

ÇAĞIN TAN EROĞLU

ON THE POSSIBILITY OF EX-ANTE CONSTRAINED MAXIMIZATION

Bilkent University 2026

ON THE POSSIBILITY OF EX-ANTE
CONSTRAINED MAXIMIZATION

A Master's Thesis

by
ÇAĞIN TAN EROĞLU

Department of
Philosophy
İhsan Doğramacı Bilkent University
Ankara
January 2026

*“The threat of all against all is not a logical necessity;
rather, it is a potentially useful device.”*

- Russel Hardin

ON THE POSSIBILITY OF EX-ANTE CONSTRAINED MAXIMIZATION

The Graduate School of Economics and Social Science
of
İhsan Doğramacı Bilkent University

by

ÇAĞIN TAN EROĞLU

In Partial Fulfillment of the Requirements for the Degree of
MASTER OF ARTS IN PHILOSOPHY

THE DEPARTMENT OF
PHILOSOPHY
İHSAN DOĞRAMACI BİLKENT UNIVERSITY
ANKARA

January 2026

ON THE POSSIBILITY OF EX-ANTE CONSTRAINED MAXIMIZATION

By Çağın Tan Erođlu

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Philosophy.

Simon Wigley
Advisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Philosophy.

Sandy Berkovski
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Philosophy.

Lars Vinx
Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences

Refet S. Gürkaynak
Director

ABSTRACT

ON THE POSSIBILITY OF EX-ANTE CONSTRAINED MAXIMIZATION

Eroğlu, Çağın Tan

M.A. Department of Philosophy

Supervisor: Professor Simon Wigley

January 2026

The problem of collective action has intuitive solutions that do not thoroughly capture the individual moral motivation of cooperating towards long-term collective goals. David Gauthier's theory of constrained maximization presents a plausible moral account that can be incorporated into this debate.

To bridge this gap, I propose ex ante constrained maximization as a rational strategy for acting first towards the production of public goods, where the necessary conditions are satisfied by the absence of a centralized resource allocation scheme. I further present ex post constrained maximization as a rational strategy and a normative contractarian obligation to sustain long-term and larger-scale cooperative outcomes.

Keywords: Collective action, Contractarianism, David Gauthier, Game Theory, Public Goods

ÖZET

SÖZLEŞME ÖNCESİ KISITLANMIŞ MAKSİMİZASYONUN OLANAĞI ÜZERİNE

Erođlu, Çađın Tan

Yüksek Lisans, Felsefe Anabilim Dalı

Tez Danışmanı: Prof. Dr. Simon Wigley

Ocak 2026

Toplu hareket problemine yönelik ileri sürülen ve ilk bakışta makul gözükken çözüm önerileri, bireyin uzun vadeli ortak amaçlar uğruna hareket etmesi için gerekli ahlakî motivasyonu yeterince dikkate almamaktadır. David Gauthier'in öne sürdüğü kısıtlanmış maksimizasyon kuramı, bu tartışmaya katkı sağlayabilecek önemli bir ahlak kuramıdır.

Bu çalışmada mevzubahis iki kavram arasındaki ilişkiyi kurabilmek adına, merkezi kaynak dağıtım organizasyonların yokluğunda kamu mallarının üretimi için gerekli rasyonel bir strateji olarak sözleşme öncesi kısıtlanmış maksimizasyonu ileri sürüyorum. Ek olarak, sözleşme sonrası kısıtlanmış maksimizasyonun hem rasyonel bir strateji hem de sözleşmeci ahlakî bir yükümlülük olarak kabulünün uzun vadeli ve geniş odaklı ortak çıkarların sürdürülebilirliği için gerekli olduğunu iddia ediyorum.

Anahtar Kelimeler: David Gauthier, Kamu Malları, Oyun Kuramı, Sözleşmecilik, Toplu Hareket

ACKNOWLEDGEMