 6.3-5: Predicted probabilities depending on 'investment type' and 'business cycle'.



Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variables on the x-axis and marked in colours. The solid lines indicate the 95% confidence interval. 'a_none' represents those rounds in which participants have not invested at all or similarly in both 'prevention' and 'stimulus' items.

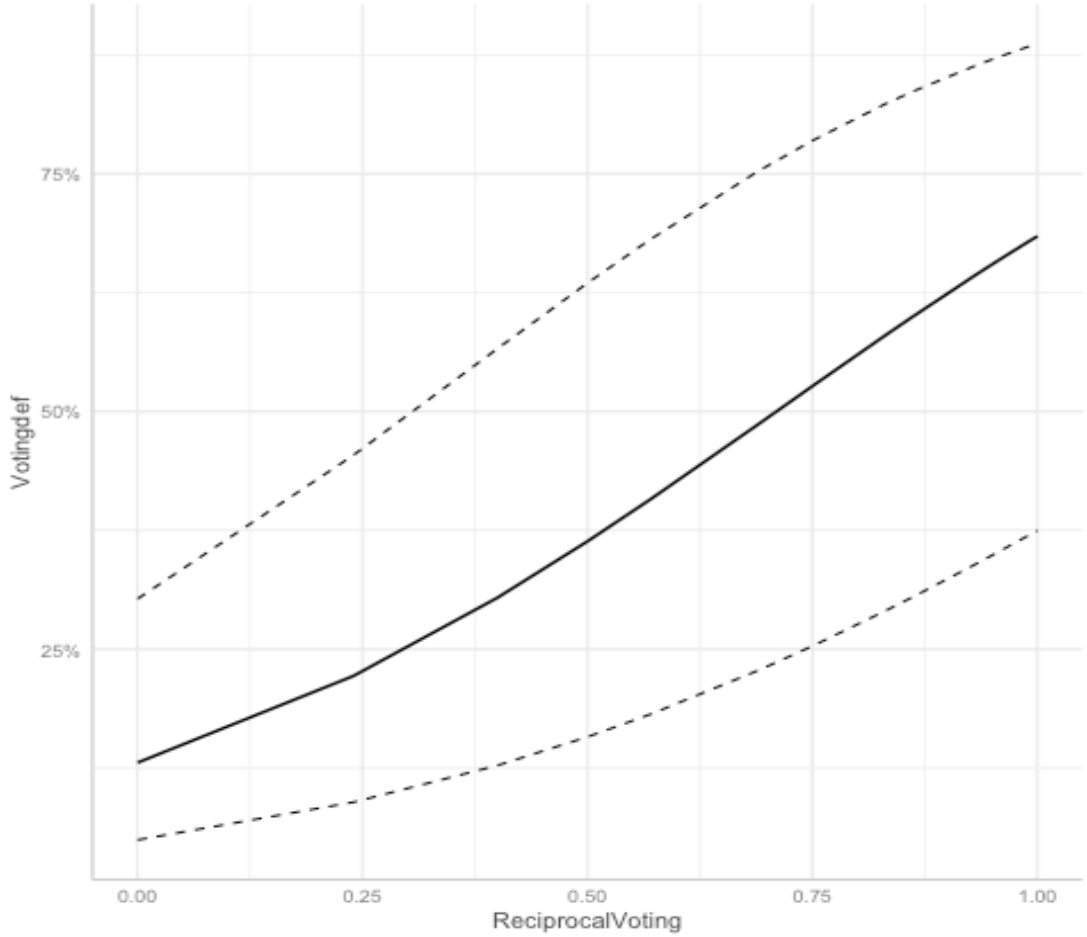
6.3.2.2.2 Risk aversion and sustainability

In order to disentangle notions of risk aversion and sustainability from economic policy ideology, there are two more hypotheses. The 'risk aversion hypothesis' states that participants who rather invest in ladders, supposed to reflect 'risk-taking'-oriented cultivation, than in shepherds, supposed to reflect 'risk-averse'-oriented cultivation, are more likely to vote incorrectly. The 'sustainability hypothesis' states that participants who never harvest small apples and always leave two fallen apples per tree as fertiliser in the field, are less likely to vote incorrectly. In all models, neither 'risk aversion' nor 'sustainable' show a statistically significant effect.

6.3.2.3 Strategic interactions

The hypotheses state that participants who invest in shared trees are less likely to vote incorrectly ('co-operation hypothesis'), and that the more deals that a participant seals over the course of the game, the more likely they are to vote incorrectly ('reciprocal voting hypothesis'). In contrast, in all models, 'reciprocal voting' has a large positive effect on 'voting incorrectly' and is statistically significant, as expected by the hypothesis. The figure below supports this. Having sealed and always complied to deals over the course of the game (reciprocal voting=1) increases the probability of voting incorrectly by 55 percentage points. This leads us to conclude that a vast part of incorrect voting can be considered due to purposefully trying to avoid sanctions and forming alliances with 'reciprocal voting'.

Figure 6.3-6: Predicted probabilities depending on 'reciprocal voting'.

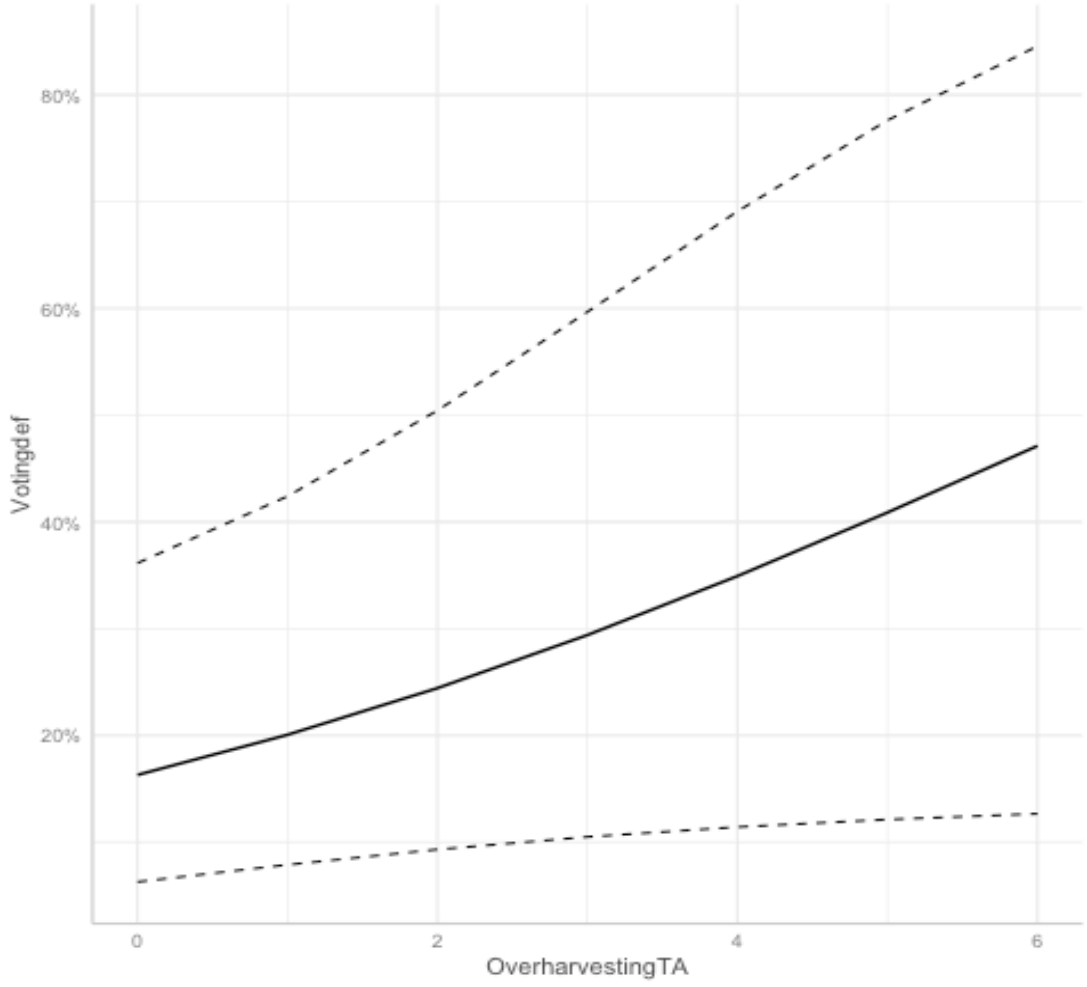


Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The dashed lines indicate the 95% confidence interval.

6.3.2.4 Rule adherence

The ‘rule adherence’ hypothesis states that participants that overharvest are more likely to vote incorrectly. In all models, ‘overharvesting’ shows a positive and statistically significant effect on ‘voting incorrectly’, as expected by the hypothesis. The analyses use a variable that displays the absolute number of rounds in which a participant overharvested (see Appendix A.6-2.2.1 for a comparison of indicators). The figure below supports this. The predicted values have 31 percentage points difference between participants who never overharvest, and those who overharvest in six rounds, which is the observed maximum. The relationship is roughly linear, and this means that one more round of overharvesting increases the probability of voting incorrectly by roughly four percentage points.

Figure 6.3-7: Predicted probabilities depending on ‘overharvesting’.

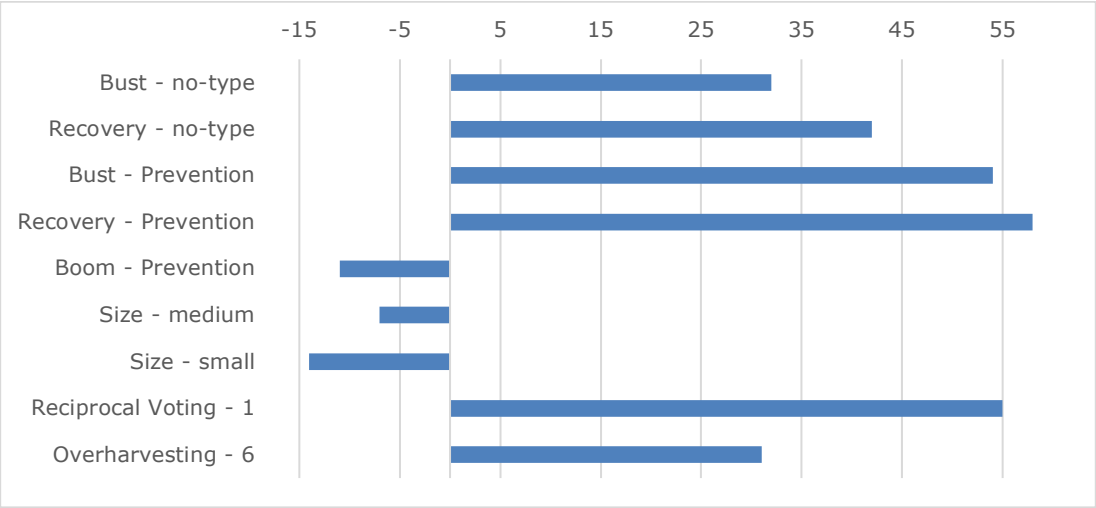


Source: own calculation based on Model 61. The y-axis shows the probability (in %) that overharvesting happens in a particular case, i.e. one participant-round, depending on the different values of the dependent variable on the x-axis. The dashed lines indicate the 95% confidence interval.

6.3.3 Summary

Two tables help summarise the results. The first table provides an idea about the size of the effects. The second table summarises the size, direction and statistical significance of the effects, and compares them to the hypotheses.

Figure 6.3-8: Predicted probabilities of statistically significant coefficients.



Source: own calculation based on Model 61; percentage points of predicted probabilities as difference towards baseline category; for the interaction of 'business cycle' and 'investment type', the baseline category is 'boom' and 'no type', respectively. For ease of reading, no confidence intervals are shown.

The table above shows the predicted probabilities of all statistically significant variables. Depending on a variable's value, and in comparison to its baseline value, this shows the probability that 'voting incorrectly' equals 1. The largest probability is related to the 'prevention'-type in 'recovery' and 'bust' rounds: such participants are 58 and 54 percentage points, respectively, more likely than 'prevention'-type participants in 'boom' rounds, which represent the baseline combination. In contrast, prevention'-type participants in 'boom' rounds are 11 percentage points less likely than 'no type'-participants in 'boom' rounds, which is the respective baseline combination. 'Bust' and 'recovery' rounds carry, in general, large explanatory power as the combinations with the investment types' baseline category 'no type' show. Compared to 'boom rounds' for 'no type', they increase the probability for voting incorrectly by 32 and 42 percentage points, respectively. This supports the 'economic policy ideology' hypotheses as 'prevention'-type participants are less likely to vote incorrectly in 'boom' rounds, however it rejects the hypothesis for 'bust' and 'recovery' rounds. Moreover, the results do not support the hypothesis for 'stimulus'-

type participants. The results strongly support the role of ‘bust’ as suggested in the ‘business cycle’ hypothesis and, additionally, show that also in ‘recovery’ rounds, ‘voting incorrectly’ is more likely than in ‘boom’ rounds. While the hypothesis did not consider such effects for ‘recovery’, in the spirit of the hypothesis, it appears equally meaningful. Moreover, this might reflect the lagging of the voting rules in the experimental design by which one first votes for a warning and then for a fine.

The second largest probability refers to when ‘reciprocal voting’ equals 1, compared to ‘0’ (55 percentage points). This corresponds to the hypothesis according to which this thesis expects participants that engage in reciprocal voting to circumvent getting sanctioned also to be more likely to vote incorrectly. Thirdly, it follows ‘overharvesting’ with a positive effect of 31 percentage points. This is in line with the hypothesis that expects a positive effect from Stage 1 breaches on Stage 2 breaches. Finally, the variable ‘size’ shows that medium-sized and small participants are 7 and 14 percentage points less likely to vote incorrectly than large ones. This supports the hypothesis that states that large participants are more likely to vote incorrectly.

The table below sums up the results. In conclusion, one can see that the most explanatory power lies with the ‘business cycle’, the ‘prevention’-type both reflecting and rejecting the ‘economic policy ideology’ hypothesis, ‘reciprocal voting’ and ‘overharvesting’. In particular, the results support the idea that the role of the investment type is mediated by the business cycle: the effect for ‘prevention’ is even reversed between ‘boom’ rounds on the one hand, and ‘bust’ and ‘recovery’ rounds on the other. ‘Size’ shows the intended negative effect for ‘small’ and ‘medium’. Moreover, the results do not support the hypotheses on ‘nationality’, ‘risk-aversion’, ‘sustainable’, and ‘co-operation’.

Table 6.3-9: Empirical findings and hypotheses – voting incorrectly.

	Size of predicted probabilities	Direction of effect	Statistically significant?	Hypothesis supported?	Hypothesis
Economic policy ideology / cultivation strategy					
Investment type	12% (no type – baseline) 13% (prevention) 11% (stimulus)	-1pp to +1pp	yes (in interaction, see table above)[235]	**Hypothesis partly supported:** supported in 'boom' rounds, rejected in 'bust' and 'recovery' rounds	**Economic policy ideology hypothesis (EH):** Participants who rather invest in beekeepers and manure than in complete water irrigation systems are more likely to vote incorrectly.
Risk aversion	18% (no type – baseline) 18% (risk-taking) 15% (risk-averse)	-3pp to none	no	**Hypothesis not supported**	**Risk aversion hypothesis (RiH):** Participants who rather invest in ladders than in shepherds are more likely to vote incorrectly.
Sustainable harvesting	18% (0 – baseline) 17% (1)	-1pp	no	**Hypothesis not supported**	**Sustainability hypothesis (SuH):** Participants who always leave two apples per tree on the ground and never harvest small apples are less likely to vote incorrectly.
Economic necessity and political importance					
Business cycle	12% (Boom – baseline) 54% (Bust) 60% (Recovery)	+42pp to +48pp	yes	**Hypothesis supported**[236]	**Business cycle hypothesis (BH):** Participants are more likely to vote incorrectly in bust rounds (Rounds 4, 7 and 10) than in recovery (Rounds 5, 8 and 11) or boom rounds (Rounds 6 and 9).

235 The regression for 'voting correctly' shows a positive effect of 'stimulus', which is statistically significant, indicating a negative effect on 'voting incorrectly'.
236 While the hypothesis did not consider such an effect for 'recovery', in the spirit of the hypothesis, it appears equally meaningful.

Size	18% (large – baseline) 14% medium 4% small	-14pp to -4pp	yes	**Hypothesis supported**	**Size hypothesis (SH):** Participants playing as small or large farmers are more likely to vote incorrectly than medium-sized ones.
Nationality	18% (France – baseline) 18% (Germany) 18% (Greece) 10% (Portugal)	-8pp to none	no	**Hypothesis not supported**	**North-South hypothesis (NSH):** Participants from member states that are geographically located in the 'South', i.e., Greece and Portugal, are more likely to vote incorrectly than those from the 'North', i.e., France and Germany.
		Strategic interactions			
Co-operation	18% (exploiting – baseline) 21% (co-operative)	+3pp	no	**Hypothesis not supported**	**Co-operation hypothesis (CH):** Participants who invest in shared trees are less likely to vote incorrectly.
Reciprocal voting	13% (0 – baseline) 68% (1)	+55pp	yes	**Hypothesis supported**	**Reciprocal voting hypothesis (RVH):** The more deals a participant seals over the course of the game, the more likely they are to vote incorrectly.
		Rule adherence			
Overharvesting	16% (0 – baseline) 47% (6)	+31pp	yes	**Hypothesis supported**	**Rule adherence hypothesis (RuH):** Participants that vote correctly are less likely to vote incorrectly.

Source: own description. The first column shows the probability assigned to a variable's value that 'voting incorrectly' equals 1, all else equal. The second column shows the probability change associated with a unit, or a value change within one variable, compared to its baseline value. The third column refers to Model 61 and, for variables which are part of an interaction, i.e., 'investment type' and 'business cycle', to Model 5 so as to account for their pure effect.

6.4 Discussion of experimental results

This experiment set out to investigate the role of economic ideology for rule compliance and an SGP-like governance of a common-pool resource in principle. In order to disentangle competing explanations, it also assesses hypotheses put forward by the literature. Section 6.2 explores breaches from the harvesting rule, and Section 6.3 explores breaches from the voting rules. The expectation was that one would find similar results for both and that they would also affect each other.

The hypothesis for the 'investment type' finds strong support for 'overharvesting', though, shows surprising results for 'voting incorrectly'. The analyses find that indeed 'stimulus' type participants overharvest more often. In particular, this effect is mediated by the business cycle, revealing the differences between 'stimulus' and 'prevention'. Instead, for 'voting incorrectly', the effect is distinct. First of all, the effect is entirely mediated by the 'business cycle'. For 'boom' rounds, the hypothesis is supported, and 'prevention'-type participants are less likely to vote incorrectly. Instead, for 'bust' and 'recovery' rounds, the effect is inverse, and 'prevention'-type are more likely to vote incorrectly. However, incorrect voting could in fact reflect the intention to make the application of the rule stricter rather than less strict. This perspective is supported by the finding that 'reciprocal voting' does not carry the majority of the explanatory power for 'voting incorrectly', and that there is also no significant collinearity. With this reading, the hypothesis would in fact be supported. This needs further research.

Moreover, as part of the cultivation strategy, the analyses find little support for the hypotheses for 'risk aversion' and 'sustainable'. While 'risk aversion' is insignificant in both sets of analyses, 'sustainable' has the intended negative effect on overharvesting, though, the effect is rather small in size. This indicates that considerations for a sustainable harvesting are relevant for overharvesting decisions, but not for implementing the voting rule. Nonetheless, conceptually, it is essential to have the respective items in the experimental design and the variables in the regression models, so as to disentangle such perspectives from the 'investment type'.

Looking at the hypotheses from the literature, one can conclude that the findings provide a nuanced picture. The analyses find strong support for the 'business

cycle' hypothesis. This shows that breaches at both stages are also about economic necessity and, as regards the voting rules, a more generous treatment of other's overharvesting in economically tough times.

The results for 'size' are intriguing. While 'size' does not affect 'overharvesting', large participants are significantly more likely to vote incorrectly, as expected by the hypothesis. This reflects the role of weighted votes for voting. Even though the participant's vote is pivotal, which of course they do not know but could infer from the voting results as they appear to always vote with the majority, they seem to take into account their voting power when making choices. Large ones apparently engage more confidently in voting incorrectly than small or medium-sized ones. In contrast, they do not seem to anticipate their political power in the voting stage when making overharvesting decisions. Or, put the other way around, there seem to be other more salient reasons to decide for overharvesting than a purely strategic perspective (on being able to circumvent sanctions).

Likewise, the results for 'nationality' appear insightful. All else equal, 'nationality' has no effect on compliance with the harvesting and voting rules. This is contrary to the hypothesis, which, however, builds on economic conditions and policy traditions in a country. Such aspects might be less prevalent for university students, or in a fictional experiment about an apple orchard. However, the thesis argued that participants from different countries would also carry general cultural and political backgrounds. The results show that these also do not matter – at least not additional to the explanatory power of the other variables. Nonetheless, the analysis of Stage 1 compliance finds indication that 'nationality' might interact with 'size' and 'investment type'. The small size of the dataset impedes further analyses in that regard. Finally, one could also argue that the none-effect of 'nationality' would be due to the specific sample of university students, and that university students would be a more homogenous group with respect to studying economic preferences and co-operation than their nationality a differing factor.

The role of strategic interactions shows informative results. The findings strongly support the 'reciprocal voting' hypothesis put forward by the literature on the role of forming alliances to avoid sanctions, the so-called 'sinners' solidarity'. The analyses find strong positive effects on breaching at both stages. On the other hand,

co-operation on shared trees shows a positive effect on overharvesting, which is unexpected by the hypothesis. This thesis hypothesised that cultivation of the common-pool resource would render compliance with the rules an intrinsic motivation. However, the analyses find the opposite. The explanation could be that these participants might consider it their 'right' to breach, in order to 'govern' the common-pool resource according to their preferences since they also invested in the trees and, thereby, contributed to the common-pool resource.

Finally, the analyses find some relation between breaching the harvesting rule and breaching the voting rule. While 'voting correctly' slightly reduces the likelihood of overharvesting, 'overharvesting' has a strongly positive effect on 'voting incorrectly'. This is in line with the hypothesis, but the size of the effect is surprisingly weak and strong, respectively. While having overharvested seems to be a strong reason for engaging in 'voting incorrectly', having shown general compliance with the voting rule over the course of the game does not seem to be a decisive factor when making overharvesting decisions. There seem to be other more salient reasons why participants decide against overharvesting rather than sticking to the principle of complying to the rules.

6.5 Applicability towards the real-world SGP

The following section discusses to what extent the experimental results could provide insights for the real-world SGP. To that end, the sections discuss and address concerns and argue that, 1) the SGP-like setting in the experiment reflects situations of scarcity that are similar in logic to the real world, 2) the conflicts of interest that participants face in the experiment are comparable in principle to those of governments, which provides insight into the interplay between the situational constraints and the decision that a participant takes, 3) that policy preferences that manifest over the course of the game can be understood as an expression of 'ideology' because the experiment measures the intention of policy choices without the uncertainty that is usual in the real world, 4) the experiment disentangles the real-world aspects related to 'size' and 'nationality' and only studies the economic importance for others, and the political power for the former and a common cultural background for the latter, and 5) despite the complexity in the real world, the more 'focused' view of the experiment reveals

that the participants' choices are not solely driven by rational considerations, but also by convictions about what action is the 'right' one to take. In addition to the dedicated section on external validity (see Section 5.4.1.2), this section seeks to put the experimental results into a meaningful context for the application to the real-world SGP.

6.5.1 Structure: SGP-like setting is different to the real-world setting

An experiment's strength lies in its level of abstraction, as well as in the extent to which it is embedded in a meaningful yet parsimonious context. The experiment mirrors an SGP-like setting and rules, but it is an abstraction and is not explicitly about economic policy, fiscal policy or monetary stability, as this would be too complex to be reflected in a concise experiment. Moreover, such a design would be too prone to social desirability biases based on recent experiences with the sovereign debt crisis. Instead, the experimental design seeks to break it down to its most basic mechanisms. The experiment is about individual business behaviour complemented by an SGP-like sanctioning procedure for dealing with conflicts over a common-pool resource. The allegory of the apple orchard as a context for decisions seeks to strike a balance between accuracy and simplicity. Apples are the essential source of means for conducting any action, and, therewith are comparable to taxes and borrowing for budget policies. Even though, overharvesting does not destabilise the value of apples as is assumed to be the case with inflation; the conflict over apples from shared parts and the related uncertainty is comparable to the potentially limited use of the common-pool resource in the SGP context, i.e., 'stable money'.

Moreover, 'voting incorrectly' is less explicit in the real world than in the experiment. While members of the Council can abstain from voting, or vote against sanctions, they cannot vote for sanctions if there is no such proposal from the European Commission. However, this thesis argues that voting in the experiment signals preferences towards or against sanctions, which is reflected to a similar extent in the discretion used for voting in the Council and is, therefore, comparable as a dependent variable. In particular, falsely voting 'green' finds correspondence in the real world, and has resulted in not sanctioning breaching member states.

The SGP-like rule limits extraction from a common-pool resource which can be essential for individual prosperity. With the overlapping of a common-pool resource with private goods, this experimental design is novel, and adequately reflects the SGP' case. With this, the experimental design goes beyond standard game theoretical experiments, such as the prisoners' dilemma or the ultimatum game, which are even more abstract and simplistic.

6.5.2 Actors: students are different than governments

While in the real world, governments and parliaments decide; in the experiment, students take decisions. There are two aspects to the difference: it is students taking decisions in a fictitious environment, versus politicians taking decisions in the real world. Therefore, the discussion has to account for the comparison between students and politicians, on the one hand, and the context of their decisions on the other. While experimental literature shows that politicians might behave similarly in experiments to students (Fréchette 2015a), it is not about to compare both when acting in a fictitious environment. Nor is it about to compare both when acting in the real world, where both face very different situations, knowledge, skills and experiences. Instead, it is about to draw insights from students in a fictitious environment and relate them to politicians in the real world.

We argue that it is the interplay between the situational constraints and the decisions that a participant takes that is informative for the real world. There is no need to draw parallels to each aspect in the real world. For instance, one can conclude that situations of economic scarcity make breaches from an SGP-like extraction rule more likely, and that this logic is insightful for the real world. There is no need to understand to what extent a weather event in the game is comparable to a specific business cycle situation. This would be too granular, and also unnecessary for understanding the comparability of the logic of the interplay between the aspects. The game is designed such that the situational environment is comparable to the real world: students face conflicts of interest, costs and benefits, whose logic is similar to that in the real world. From these aspects, one can draw insights from the experimental evidence to suggestions for how the real world would likely operate.

In contrast to students playing a fictitious game, governments face more complex and existential decisions. The real world is more intertwined and thicker than the abstract game could possibly mirror. Moreover, the scope of actions is much larger compared to the few simple options in the game. Besides the re-election threat that creates a particular responsibility of governments towards the electorate, and also drives government decisions, governments also face pressure from financial markets. Concerns about debt sustainability, long-term growth, and competitiveness sometimes stand in conflict with social policy and responsibility for society's wellbeing. Accordingly, the government considers a much bigger picture and, hence, decisions operate in a more complex environment, with potentially unforeseen adverse effects. In that sense, having students decide about whether to harvest one apple more or less seems simplistic in comparison.

However, the principles that both students and governments face when taking a decision are comparable: the participant also faces potential losses and reduced revenue due to weather events, and has a 'budget constraint' for decisions. Similar to governments, participants find their scope of action restricted by the harvesting rule. Moreover, they have the opportunity to advance the economy by investing in items of two types that mirror 'stimulus'- and 'prevention'-oriented policies. They interact with others in order to cultivate the common-pool resource together, and they can sanction misbehaviour. Accordingly, the questions that this thesis addresses to the experimental results are similar to the ones on the real-world SGP: How do external events (business cycle hypothesis), country characteristics (nationality, size), perspectives towards 'stimulus'- and 'prevention'-oriented policies, or co-operative behaviour on a common-pool resource affect compliance?

There is, however, one important noteworthy difference: while students harvest more from shared trees in an economic downturn, they harvest equivalently less in better economic circumstances. For a government, experience for most euro area member states shows that maintaining primary surpluses is difficult as political pressure might rise to use this fiscal space for tax reductions or policy goals (Wierts 2008: 99ff., Afonso and Hauptmeier 2009). This relates to the salience of taxes and state action to the electorate, also in competition with other political parties. Such pressure is, of course, difficult to replicate in an experiment, and this is also the reason why this thesis has not formulated a respective hypothesis.

6.5.3 Choices: preferences are different than ideology

It is crucial to understand that with the experimental results, this study directly measures the intention of policy choices. In the real world, however, there is uncertainty. Choices for a specific policy (i.e., item) are based on beliefs about their effect, but are subject to risk and uncertainty about the transmission success. Instead, the experiment measures preferences for the 'result' of a policy, and there are no transmission problems for the intended effect. This is a useful abstraction to reveal preferences and not to overlay them with assessments of risks and uncertainty. This allows the experimenter to measure such preferences properly.

This direct impact of items is also important to properly measure the assessment of a choice for or against compliance in an economic downturn. The potential repercussions of choosing compliance with the deficit rule in the experiment are rather abstract in comparison to the real world. However, they are comparably harsh in their respective contexts. In the experiment, harsh weather halves the available apples for harvest, representing an approximation to the loss of revenue due to a shock in the real world. On the expenditure side, investment decisions in the experiment are free of path dependency (except for the water irrigation system), and can be made flexibly. However, some items break during shocks and the participant might like to fix them. This mirrors the limited flexibility that governments usually face for the expenditure side, in particular for social policy, and also the increased expenditure pressure during an economic downturn due to automatic stabilisers. Even though in the experiment there is no potentially disappointed electorate, when the farmer spends less energy points during an economic downturn, not investing at all in items leads to a loss of economic strength.

This is also the basis for why one can assume to infer 'ideology' from measuring preferences to some extent. In particular, this thesis argues that 'beliefs' about the 'right' strategy to cultivate the field are revealed over the course of the game. For that reason, the study uses the aggregate variables for the analyses. The thesis argues that such a manifestation of 'beliefs' reflects 'economic policy ideology' in the narrow sense put forward in the hypothesis. From the thick theoretical background outlined in Chapter 3, the thesis uses the basic difference of 'stimulus'- versus 'prevention'-oriented policies. This thesis argues that they likely indicate the

respective thick ideological background and the respective perspective towards constraining fiscal policies with the SGP. As argued above, the experimental design reflects an economic setting with an SGP-like rule that constrains the use of the common-pool resource comparable to the actual SGP case. Within this design, investing in 'stimulus'- as opposed to 'prevention'-oriented items is equivalent to such economic policy decisions. Accordingly, the hypotheses set out in Section 5.3.4 relate such policy preferences to SGP-rule compliance behaviour more equivalently than in the real world. This justifies that the experiment is an abstraction of the real-world SGP case. However, to what extent can one draw from the experimental results back to the real-world case? As argued above, this thesis intends to provide insights into the logic of the interplay of specific variables for rule compliance. Accordingly, one can infer that policy preferences that manifest over the course of the game, towards 'stimulus'- as opposed to 'prevention'-oriented items, result in less compliance with the SGP-like extraction rule. As this supports the hypothesis, that has been drawn from the real world, one can conclude that this relationship might also be observed in the real world, all else equal.

Nonetheless, there is a typical caveat for experiments. Participants might understand rule compliance as a principle rather than an expression of 'economic policy ideology'. This risk exists even though one would not expect a significant number of participants to consider it that way. Therefore, the instructions are formulated in such a way as to avoid this as much as possible. There is no particular strategy prescribed that needs to be followed. Even the voting rules are formulated so that the participant can decide whether or not to follow the rule without feeling social pressure for compliance or for breaching. Moreover, embedding the decisions for compliance in a thematic context gives meaning to the decision. Hence, compliance as a principle would not necessarily make economic sense in the context of the game, and therefore needs to fit the individual strategy. Nonetheless, there might be different perceptions of the usefulness of rule adherence per se to which participants might feel inclined, which might be affected by their cultural background ('nationality'), the field of study, age and living experiences ('capital', 'international'). The experimental analyses seek to control for such aspects. Likewise, the analyses seek to control for participants' attitudes towards 'risk aversion' or long-term planning ('sustainable') in order to disentangle such aspects from measuring 'economic policy ideology'.

Similarly, the analyses control for the self-positioning on the left-right political spectrum.

6.5.4 Country characteristics: 'size' and 'nationality' are different in the experiment than in the real world

In the real world, the size of a member state entails a number of aspects that might affect rule adherence, such as economic capacity, the economic importance for others, and the political power through voting weights. In the experiment, one is able to disentangle these aspects. Firstly, economic strength as measured in GDP per capita does not necessarily coincide with the size of an economy (see, for instance, the case of Luxembourg). Secondly, the economic importance for others directly relates to the size of an economy. Thirdly, even though the political power in the Council depends on the population's size, for member states, this also coincides with the economy's size.

Firstly, in the experiment, 'size' does not reflect economic strength per se. Small farmers are not poorer; they just have less capital than large farmers. The economic scope of action is the same relative to the number of trees. Therefore, in the experiment, the thesis does not assess the role of economic strength per se, but seeks to control for it and allow all participants to have similar relative starting positions so that their choices are not unequally constrained by economic strength at the beginning. Instead, 'size' reflects the ability to engage in the diversification of the economy. Accordingly, this might make large farmers less vulnerable to shocks in cases where they cultivate their trees with different strategies each. As a result, one might see fewer breaches based on economic necessity. Secondly, the economic importance for others is comparable to the real world. In the experiment, this transmits via the number of trees that are shared between two neighbours.

Thirdly, in the experiment, 'size' relates to political power, i.e., voting weights, and shows whether political power as such affects compliance. For instance, whether small states do not engage with the rules because they do not see themselves as being able to get their preferences passed anyway, or whether large states assume responsibility towards the common-pool resource. The application of political power using 'reciprocal voting' can be measured directly and separately from 'size'. In the

real world, there is little available data to check the role of 'reciprocal voting'. The literature assumes that this allows especially large states to leverage their economic importance for others to avoid sanctions. In the experiment, the thesis is able to disentangle both aspects.

Similarly, 'nationality' in the real world refers to a large variety of political, societal and economic aspects: the policies of the current government, the society's (and electorate's) preferences for policies or political parties due to collective norms and traditions, perspectives stemming from historical imprint and collective experiences, particularities of the economy's structure, economic development, the welfare state and regulation. This provides quite a unique set, from which the 'North-South' hypothesis derives a level of economic capacity, or political will, to follow the SGP rules.

In the experiment, this thesis is able to disentangle some of these aspects. 'Nationality' only reflects a common cultural background. The variable seeks to control for any other aspects that might relate to being brought up in the respective society, and might affect the style of play. The game design abstracts from the thick economic setting in the different member states, and instead provides all participants with an equal economic situation. Moreover, the study measures economic policy ideology directly. Finding any differences across nationalities that go beyond differences in the variables that measure economic policy ideology would show that there is apparently something more to cultural differences than this experiment can measure. It is particularly intriguing that there are indications that the role of economic policy ideology for overharvesting seem to play out differently across size and nationality. One would need to study a four-way interaction, for which, however, the dataset has too few participants. This study was not expecting this beforehand, and therefore did not target more participants. Therefore, this would be a fruitful avenue for future research.

6.5.5 Rejecting the null hypothesis: reflections on rationality

The experiment allows this thesis to study participants' behaviour under laboratory conditions. Rational choice assumptions expect that participants make similar decisions under similar conditions. In particular, the idea is that there would be one

‘best’ strategy that reflects rational considerations of costs and benefits. The results show that this is not the case. Participants do behave differently in comparable situations. This reveals that the choices are not solely driven by rational considerations, but also by convictions about what action is the ‘right’ one to take. The analyses find that participants do engage in different strategies, which reflect in different typologies and also differently in rule adherence. In particular, the analyses find that rule adherence does not reflect a pure apples-maximising strategy, but that harvesting and voting have to make sense for advancing the economy. This is particularly visible in the harvesting behaviour in recovery and boom rounds. Harvesting from shared trees shows that participants do indeed take more apples in bust rounds, but also reduce harvesting from shared trees in better economic circumstances. This is a remarkable finding that goes against the general rational perspective, which assumes that harvesting from shared areas follows a maximisation strategy to extract as much as possible, and trying to avoid sanctions. Instead, reflected through economic preferences, most participants do not understand harvesting as a goal in itself, but as a means to advance their economy.

In the experiment, rationality is much easier to grasp and to measure than in the real world. In the real world, as Chapter 2 has demonstrated, cases are complex and it is difficult to extract whether a government decision was ‘rational’ given the respective conditions. Additionally, in the real world, there are more than just the one point of view for which one can assess rationality. As was said previously, the relevant decisions for complying with the deficit and the voting rules are not done by a single actor in a member state, but by a collective actor, and they are not necessarily done by the same actor for both rules. Accordingly, the decision by a collective actor does not necessarily reflect rationality, but a compromise between its members. Preferences are also formed with respect to different audiences, such as the electorate, the own political party, businesses and the financial market, for instance. All this adds complexity, and, therefore, one cannot conclude that, in the real world, one could also reject the null hypothesis. Instead, one can say that given that the results show that basic preferences such as ‘stimulus’- versus ‘prevention’-orientation as measured in the experiment have an effect on rule adherence, one could also assume that politicians’ assessment of costs and benefits could be coloured by their ‘economic policy ideology’.

Applying the null hypothesis to governments, it states that governments would follow similar fiscal policies in similar economic settings, and variety is only to be explained by the different situations their economies are in at a specific point in time. The experiment gives indication that this might not hold in the real world. The section on the compliance track record already shows that governments decide differently in comparable business cycle situations. In particular, the financial, economic and sovereign debt crisis revealed that governments indeed do not decide on similar policy options in comparable business cycle situations. However, one could not conclude that in fact such situations can be considered comparable without taking the general economic conditions of a member state into account. For this, more research would be needed to disentangle the potential aspects at play. Moreover, the null hypothesis states that governments would 'free-ride' on the other member states' provision of the common-pool resource, and exploit comfortable borrowing opportunities. There is some indication that such considerations could indeed be at play, in particular for member states that enjoyed windfall gains at the beginning of their euro area membership.

As regards the voting rule, the null hypothesis would basically predict that voting only happens to avoid getting sanctioned. Otherwise, in particular in such a large group of 19 member states, casting the vote may seem like too much effort, and harmful for potential future non-punishment alliances. From the results of the experiment, one can see that voting correctly indeed takes place despite such 'rational' concerns. Moreover, again with a view to the sovereign debt crisis, one could assume that member states indeed engage in voting correctly, as compliance by other member states becomes a salient issue for them. While this holds true especially for large member states that enjoy relatively higher political power, the crisis has shown that even small member states, such as Finland, have been rather vocal in pushing for rule compliance. Still today, one can see that medium-sized member states, such as Austria and the Netherlands, form a 'frugal' coalition together with Denmark and Sweden. While this might express perspectives based on 'economic policy ideology', it also makes sense from a rational perspective, as this makes the issue salient to them. This shows that future research could engage in disentangling the effects at play, for which the experiment can only yield indications for the real world.

6.5.6 Summary

The results indicate that some factors, as discussed in the analyses above, play a role for the logic of an SGP-like rule. Of course, one cannot conclude that the results explain real-world SGP cases. While the SGP-like setting and the conflicts of interest for choices of the participants are an abstraction of the real world, this thesis argues that they reflect the real-world logic sufficiently accurately for the purpose of an experiment. In this framework, measuring preferences are coherent and manifest over the course of the game. Given the economic context of the game, these manifestations can be interpreted as 'beliefs' about what strategy the participant perceives as most meaningful to cultivate and govern the field. The simplification of 'stimulus'- versus 'prevention'-oriented policies is drawn from the thick theoretical background, presented in Chapter 3. Moreover, it is associated with being more or less likely to breach the SGP-rules, respectively. Analogously, the experiment reflects such investment options, that build into a strategy over the course of the game, and shows that they similarly link to rule adherence as suggested by the hypothesis. As a result, one can infer, from how these two variables interact, general conclusions on how they might likely interact in the real world, as well. Finally, the experiment allows us to disentangle notions of 'size' and 'nationality', which in the real-world case are intertwined with several aspects and difficult to measure separately. Finally, as usual with experiments, one has to be cautious about rejecting the null hypothesis in the real world analogously to the experimental results. All in all, in spite of that, the experimental design has been drawn from the real world with the intention of most accurately abstracting from the real world with the caveat of neglecting some important actors, such as the financial markets, the ECB or domestic political actors, one can draw general considerations from the logic of the interplay between variables from the experimental results to the real-world case.

7. Conclusion

7.1 Summary

While the SGP has been said to be dead countless times, it remains central to the EMU institutional design. The SGP has been reformed several times, and the European Commission just recently launched yet another reform process. The literature review presented in this thesis concludes that the disciplining effect of the SGP has been flawed despite these reforms, especially in the decade before the sovereign debt crisis. This concerns compliance with both the deficit and the voting rule, i.e., what this thesis calls compliance at the national and European levels. In an environment of persistent breaches, why would a member state comply at all? In particular, as breaches at either level do not consistently result in consequences, why would a member state engage in potentially huge efforts to make the state budget comply with the deficit rule? Similarly, why would a member state vote for sanctioning their peers and, thereby, risk losing their support in the future? As a result, why should a member state not breach the rules? With these research questions in mind, this thesis studies the incentives that the SGP sets for compliance and breaches.

The literature provides several explanations for breaches from the deficit rule stemming from a lack of 'capacity': economic necessity in a downturn, different political and economic capacities due to 'size', and general dispositions of a member state as regards their inclination to follow the rules ('North-South' divide). In contrast, complying with the voting rule is said to hinge on the member states taking 'ownership' of the rules. In particular, the SGP's institutional design is considered to undermine peer pressure by facilitating circumvention of the rules with reciprocal voting to form non-punishment alliances (so-called 'sinners' solidarity'). With this, national- and European-level compliance are linked, making enforcement even more difficult. Linking the European integration literature to the regular implementation of the SGP, this thesis investigates the role of economic policy ideology as a complementary piece to explain adherence to the SGP rules.

Using an experimental approach, this thesis provides three notable insights into the real-world SGP: Firstly, and in particular, in the absence of successful extrinsic disciplining, fiscal rules shall be based on a common economic thought. The analyses

find that economic policy ideology, according to the narrow definition used in this thesis, affects compliance, and this provides more substance to the meaning of 'political will'. Secondly, the experiment also shows that insufficient economic capacities can cause breaches. Thirdly, there seems to be strong support for the role of reciprocal voting, which highlights the disincentives that the SGP rules set for compliance. Accordingly, 'political will' could manifest in both economic policy ideology and the institutional opportunity to leverage one's own interests, thus circumventing the threat of punishment.

While the limitations of the experimental method in social sciences are widely discussed, the experimental results can inform the real-world SGP. In particular, it gives indications about certain aspects, and how they impede the SGP-like rule's smooth functioning. The idea of the experiment is to understand what, in principle, drives adherence to an SGP-like rule for governing a common-pool resource. It is not to explain member states' compliance track record. One can learn from the experiment that economic preferences associated with economic policy ideologies affect compliance with an SGP-like rule. This indicates that one could also expect such a relationship to reveal in the real world because it relates to the understanding of such a rule as being 'appropriate'.

7.1.1 Economic policy ideology drives compliance

Inspired by the intense debate between the two economic policy ideology sides that dominated during the sovereign debt crisis, the thesis investigates their role for regular SGP implementation. The analyses find that economic policy ideology drives taking ownership, and affects compliance at both levels. Given that the literature diagnoses insufficient extrinsic incentives stemming from the SGP, the results suggest intrinsic incentives as a complementary explanation for compliance at both levels. In particular, economic policy ideology drives the understanding of the SGP rules as 'appropriate' to follow. Perceiving the SGP rules as part of the 'right' economic policy encourages compliance, even in difficult circumstances.

The experimental evidence shows that the investment strategy as linked to the economic policy ideology affects compliance, especially in an economic downturn. In line with the hypothesis, the results show that 'stimulus'-oriented participants breach

the national-level rule more often, in particular in an economic downturn, and that 'prevention'-oriented participants comply with the European-level rule more often during economically good times. Speaking in terms of the experiment, 'stimulus'-oriented participants overharvest more often in general, and in particular during 'bust' and 'recovery' rounds, and 'prevention'-oriented participants vote correctly more often during 'boom' rounds. For 'voting incorrectly', the effect is more surprising. For 'bust' and 'recovery' rounds, the effect is inverse to the hypothesis, and 'prevention'-type participants are more likely to vote incorrectly. However, incorrect voting also reflects the intention to make the application of the rule stricter rather than less strict. With this reading, the hypothesis would in fact be supported. This would benefit from further research. Additionally, the analyses find little support for the hypotheses regarding 'risk averse' or 'sustainable' cultivation behaviour, which underlines that rule adherence does not seem to be significantly driven by personal psychological traits or attitudes; and instead considerations based on the economic setting of the game are more likely to drive compliance to the rules.

As insight into the functioning of the real-world SGP, this thesis indicates that economic policy ideology does not, in principle, stand in the way of compliance at the national level. The experimental evidence shows that in principle, adherents of both manage to comply, even though less so in an economic downturn. For these cases, though, there is already an exception clause in the SGP. Hence, in principle, one could assume that this is also possible in the real world, however, with the caveat that the budget balance needs to equivalently improve in economically better times, which easily happened in the experiment, but which is less likely in the real world. In that regard, the establishment of other institutions that satisfy demands for a stabilisation capacity could help, and possibly also increase, compliance with the voting rule. Section 7.2 discusses the conclusions to draw for EMU's institutional design and Sections 7.3 and 7.4 discuss the contribution to the academic debate and areas for future research.

7.1.2 Compliance also hinges on economic capacity

The experimental results show that the business cycle carries significant explanatory power as an indicator of economic necessity. Overall, compliance appears more difficult in severe economic times, such as 'bust' and ' recovery' rounds, and less so

in economic boom times. Moreover, the business cycle appears to mediate the role of economic policy ideology as discussed above. This shows that it is not just a difficult economic situation, as such, but also the ideological perspective on it that drives rule behaviour in the experiment.

This supports the literature's hypothesis on the role of economic necessity in a downturn for public finances. As part of 'capacity' issues, it defines the (perceived) need to react to economic circumstances, as well as the fiscal capacity to engage in it. This finding can be meaningfully transposed to the real world as the experiment mirrors an economic setting and business cycle developments that induce periods of scarcity. Decreasing tax revenues, and increasing expenditure for automatic stabilisers and additional policy measures, compress the budget balance and make a breach of the deficit rule more likely. Moreover, the ideological perspective on the need and means to react to a downturn appear to impact compliance, as discussed above. This is a new finding in studies on the SGP, and confirms that the relevant theme in the European integration literature is also relevant for regular SGP implementation.

7.1.3 The SGP voting rules impede compliance

The experimental evidence shows that reciprocal voting plays a significant role for rule compliance. In the experiment, the option to seal deals has been used by participants that did not comply with the deficit-like rule to circumvent sanctions. This supports the literature's hypothesis. This finding can also inform the functioning of the real-world SGP as the decisive aspects of the setting are comparable: the experimental design mirrors the SGP rules and the collective action problem in the euro area by simulating the overlapping of a common-pool resource with private goods, and a non-hierarchical self-governance mechanism. By allowing strategic alliances to form among non-compliant members, the voting rules in the Council provide disincentives for compliance. Moreover, a reluctant enforcement by the Council reduces the threat of sanctions, and members might have internalised this already when discussing the budget at national level. While in the experiment, these disincentives reflect in breaches, one can only suspect this to be the case in the real world: despite the lack of data, this has long been suspected in the literature, and there have been several suggestions on how to reduce the role of alliances (as discussed in Section 2.3), such

as the exclusion of non-compliant members from voting on other members, or decreasing the role of the Council to open or proceed in the sanctioning procedure.

Moreover, the results for the variable 'size' are intriguing, reflecting the role of weighted votes for voting rule compliance. In the experiment, large participants are significantly more likely to vote incorrectly, as expected by the hypothesis. In contrast, they do not seem to anticipate their political power on the voting stage when making overharvesting decisions, since 'size' does not affect 'overharvesting'. Or put another way, there seem to be other more salient reasons for deciding to overharvest than a purely strategic perspective.

The experiment provides some relevant insights into the 'size'-hypothesis for the real world. The results support the literature's hypothesis that small member states might find it more difficult to get their preferences passed in the Council and therefore revert to compliance. While the literature suggests a number of aspects that might relate to 'size', I break these aspects down. The experiment controls for economic strength per se, and instead creates a homogenous definition of 'size' of the economy: it reflects the ability to engage in diversification, one own' economic importance for other members, and political power, i.e., voting weights. Accordingly, for the real world, this indicates that compliance with the deficit rule might not relate to the size of a member state in terms of voting weights. It rather indicates that it relates to what the size usually entails for states in the euro area according to the literature (see Sections 2.3.2.1.3 and 3.3). This relates to the economic development, the economy's trade position and related experience with global competition, and to the domestic growth model. In conclusion, one could say that the experiment's results support the literature's assessment that weighted votes might impede the voting rule as large states find it easier to exert their power in the Council and vote against the rules.

7.1.4 No indication for a 'North'-'South divide

In contrast to popular opinion, the results do not show a general 'North-South' divide. In the experiment, participants from northern countries are not, in general, more likely to comply with the rules than participants from southern countries. This non-effect might be due to the specific design of the experiment, which disentangles and controls for the economic aspects of this differentiation. With this reading, the results suggest

that in addition to such economic aspects and economic preferences as reflected with the 'economic policy ideology' hypothesis, the remaining cultural differences, which students also carry as participants, are not a detrimental factor for compliance with the SGP in principle. Additionally, the non-effect could be due to the specific sample of university students. It might be that university students are a more homogenous group as regards their perspectives on economic preferences and co-operation than their nationality would be a differentiating factor. With this, one could conclude that SGP compliance is likely not driven by cultural aspects, but by economic considerations only.

7.1.5 Principled compliance and co-operative behaviour need further research

The experimental results show that principled breaches happen rather than principled compliance. Speaking in terms of the experiment, adherence with either rule affects the other: 'voting correctly' slightly reduces the likelihood to overharvest, and 'overharvesting' has a strong positive effect on 'voting incorrectly'. This is in line with the hypothesis, but the size of the effect is surprisingly weak and strong, respectively. While having overharvested seems to be a strong reason for engaging in 'voting incorrectly', having shown overall compliance with the voting rule over the course of the game does not seem to be a decisive factor when making overharvesting decisions. There seem to be other more salient reasons to decide against overharvesting than sticking to the principle of complying to the rules as reflected in the voting behaviour.

While there is no data to verify the interaction of both rules' compliance for the real world, the experiment provides insights into the logic of the interaction. It seems plausible to assume such logic also for the real world – in particular, with a view to voting incorrectly while breaching the deficit rule that could rather result from other factors such as those named above. Despite the effort to vote – instead of simply abstaining – and despite the potential reputational repercussions of forming non-punishment alliances in the future, compliance with the voting rules happens frequently in the experiment. I assumed that voting behaviour would signal general agreement or disagreement with the rules, while abstaining represents the default option with the least costs. In the real world, this assumption might be even more pronounced given the associated costs of voting. Instead, voting correctly would indeed signal agreement with the rules. In addition, in the experiment, breaches of the

rules are intentional, and result from choices. Drawing from the experiments' results, one might overestimate the role of intentionality for compliance with the budget rule in the real world. Therefore, the interaction between the compliance of both rules in the real world could benefit from further research.

Finally, the experimental results suggest that contributing to the common-pool resource seems to lead to the conviction that this also gives the 'right' to extract more than the rule allows, instead of rendering compliance an intrinsic motivation. These findings are rather difficult to transpose to the real world. One could suggest that a member state with a strong economy also contributes to maintaining the common-pool resource of monetary stability. Even more so, with a strong and sustainable economic development, it would contribute to 'increasing' the common-pool resource. In the experiment, this, in turn, fuelled the idea that there is an 'equivalent' right to also (temporarily) extracting more from the common-pool resource than the rules allow. For the literature on common-pool resources in general, this is an insightful finding. So far, the suggestion was the other way around: those who contribute to the common good would also care about its rules. However, in the literature on the SGP, this motivation has not yet gained prominence. It might be rather implicit if it exists at all. It seems worth devoting more research to this other kind of motivation, which will also be discussed in more detail in Section 7.3.2.

7.1.6 Validity of the experimental design and SGP reforms

Finally, the experimental design and accordingly, the central arguments of the thesis, remain unaffected by previous SGP reforms. The experimental design reflects the basic elements of the SGP functioning and is, therefore, unaffected by the numerous reforms that have happened since its establishment. Despite these reforms, the fundamental logic of constraining fiscal space did not change, neither did the voting rules, in principle. Establishing reverse qualified majority voting in 2011 only relates to pecuniary sanctions, while all other steps of the sanctioning procedure, in particular, the opening of an EDP, remain with qualified majority voting. Moreover, the reforms introduced an acceleration of the process with shorter deadlines, but did not change the political power in the voting weights. While voting weights have, in principle, been abolished, the new 'double majority' (55% of member states and 65%, population)

still reflects weights according to the size of the population. Therefore, one can learn from this SGP-like experiment for the current real-world SGP.

Moreover, the experiment also reflects the different notions of compliance at the national and the European level. While breaching the national-level rule triggers a voting procedure, not adhering to the European-level rule, i.e. not voting in the spirit of the SGP rules, does not trigger any adverse consequences in the experiment. This reflects the different legal quality of the 'rules': whereas the deficit and debt rules are binding EU law, voting in the Council hinges on political sovereignty of the members and is not constrained by a legally binding rule. Adherence is not even made publicly available as there are no voting records of the respective items in the Council meetings. Reforms of the SGP have not yet changed this. Nonetheless, despite the fact that there is no binding legal rule, one could argue that there is a political notion of compliance as member states are requested to vote and to consider the spirit of the rules for their voting assessment. The two different notions of compliance, the legally binding and the political one, are reflected in the experimental design and the analyses of this thesis.

7.2 Insights for economic and fiscal governance in EMU

7.2.1 General political debate surrounding the Stability and Growth Pact

For assessing areas for SGP reform, this thesis helps with the distinction between extrinsic and intrinsic motivation. Moreover, this thesis conceptualises that intrinsic motivation relates to economic policy ideology, and the experimental results show that it becomes especially salient in times of economic necessity.

Starting with the first aspect, what could extrinsic incentives possibly look like in practice? Well-known means to encourage desired behaviour extrinsically among individuals are prices, public attention and judicial restrictions of sovereignty. Scholars have already made several suggestions in that regard for the SGP to increase financial market discipline (for instance, by abolishing the 0% risk weight of sovereign debt, Bénassy-Quéré et al. 2018), by reducing the complexity of the fiscal rules to allow easy tracking by the electorate (cf. EU Independent Fiscal Institutions 2021), or by removing voting rights for members in an excessive deficit procedure (cf. Irlenbusch and Sutter 2006), to name just a few.

Given the delicate nature of fiscal policy for sovereignty, it seems most likely that the SGP will remain a non-hierarchical self-governing framework in the near future. It seems unlikely that calls for more automatic sanctions would be implemented, as this would further reduce the member states' scope of action. Automaticity refers to getting decisions done with less influence of (breaching) member states in the Council, for instance, through introducing reverse qualified majority voting for all steps of the sanctioning procedure, removing voting rights for states in an EDP, or through abolishing weighted votes (cf. Irlenbusch and Sutter 2006). The latter proposal would increase the political power of small states relative to large- and medium-sized ones. These proposals seek to increase the threat of punishment and the costs of a breach. The experiment's results indicate that small participants vote significantly less incorrectly. This indicates that such reforms could increase compliance.

Likewise, it seems unlikely that budget rules would become stricter: the key lesson from the sovereign debt crisis is that trying to impose conditionality against the political will of the member state's government risks losing legitimacy and breaking-up the euro area. While the debate at the time overlooked that it was also about all other member states' sovereignty and tax payers' money, the overall objective now is to avoid conditionality, and keep it for ESM programmes only as the last resort ('ultima ratio'). While the intention is that conditionality also serves a disciplining purpose with a strong deterrent effect, one could plausibly assume such an effect for small- and medium-sized member states only, as the ESM's capacity is considered too small for large member-states.

Additionally, even though one of the current reform debate's goals is to simplify the rules so as to increase transparency and easy tracking by the public, the European Commission's increasing role as a 'political' actor might occasionally run counter to this goal. As guardian of the treaties, the Commission shall ensure a transparent surveillance process of SGP implementation. However, at the beginning of his presidency, Juncker announced that he would head a 'political Commission' (Kassim and Laffan 2019). This might fall short of ensuring a diligent surveillance, as has been the case, for instance, for France in the aftermath of the sovereign debt crisis (Juncker 2016). Moreover, increasing transparency and involving the public more, so as to increase the re-election threat, does not necessarily mean that the public would

be in favour of SGP compliance. It might also be that a government could especially rely on the public for engaging in a breach – as was the case in Greece when negotiating compliance with the support programme in the first half of 2015.

While the above aspects refer to 'negative' incentives, i.e., incentives through the threat of costs in case of non-compliance, most recently there is some indication of positive extrinsic incentives with the NGEU, i.e., incentives through rewards in case of compliance. After assessing compliance with the reforms in the Recovery and Resilience Plans (RRP), funds are released to the member states. These funds are to be used for stimulus policies and investments, on the one hand, and financing structural reforms on the other. This approach marks a turning point in the incentive structure of EMU institutions. While there is not likely to be a change in the SGP's negative extrinsic incentives, this change in positive extrinsic incentives could facilitate compliance for some governments, especially those facing challenges regarding economic necessity. It might not help for those cases where a breach is due to political will.

As regards the second aspect named above, the role of political will for intrinsic compliance, the pandemic has triggered an unprecedented shift towards more state intervention in the euro area. By building a common stimulus mechanism such as the NGEU, the solution for the ideological debate about the SGP was sought outside the SGP. This could help to make the SGP rules leaner and have other institutions providing means for stimulus in times of crisis. The idea could be to keep the 'ordoliberal' SGP, and to have another institution that satisfies state-interventionist demands. Such an approach would be supported by this thesis' experimental results that highlight the role of economic necessity for breaches at both the national and European levels.

For such a compromise to be sustained in the long-term, it would need a fruitful exchange between both camps. And for that to work, it would need a proper understanding of each other's convictions. The goal is not to share the other's economic policy ideology, but to understand and address the related concerns about liability and scope of action. For the ordoliberals, rules provide stability, especially in times of recovery, to facilitate economic actors' investments. For the state-interventionists, discretion provides stability in times of crises with the state providing

bold action. The former warns about the institutional design providing false incentives for moral hazard, and the latter warn about limitations to sovereignty for fiscal policy. In the end, solidarity is not a one-way-street, as those who provide support also trust that the recipient uses it responsibly so as to avoid ever more support in the same direction.

In conclusion, the core interest for a functioning SGP is not to define what economic policy strategy is optimal, but what is acceptable within the scope of the SGP rules. By opening doors to institutionalised 'solidarity' through a stimulus function such as the NGEU, it seems fair to equivalently improve 'liability' with, for instance, increasing market discipline to better reflect borrowing costs stemming from domestic politics that are not associated with sustainable fiscal policies. Establishing positive as well as negative extrinsic incentives outside the SGP could 'free' compliance from dependence on 'economic policy ideology' to some extent. These extrinsic incentives could complement the SGP, and support its goal even in the absence of intrinsic motivation. Nonetheless, as this thesis has shown, European-level compliance seems to inherently need intrinsic motivation as extrinsic motivations are largely absent unless there is a significant majority that exerts peer pressure. For such a non-hierarchical case, there does not seem to be an 'ideological neutral' incentive design. With respect to political feasibility (Schelkle 2017: 2), this avenue could pave the way for an interim institutional balance rather than representing an economically 'flawed' institutional design, as long as the monetary union is not complemented with a political union (also see Featherstone 2012: 25, de Grauwe 2006). This, however, hinges on, as a common denominator, taking 'ownership', at least to avoid the very high debt burden in some member states endangering the union.

7.2.2 Current debate on the reform of the EU fiscal framework

After intense discussions among member states in the Council, in April 2023, the European Commission presented legislative proposals to reform the fiscal framework (within the economic governance review). The key objectives of the reform are to reduce high debt levels effectively and to account for future reform and investment challenges to improve competitiveness and to secure wealth. Since the last reforms, high debt levels have not been brought down significantly, and fiscal policy has not been sufficiently supporting growth in a counter-cyclical manner (for a discussion of

the economic impact of the rules, see Lane 2021). The pandemic and the energy prices crisis have increased debt levels even more. The starting point of the reform debate was that the European Commission engaged in an in-depth review of the rules, as is prescribed in the current rules, in 2020. They found that, besides the complexity of the rules, insufficient implementation of the rules could also be attributed to a lack of ownership.

Central to the reform is to operationalise the 3% deficit and the 60% debt thresholds, prescribed in primary law, with new rules. The suggestion is to replace the current structural deficit and the debt reduction rules with an expenditure rule based on a debt sustainability assessment. From the debt sustainability assessment, the Commission would derive a technical trajectory of a fiscal adjustment path with yearly expenditure targets for a four-year horizon. This path would ensure that debt is kept below 60% of GDP or is on a downward path, and that the deficit level is below 3% of GDP at the end of the horizon. Member states would present medium-term fiscal-structural plans, that also include major reforms and investment plans, which might justify revising the fiscal path and extending the plan up to seven years. Additionally, high debt countries would be required to reduce the debt level by 0.5% of GDP per year. The Commission would assess and the Council would endorse the final plans. For member states that have deficit and debt levels below the thresholds, the technical trajectory would serve as an informative fiscal path to secure that both indicators remain below the thresholds. Moreover, the Council could activate an escape clause to suspend requirements to cater for exceptional circumstances, outside the scope of a government, that occur either EU or euro area wide, or are specific to a member state. The process would be integrated into the European Semester, and annual progress reports would be assessed by the Commission.

Completion of the legislative procedure is envisaged for the end of 2023, before the European Parliament elections are held in spring 2024. Observers consider this to be an ambitious timeline, in particular because member states' positions still differ on a number of essential issues. The following section discusses the main aspects of the suggested reform, with a view to the thesis' results, linking it to previous reform debates.

7.2.2.1 The economic rationale and the role of economic policy ideology

The economic rationale of the fundamental rules, the 3% deficit rule and the 60 % debt rule, has been subject to intense debate ever since the establishment of the SGP (also see Section 2.3.1). While the original pact understood the 3% threshold as a sufficiently high upper limit, while maintaining a balanced budget in 'normal times', the first reform in 2005 already took account of cyclical developments that might allow divergence from the rules in specific circumstances and for a period of time. While contestants perceived this as insufficient for granting the necessary scope of discretion during an economic downturn, in later reforms, the aim was to operationalise the rules with a structural balance indicator, so as to identify the part of the budget that is under the scope of the government and reflects policy actions, while allowing usual business cycle fluctuations. The idea was also to allow more country-specific flexibility, and to take account of differing economic developments. At this point, adherents of simple and harmonised rules, as in the original pact, requested more automaticity in the enforcement of insufficient implementation so that discretion would not expand too much. As a result, the rules became more and more complex, and both sides criticised the rules either for their continuous lack of economic viability or for the lack of rigour in enforcement.

At first sight, the suggested new rules seem to strike a great deal in reducing previous contestation of the economic rationale of the SGP. While the current rules have been accused of not allowing different responses to an economic downturn, the suggested reform seems to allow more autonomy for governments to follow their preferred economic policy strategy. Whereas the current rules seek to define fiscal space as the part of the budget balance that can be attributed to be within the scope of a government, irrespective of business cycle fluctuations, the new rules suggest linking fiscal space to debt sustainability. This has the potential to 'free' the main indicator of compliance with the rules from economic policy ideology to some extent. Linking fiscal adjustment effort to debt sustainability clearly shows a government what is at stake, and how their political agenda might put this at risk in the medium-term. So far, governments could easily contest the structural deficit targets, also based on measurement issues (see discussion in the section below) and defend breaches, arguing that their policy priorities require this to address a particular domestic situation. The new rules have the potential to reduce the conflict of interest between stimulating

growth (both in the short- and medium-term) and fiscal sustainability as the new fiscal-structural plans allow governments to present an economic narrative of the original situation and their policies' impact, also identifying policy priorities, reforms and investment that are considered to be costly in the short-term but growth-enhancing or wealth-securing in the medium-term (for reasons on considering investment needs see for instance Wolff 2021). This goes far beyond the current assessment of numerical rules only. With this, the new rules satisfy calls from both predominant economic policy ideologies to engage in potentially costly structural reforms (ordoliberal perspectives) and public investments (Keynesian perspectives). Both could be viable economic responses to a crisis or challenge, justified by the respective member state, assessed by the Commission and endorsed by the Council.

In light of very high debt levels in several countries, anchoring the rules to a debt sustainability assessment, instead of simply looking at the debt level, would also make the rules more plausible, and encourage governments to engage in growth-enhancing policies, as well. As Beetsma (2022) shows, several high-debt countries have not managed to reduce their debt levels since the financial, economic and sovereign debt crisis, and they rose even more during the pandemic, making the 60% reference value unachievable for decades to come as the fiscal effort would be immense. The current debt rule prescribes that the difference between the debt level and the reference value should be reduced by 1/20th per year. The debt sustainability assessment, instead, takes into account much broader fiscal, economic and financial conditions and not just the business cycle. This allows governments to target all these aspects to achieve debt reduction, and not necessarily give fiscal aspects priority. For instance, as could be seen in 2022 for many euro area member states, debt reduction also happens very efficiently through the denominator – though, in 2022, due to inflation driving nominal GDP. This again has the potential to reduce the perceived conflict of interest between fiscal and economic policy, and could even encourage governments to engage in growth-enhancing reforms and investments.

Moreover, differentiating member states based on their debt sustainability assessment could allow fairer treatment of countries, which would make the rules more plausible and reduce contestation. Lane (2021) notes: "The limitation of the current framework is well illustrated by the adjustment requirements for 2019 that resulted in Latvia, a country with debt well below the 60 percent of GDP debt reference level of

the Treaty, facing the same adjustment requirements as countries with debt ratios at or above 100 percent of GDP". The current rules differentiate countries only based on the debt level (being above or below the 60% threshold) and the level of the output gap indicating the business cycle. A matrix defines the required fiscal adjustment to reach sustainable fiscal positions. In comparison, debt sustainability assessments seem to provide a more meaningful differentiation.

The new rules can also, to a lesser degree, be accused of not being economically viable. With the focus on debt sustainability, the rules target the most prominent indicator during the sovereign debt crisis for funding costs, divergence of government bond yields within the euro area, and credit ratings. Using this indicator not only shows governments what is at stake, but is also a strong signal to financial markets and the public, who could, in turn, better engage in scrutinising government action.

On closer examination, though, the reform risks allowing imprudent fiscal policies and debt levels to persist for longer, in-transparent assessments of fiscal requirements, inconsistent treatment of member states due to the many assumptions to make in the assessment, and the politicisation of the debt sustainability assessment tool. Firstly, there is the risk of allowing imprudent fiscal policies to persist because macroeconomic policies could be given greater importance, or even priority, in specific instances, based on the economic narrative in the fiscal-structural plans. This is not the original idea of the SGP, which seeks to secure fiscal sustainability, to avoid an impact of public spending on inflation, to secure debt sustainability, and, with this, to complement the no bail-out clause. Previous reforms did not change the fundamental logic of constraining fiscal space for this purpose. With the suggested new rules, there could be a risk of randomness and a greater dependence on the political priorities of the Commission and the majority in the Council (see discussion below in Section 7.2.2.3): when assessing or revising fiscal-structural plans, several other than fiscal aspects could gain salience.

While the reforms during the sovereign debt crisis also underlined liability of government actions for fiscal positions, the new rules suggest that many different conditions, challenges and plans could justify backloading of fiscal effort. The experience with a few countries that resisted the Commission's calls for more fiscal

prudence over the past decade, seem to have fuelled embracing their arguments and embedding them in the national fiscal-structural plans. The Commission might hope that this allows them to scrutinise these arguments much better than is currently possible, but it could also risk justifying imprudent policies. Moreover, debates about the economic rationale might pop up again in discussions about the viability of a fiscal plan. This could happen in case the plan could also be revised on a yearly basis, like the medium-term fiscal strategies presented in Spring for the European Semester, or in the event of yearly updates of the debt sustainability assessment showing significant differences. In this regard, one could also argue in favour of revisions based on budget execution data, reform success or delay, significantly revised GDP data (denominator effect), or better placings at financial markets (changes in interest rates or ratings). Such instances risk undermining the viability of a plan if it remained unchanged – but also the viability of the rules in case a plan would change several times.

Secondly, while the conflict between member states might reduce with respect to economic policy ideology and the question if a government had to consolidate public spending, it might increase with respect to the question on how much consolidation would be necessary based on the debt sustainability assessment. Given that the assessment hinges on many assumptions and could be sensitive to slight changes, the debt sustainability assessment risks becoming subject to political discretion. Originally, the debt sustainability assessment was seen as a technical tool used for country surveillance, also by many international institutions, national debt management agencies, or financial market actors. For the European Commission, it might be difficult to continue having yearly updates of the debt sustainability assessments for surveillance purposes, but not using them to scrutinise governments' fiscal effort and to assess compliance with the SGP.

Moreover, several observers to the political debate see a risk to using the debt sustainability assessment as a 'fig leaf', and to delay debt reduction towards the 60% threshold (for a discussion of reasons for debt reduction, see Beetsma 2022). Additionally, in the past decade, interest payments on public debt have compressed significantly, which make the 60% reference value appear meaningless to reach in order to gain debt sustainability. Low interest payments also secure debt sustainability at much higher debt levels. For some countries, debt reduction is essential, for instance those who received financial support during the sovereign debt crisis have decade-long

obligations towards official counterparties in the euro area. This covers actual market assessment of some parts of the debt stock, creating roll-over risks in the future. These interdependencies could potentially burden monetary policy or economic policy co-operation among member states during economically bad times.

7.2.2.2 Fiscal targets and economic capacity

The intention of the new rules is to better address the requirements of business cycle developments for fiscal targets with simpler rules that are less prone to measurement issues. This thesis suggests that business cycle flexibility is warranted to facilitate compliance. Instead of looking at the structural deficit, trying to assess the current point in time of the business cycle and what part of the fiscal position is within the scope of a government, the reform suggests having an expenditure rule as a single indicator. Previous reforms tried to define business cycle developments more accurately, but using the output gap as level and the related structural balance as fiscal indicator entails some measurement issues (for a discussion, see Benalal et al. 2022). This also makes the rule prone to contestation in terms of economic rationale. Moreover, a particular level of the structural balance has to be defined as adequate to represent state intervention in normal times (so-called medium-term objectives, and fiscal requirements to reach this dependent on the output gap level and the debt level in the so-called 'matrix'). While this was the operationalisation of the SGP's spirit of 'close to balance or in surplus', it was criticised for being blind to addressing future challenges, for instance. The reform instead suggests using as single indicator net primary expenditure growth, which should not grow faster than average potential growth in the medium-term. The indicator was first introduced with the reforms in 2011, and excludes interest payments, expenditure financed by EU funds, one-off policy measures and cyclical spending and includes changes in revenue policies, such as tax rate increases. The indicator also considers potential output but as a growth variable over several years, which is considered to be less volatile (see also Lane 2021). This is likely to reduce discussions about the economic rationale with respect to business cycle developments or growth-enhancing policies as the concept is that current state expenditure should, in principle, not grow more than the economy. Otherwise, state demand might risk crowding out private demand, or too high tax

levels could jeopardise economic growth. The idea is to place state engagement in relation to economic growth.

While this indicator has several advantages, it also risks getting eroded as interest payments and investments would not be considered. Firstly, interest payments represent an important disciplining effect for fiscal prudence stemming from financial markets. Moreover, there is heterogeneity across countries and by removing it, there is less of an incentive for governments to engage in reducing these costs. Secondly, in cases where many policy objectives (such as addressing the energy transition, climate change, digitalisation, or defence expenditure) would require increased investments, a significant part of the budget would not be considered under the rule. With this, the actual deficit risks being much higher.

7.2.2.3 Enforcement in the Council and reciprocal voting

With the suggested new rules becoming more encompassing of other policy areas, monitoring and enforcement in the Council risks becoming more difficult. Firstly, effective evaluation of the fiscal situation in other member states would need more resources in finance ministries. Secondly, with the Council becoming involved in more assessments during regular country monitoring, i.e., in the preventive arm of the SGP, on top of those regarding enforcement, i.e., in the corrective arm, there would also be more opportunities for reciprocal voting, which as this thesis has shown, is likely to increase breaches. Thirdly, despite the intention of making the rules simpler and easier to understand for the public, scrutiny by the electorate and financial markets might risk focusing only on those countries where the debt sustainability assessment shows a high risk.

Firstly, monitoring in the Council would become more difficult because tracking compliance or a risk of breach risks of not anymore being meaningfully assessed by other member states. They would need to be able to replicate the Commission's debt sustainability assessment and the technical trajectory, to assess the national fiscal-structural plans, the plausibility of assumed costs and gains from reforms and investments, the economic forecast, and to understand deviations leading to a revised fiscal path, including possible extensions. This increases member states' reliance on the Commission's assessment. This would further increase if the bilateral

discussions between a member state and the Commission happened before the Commission issues their technical trajectory.

Secondly, the potential risk of reciprocal voting to abate enforcement in the Council and, with this, compliance with the national-level rules in the first place could increase if the Council has the opportunity to decide on more instances of the surveillance procedure. As the Commission proposes, the Council would have to endorse the member states' fiscal paths and plans, extensions and revisions. This risks creating new opportunities for reciprocal voting already before the fiscal path would start. Accordingly, there is a risk of watering-down fiscal effort at the outset because it makes the entire process dependent on a favourable Council majority. However, as has been documented in the past, the Council might also consist of a majority that votes against the SGP spirit. Similar to previous reforms, these political economy aspects have remained untouched in the suggested reform.

This underpins that enforcement in the Council has a very different legal quality than compliance with the national-level rules: sovereign member states in the Council simply cannot be forced to enforce the rules. However, the role of political economy aspects for incentivising compliance seems to have been missed in the Commission's reform proposal. The only suggestion in that respect is to decide on sanctions earlier in the excessive deficit procedure in order to reduce their size and have them accumulate every six months up to a maximum of 0.5% of GDP. While the intention to make sanctions more applicable seems good, it risks avoiding the opening of a procedure altogether, also in view of loss of reciprocal voting power due to increased automaticity. To avoid this, the Commission suggests that an excessive deficit procedure would open automatically if a member state with a severe debt challenge deviated from the agreed fiscal path. While this would likely improve enforcement, it already seems as harsh as conditionality under financial support programmes during the sovereign debt crisis. For this to work, the debt sustainability assessment would have to be shielded from political discretion.

Thirdly, despite the intention of making the rules simpler and easier to understand and to conceive for the public, public attention might narrow to the cases with a severe deficit or debt challenge. In that moment, public scrutiny might come too late to avoid such a situation. Moreover, focusing on debt sustainability, which

risks becoming unstable only in very difficult situations, and less on fiscal sustainability, which could already be discussed in an economic downturn, might catch public debates by surprise and too late. A meaningful and timely deliberation would not be possible, as such a situation might have to be addressed rather quickly. In order to improve public scrutiny and to inform financial market actors, the involvement of independent fiscal intuitions during several steps of the process could help. They could serve as a politically-independent body to assess the complex issues and to inform the public. This could shift the attention back to the regular monitoring, at least for the interested public and involved actors.

7.2.2.4 Expectations about ownership and enforcement

This thesis has argued that with increasing ownership, both national-level and European-level compliance could increase due to intrinsic motivation. This would also make enforcement in the Council much easier. The suggested reform shares this idea. With the aim of increasing national ownership and, at the same time, reducing the impact of economic policy ideology on compliance as demonstrated in this thesis, so as to ensure that both prominent perspectives could, in principle, comply to the rules, the Commission suggested replacing the current one-size-fits-all approach with a common framework and country-specific fiscal requirements to meet the same goal, i.e., debt sustainability. However, with the aim of embedding fiscal requirements in an economic narrative that is endorsed by a member state's domestic sphere, there is a risk of shifting the focus away from fiscal sustainability. This is particularly necessary in an incomplete monetary union, with monetary policy being reliant on member states to work on reducing heterogeneity, so as to facilitate monetary transmission and not to unevenly impact domestic inflation developments as, for instance, the most recent developments have made clear.

The suggested reform is likely to improve national ownership, in particular, with four changes. Firstly, it allows member states to design a fiscal-structural plan that includes their policy priorities, enabling governments to embed fiscal requirements into the member state's particular economic situation, reform and investment needs. However, national ownership might be jeopardised when a new government comes into power and would want to revise the plan to include their political priorities. Just because the plan spans across many policy areas does not mean

that a new government, keen to include their own policy priorities, would not want to make revisions. Secondly, more autonomy is given to member states where both deficit and debt levels are below the 3% and 60% thresholds. However, this entails the risk that potential breaches are not detected and addressed sufficiently in advance. Thirdly, bilateral discussions between a member state, and the Commission on the plan and the fiscal path, increases the possibility of the member state explaining their position and, on the other hand, understanding and potentially embracing the spirit of the rules that the Commission is supposed to convey. However, this risks increasing opacity for the Council to understand and scrutinise the results. Finally, involving domestic actors, such as independent fiscal institutions that are considered to be defenders of the rules, could also facilitate domestic implementation with leveraging the public debate. However, given the many new objectives of the rules, not just fiscal sustainability, but also economic development, reforms and investments, it might also make it more difficult for an independent fiscal institution to make an assessment which is not prone to taking a (political) stance on assumptions.

From these potential improvements in national ownership, what could we expect for compliance and enforcement? For a member state, it might become easier to justify a potential breach or backloading of fiscal effort. So far, member states could say that there are many fiscal indicators, with measurement issues and budget execution slippages, or that potential reforms, or other domestic obligations justify delaying fiscal effort. The reform suggests having one single fiscal indicator, reduced measurement issues, and potentially increased national ownership, but allows significantly more room to justify 'surprises' for instance with inaccurately quantifying costs of reforms or investments, or unexpected delays of their implementation. In particular, the latter might happen in case of very long fiscal paths because seven years would be longer than a usual legislative term. By allowing political discretion to define country-specific fiscal paths, the idea seems to be to embed any potential argumentation that a member state might bring up inside the rule to be able to assess and scrutinise it. However, the loss of common benchmarks risks resulting in a backloading of effective fiscal consolidation and debt reduction. A more feasible debt rule could have also been designed in the form of a common benchmark (see, for instance, Lane 2021). Wary of the arguments presented above on the Commission's willingness to defend strict rules amidst their own political agenda, and

on the Council's ability or willingness to enforce the rules, the reform also offers an opportunity to better differentiate between technical assessment and political discretion. If both the technical trajectory and the member states' suggested paths were published, the new rules could help to make enforcement more transparent than is currently possible.

7.3 Academic contribution

This thesis contributes to the academic debate by providing a different theoretical, methodological and empirical perspective: 1) theoretical through considering the SGP as a mechanism to govern a common-pool resource, and with economic policy ideology to endorse the rules as 'appropriate'; 2) methodological through a novel type of common-pool-resource experiment, which I call the 'cultivation game', to study the SGP, and which could also serve to study other cases of co-operation; and 3) empirical through testing competing explanations to better understand how the SGP functions. With this, this thesis informs academic debates in the field of the SGP, co-operation of the euro area member states in the Council, the governance of common-pool resources, in general, and experimental methodology. Moreover, the results open up avenues for future research, notably in the field of economic policy co-ordination in EMU.

7.3.1 Theoretical contribution

This thesis contributes to the literature that conceptualises the SGP as a governance mechanism to govern a common-pool resource that faces collective action problems (cf. Schelkle 2017). I add that the common-pool resource overlaps with private goods in order to reflect the setting more accurately in the experimental design. This also allows us to link both national-level and European-level compliance. With this, the SGP does not only face a collective action problem in the Council, but also moral hazard for compliance at both intertwined levels. In addition to the literature, this thesis shows that in a non-hierarchical setting, self-governance needs a common understanding of the rules as being 'appropriate'. This argument stems from the common-pool resource literature (Ostrom et al. 1994: 329), and helps to provide a

content-related argument for the more process-related argument of a collective action problem in the Council decision-making, used in the literature so far.

I conceptualise the SGP as a mechanism to govern the common-pool resource 'monetary stability of the Euro'. Understanding monetary stability as relating to a limited amount of 'Euros' given a specific level of economic output, its use is rival and non-exclusive (once entered to the euro area). I use this conceptualisation for two reasons. Firstly, it helps to illustrate interactions between members based on the required compliance at the two intertwined levels, i.e., the deficit and the voting rule. Secondly, it reflects that in a non-hierarchical and self-governance setting, considering the rules as 'appropriate' facilitates members to follow them (Ostrom et al. 1994: 329). Accordingly, this framework emphasises the role of intrinsic incentives for compliance at the community level. Otherwise, there would not be enough members in favour of implementing the rules, and to exert peer pressure.

Additionally, I consider that the common-pool resource overlaps with the private goods of each of the members, i.e., the domestic economies. Contributing to the common-pool resource is, hence, not just based on preferences towards the common-pool resource, but also put in the context of its use and applicability for the domestic economy. At times, members might perceive it as a trade-off to engage in maintaining either the common or the private good, while for others, both go hand in hand. Member states might perceive a 'trade-off' between preserving monetary stability and stimulating economic growth when deciding about public finances and deficits. This trade-off has two aspects for the conceptual design: first, 'using' stable money when engaging in public borrowing is not an end in itself, but follows domestic policy strategies. The money would be used for the domestic economy. Hence, there might be a domestic limit to borrowing –irrespective of framing it as a common-pool resource in a monetary union – depending on the level of taxes acceptable to the electorate. Secondly, at times, there might be economic necessity to focus on stabilising, or stimulating, the domestic economy at the expense of 'over-using' the common-pool resource. Accordingly, non-compliance might not just happen because of a disregard for the rules, but also because the domestic good is perceived as more salient at a specific point in time. As regards European-level compliance, the Council's (non-) decisions might reveal that the goal of monetary stability is valued less than the aim to foster economic growth. There might be times where breaching the deficit rule

might actually be favoured by a majority of member states, although this would result in a double-breach of the SGP's rules, i.e., a breach at the national and at the European levels.

By introducing the factor of economic policy ideology to the debate about SGP compliance, I link the day-to-day implementation of the rules to the major theme in the EMU integration debate (McNamara 1998, Dyson and Featherstone 1999, Brunnermeier et al. 2016). I argue that, to some extent, the distinction of when and who might perceive the trade-off or not, depends on the economic policy ideology shared among the government, or reflected in the electorate as a deeply rooted tradition in the member state. The SGP was built in an ordoliberal spirit to promote a 'stability culture' (term used by Artis and Winkler 1997: 23). Monetary stability was seen as a major prerequisite for private investment and economic growth and was, therefore, put at the centre of monetary (and also fiscal) policy. By constraining public deficits, the government's share of the total 'amount' of stable money would be limited. The very idea of constraining fiscal policies goes against the Keynesian perception of the role of the government for economic growth. This perspective objects to the neoclassical understanding that there is a specific 'amount' of stable money. In contrast, inflation depends far more on other aspects than just government spending and thus, could not be guaranteed by restricting government borrowing. Government spending, and also deficit spending, are considered indispensable to restore economic growth in a crisis in order to break the 'savings paradox'. When private actors save money in a crisis, it is the government's obligation to spend and to stimulate production and, thereby, private consumption.

From the theoretical perspectives, I extract a tangible and measurable line of conflict: preferences towards 'stimulus'- or 'prevention'-oriented policies. In the experiment, these policy goals are equivalent to investing in 'stimulus'- and 'prevention'-oriented items. Accordingly, the main hypothesis of the thesis is that governments that prefer economic policies targeted at stimulus, as opposed to prevention, are more likely to run higher deficits and are more reluctant to sanction breaching governments. The argument is that for member states that do not share the ordoliberal spirit of the SGP, there are no such intrinsic incentives for compliance, which, hence, would rest only on extrinsic incentives. As long as a rule does not evolve

‘intrinsically’ at the domestic level, but is ‘imposed’ externally, a breach seems more likely than compliance.

As regards the continuous implementation of the SGP, the results also contribute to the EU compliance literature by providing a qualifying element to what has been attributed as ‘misfit’ between domestic conditions and EU law (cf. Börzel et al. 2010, Mastenbroek and Kaeding 2006): economic policy ideology. Additionally, the results suggest that aspects relating to country-specific economic capacities might also be attributed to ‘misfit’. The non-effect of ‘size’ for ‘overharvesting’ could indicate that, in the real world, other aspects of an economy’s size, such as the economic development, the trade position or the domestic growth model, might explain differences in the compliance track record. This opens avenues for future research.

This theoretical perspective could also be applied to other areas of co-operation that rest on a common understanding of the goals and, accordingly, the adequate means, and that require continuous implementation. As long as there is no common understanding, it would be difficult to overcome capacity issues in a non-hierarchical self-governance. Anecdotal evidence shows this, for example, when addressing the pandemic or the recent energy prices crisis. Shifting the focus on health support and the economic impact of the pandemic, member states managed to agree on the NextGenerationEU programme. The debates leading up to the decision have shown that it was essential to understand the issue such that no legacy of ‘unreasonable’ policy choices could be attributed to the current misery. The crisis affected regions seemingly arbitrarily, and every member state could in principle be affected. Moreover, there was a clear understanding that fragmentation in the economic recovery needs to be avoided as it has the potential to delay the recovery of the entire euro area. It would be intriguing to study whether this indeed made a common response possible, which also necessitates continuous implementation as repayment of additional debt is scheduled until 2058.

7.3.2 Methodological contribution

The thesis contributes to the literature on common-pool resources in general, and suggests a novel ‘cultivation game’ alongside the traditional ‘investment game’ (see

Ostrom et al. 1994, Deadman et al. 2000) and 'request game' (Suleiman and Rapoport 1988, Budescu et al. 1995). Introducing private goods that overlap with the common-pool resource provides an analytical framework to study common-pool resources that are used as a means and not as an end in itself. In traditional games, the common-pool resource is usually a natural resource, such as fisheries, and the preferences only relate to extracting 'the more fish, the better'. Instead, having overlapping private goods for which the common-pool resource would be used, and on which the common-pool resource also depends to some extent, allows a more granular perspective on preferences for extraction. This induces intrinsically motivated limitations to extraction. With this analytical framework, the experimental design could also be used to study other common-pool resource contexts.

With this analytical framework, this thesis also adds to the experimental literature in economics and political science (cf. Druckman et al. 2012a, Roth 2015, Duffy 2015, Palfrey 2015). Experiments are a common tool in economics and political science, used to study preferences and institutions. In economics, experimental designs are usually highly abstract, and stress the mechanism of an interaction most parsimoniously. In psychology, experimental designs are rather sophisticated and entail a relevant context. In political science, experiments rather use vignette studies included in surveys. This experiment provides a context for decisions. For the purpose of this study, providing a context for an economic decision has the advantage of allowing the experimenter to infer proper interpretations of actions. Usually in economics, experiments seek to make an abstraction from the context in order to reveal the basic mechanism, and not to have any potentially intervening aspects other than the numerical payoffs. However, this makes interpretations rather abstract, and makes it difficult to link the results to the real-world case of a study. Instead, this thesis underscores the critical importance of a context to meaningfully locate decisions for compliance. With this, the game is about more than just personal preferences for rule adherence, and more appropriately links to the real-world SGP case. This shows that such experimental designs could be used more broadly, also in economics and political science.

The core allegory is an apple orchard that is cultivated by 19 farmers together over several rounds. The participant cultivates own (private goods) and shared trees (common-pool resources) that symbolise 'the economy'. The trees grow apples, some

of which participants need for their own purposes. Weather shocks, which symbolise economic shocks ('business cycle' hypothesis), harm the trees. The participant can harvest apples ('tax the economy'), which yield energy points. These can be used, in turn, to harvest more apples, and to invest in cultivating items. The participant can choose between different items that either focus on increasing the growth of new apples, or on protecting the tree against future shocks ('economic policy ideology' hypothesis). Moreover, there are items to reveal risk aversion or big gain seeking. In this way, the participant 'governs' the trees, and looks after their prosperity. Harvesting more than half of the apples on a shared tree (simplified deficit criterion) corresponds to a breach, and participants can vote on sanctions (simplified voting rule). To avoid sanctions, the participant can seal deals with other farmers ('reciprocal voting' hypothesis). The experiment is conducted as a computer-based game. Participants play it individually and separately against 18 'computer'-farmers (programmed and fixed set of responses). Participants get randomly assigned to one of three different sizes of farmers ('size' hypothesis) and are sampled from students from four euro-area member states ('North-South' hypothesis). 327 university students from France, Portugal, Germany and Greece played the game. There is no specific goal given to participants so as not to bias their choices while playing. They are told to find their own strategy to cultivate the field.

The experiment puts 'economic policy ideology' at the centre of the analysis, controlling for economic downturn. Moreover, it controls for country-specific conditions, and offers a 'laboratory' setting of 'equal' opportunities where participants can reveal cultivation and co-operation preferences. Furthermore, the experiment allows this thesis to evaluate the actual role of co-operative behaviour on the one hand and strategic interactions for 'reciprocal voting' on the other for breaching, and to disentangle them from 'economic policy ideology'. This allows the thesis to study whether in the same situations and under the same conditions, participants would behave in the same way, or whether the results would show systematic differences based on the named aspects.

For the purpose of this study, an experiment is especially suited as it provides a controlled setting to test competing hypotheses. The strength of the method also lies in the limitations of other methods. Conducting quantitative analyses would not have been possible, as there is no official record of the member states' voting behaviour in

the ECOFIN Council. With this, an essential part of SGP compliance could not be examined, namely strategic considerations for compliance. The main challenge of using interviews would be to make an abstraction from the complex cases in order to find generalisable conclusions on what aspects drive rule adherence, in particular. Euro area member states are different in many respects, and it would be challenging to find cases to compare to each other, and to disentangle the competing explanatory variables that this thesis intents to study. Moreover, relying on public officials' or politicians' accounts might underscore the role of reciprocal voting or intentional breach of the rules as they might refrain from openly admitting that other considerations were more salient than SGP rule compliance at the time.

Moreover, the analyses presented in the thesis could be amended with further research to substantiate several findings. On the one hand, the dataset of this study could be used to engage in more granular studies. The dataset offers a wide range of further analyses, for instance, on the rule compliance typology, the treatment of a common-pool resource, and co-operation in general. A multinomial regression could yield insights into the explanatory factors for the combination of both levels' compliance. The analysis could show which aspects explain rather principled behaviour in contrast to, for instance, ignorant behaviour. This would provide a more granular picture of the role of the hypotheses for different types of adherence with both rules. Further analyses could also provide more insights into the treatment of the common-pool resource, and co-operation in general. One could study 'co-operation' as a dependent variable to see what factors explain co-operative versus exploitative behaviour. Moreover, it would be interesting to see whether behaviour is different towards some farmers as opposed to others. One could also pursue more granular analyses on voting correctly. One could study what type of participants sanction what type of computer-farmers, relating to the pre-programmed harvesting and voting types. Moreover, voting on direct neighbours could yield insights into the role of the large neighbour serving as an economic anchor, or whether proximity plays a role for rule engagement. Additionally, one could study the role of reputation, and accordingly whether participants refrain from abstaining, but correctly vote green instead. Participants might consider their reputation more if their performance track record is public information.

On the other hand, the case selection could be broadened to include more countries, or a larger sample of the population. I select the member states from which students are invited to participate based on the independent variable 'economic policy ideology'. The debate between ordoliberal thought and the state-interventionist perspective has been accompanying the SGP ever since its establishment. At the political level, Germany and France have taken either side based on their traditions and collective ideology towards money, the state and economic policy. During the financial, economic and sovereign debt crisis, Portugal and Greece joined on either side of the debate with their perspective towards the financial assistance programmes. One could broaden the selection of member states, and invite students also from other member states that supported either perspective, such as Italy or the Netherlands, to replicate the experiment. Alternatively, one could engage in finding larger or more specified samples of the population, even though this requires either more effort to find comparable samples, or more financial effort to pay laboratories. For instance, the thesis tried to invite finance ministry officials, but the response rate was very low. One could also invite officials of other public institutions, such as fiscal councils, or national central banks, or Members of Parliament, to collect the political perspective.

The experimental method also comes with a number of limitations to generalisability and, accordingly, applicability to a specific real-world case (see detailed discussion in Section 6.5), and this shows avenues for future research using other methods. The usual drawback of experiments is that experimental evidence cannot explain real-world cases. It is simply not observational evidence. In contrast to governments, who face responsibility towards the electorate, the students play in a fictitious environment without any consequences for the real world. Moreover, the experiment requires a level of abstraction that necessarily reduces accuracy relative to the real-world case. Instead, the experiment is a tool to provide insights into the logic of an SGP-like setting, and how the relevant aspects impact on rule compliance in principle. With this, the analyses show indications on aspects that might drive rule adherence also towards the actual SGP. These would benefit from further studies using real-world evidence, as discussed below.

7.3.3 Empirical contribution

This thesis provides a systematic approach to testing competing hypotheses that were put forward by the literature and had not yet been tested together, in particular because data on European-level compliance is missing. This thesis shows the role of several hypotheses for the functioning of the real-world SGP. In particular, the thesis supports the role of economic necessity stemming from the business cycle in line with the literature (cf. Hansen 2015). While I find a general effect of economic necessity driving breaches, I also find that such situations are differently addressed, depending on the prevalent economic policy ideology (see discussion above in Section 7.1). Furthermore, the study provides insights into the role of a member state's size (Buti and Pench 2004), and the experimental design provides more clarity on the different aspects attached to this factor.

Moreover, the results support the role of reciprocal voting, which the literature on the Council suggests shapes Council decisions (see discussion in Nugent and Paterson 2010). This thesis finds that the SGP-like voting rules facilitate reciprocal voting. Moreover, it shows that coalition building, along the lines of shared economic policy ideology, could be plausible – while a general 'North-South' cleavage cannot be supported based on the experimental results. The results show that ideology affects voting behaviour, based on which coalitions could be formed in the Council either to avoid or to enforce sanctions. With this, this thesis adds to the literature that suggests that coalition building in the Council depends on the policy area (Thomson et al. 2004, Hayes-Renshaw and Wallace 2006: 259-279, Thomson 2011).

The results of this thesis open a number of avenues for future research on economic policy co-operation in EMU. On the one hand, the experimental design could be amended so as to study institutional change and the effect of introducing different institutional designs on incentivising compliance. In particular, this could address the question: Under which conditions does reciprocal voting lose its relevance? On the other hand, this thesis opens up questions that could be investigated using other methods as will be discussed in the next Section.

There are several amendments to the experiment that could yield further insights into the incentivising effect of the institutional design of the SGP: one could introduce different versions of the game (to study different sets of rules or institutions),

have groups play the game (instead of individuals and computer-farmers), or include another actor, such as the European Commission, the financial market, or a 'bail-out'-mechanism. In order to study dynamics of self-regulation, one could substitute the computer-farmers with actual participants, and have a group of 19 people play the game. One would not be able to compare individual actions within a group, but groups against each other. This could yield insights into different versions with different institutional designs.

More substantial amendments could include the establishment of other institutions that satisfy demands for a stabilisation capacity. With this, one could study whether even with such a new institutional design of EMU, reciprocal voting would continue to impede compliance. One could also include an option for a mutual 'bail-out' to provide additional apples, with or without interest, to be paid back and with or without political conditionality (for instance, to prescribe investments in the following round). Accordingly, one could study how voting behaviour changes (for the former) and how harvesting behaviour changes (for the latter). One could also test reform proposals and their incentivising effects on actors' behaviour. Moreover, as mechanisms in other policy areas were recently set up, which are similar to the SGP (Macroeconomic Imbalances Procedure, Rule of Law Framework), the experimental design could also be applied to study compliance with those mechanisms.

7.4 Future research

Despite the fact that compliance with the Stability and Growth Pact has already been studied extensively, there are a couple of aspects that would benefit from more research. This thesis opens up questions that could be investigated using other methods, in particular, to investigate actual cases of the SGP compliance track record to understand when economic necessity or instead political will drove a breach. In particular, it is interesting to understand compliance despite a difficult economic situation. Moreover, it would be interesting to understand how exactly economic policy ideology transmits to government behaviour, and how deeply rooted or anchored it is in a member state's society.

Firstly, it would be interesting to understand what exactly explains the compliance track record and what aspects drive European-level compliance. The

former is challenging because of the many different hypotheses that are at play. This would require case studies. The results of this thesis could yield an orientation for the selection of hypotheses. The latter is challenging because there are no official voting records, and negotiations happen predominantly on an informal level. This would require confidential interviews with relevant politicians or relevant public officials. In particular, it would be interesting to further investigate this thesis' indication that the role of economic policy ideology for compliance seems to play out differently across sizes and member states.

In order to understand the track record, country case studies could provide insights into the relevant hypotheses at play, to understand national-level compliance. In particular, case studies could reveal what exactly drives the role of 'size' in the real-world case: economic development, such as catching-up or established economies, the economy's trade position and related experience with global competition, or the domestic growth model. The national-level compliance perspective could be complemented by interviews for examining European-level compliance. However, the aforementioned caveats (see Section 4.2.2) apply given the sensitivity of the topic. An interesting case could be Italy in 2018 and 2019, when the government openly announced breaching the deficit rule, and the European Commission had to take a delicate negotiation stance, as it did not intend to escalate the sanctioning procedure.

For compliance at the European level, this thesis shows that 'reciprocal voting' potentially plays a significant role, but might be too short to explain large parts. In particular, it seems that a shared ideology also explains voting, to some extent. Given that a Council decision marks the last step of the bargaining and the Council president usually seeks to have solved all disagreement at this stage, it might be fruitful to have research focus on the negotiation stage where disagreements are debated. However, collecting information might be cumbersome as negotiations happen predominantly informally. This could also be used for studying coalition building in the Council, and provide substantial evidence for the currently ongoing SGP reform process.

Finally, it seems fruitful to engage in further research to understand the role of 'nationality'. While the experiment disentangled the role of 'nationality' from economic strength, political power and 'economic policy ideology', it remained rather superficial on this aspect. The results also provide an indication that the named

variables play a different role among member states. Future research could delve into linking cultural and ethnographic research to their role for economic policy ideology and, in particular, the role of the state for the economy. This seems especially warranted as the experiment risks measuring psychological traits instead of deeply-rooted cultural aspects. Moreover, this could be contrasted with data on governments' economic policy ideology to see whether, in specific cases, it is rather the deeply rooted cultural background or the temporary party ideology that explains compliance. Such research could complement the above-mentioned case studies, to verify the compliance track record on the one hand, and studies on European-level compliance and coalition building in the Council on the other.

With this, future research could inform and stimulate the political debate about reforming the EMU institutional design. Even though in the past, political decisions have fallen short of the academics' (economically) optimal suggestions for the institutional design, their insights have been valued, requested and also used as arguments by politicians in negotiations. Intriguing from an academic perspective is to look out for politically feasible solutions to a multi-faceted economic, political and cultural problem. This thesis opens up future avenues to study how to reconcile the two prevalent economic policy ideologies. Recently, politics have tried to embark on a new route that could appease the ideological impact on the SGP by establishing other institutions that seek to anchor a stabilisation function with the ESM and the NGEU. It remains to be seen whether this strategy bears fruit, or just increases the costs for the 'frugal' member states that already see their options to incentivise the 'state interventionists' to contribute to the common-pool resource decrease. On the other side, the member states adhering to the 'state interventionist' perspective finally see their scope of action increasing, however, at relatively low speed given the economic challenges ahead. This shows that the future rules need to account for both aspects studied in this thesis: economic necessity, which might be leveraged by capacity issues, and political will, which might be leveraged by ownership issues. The common understanding among members is crucial, and the handling of the sovereign debt crisis has put this in the spotlight: it is not only important to understand each other's economic policy perspective (and to try to find a compromise that both sides agree with), but also to understand and trust each other, in general, and not consider European co-operation as a game of who can extract the most.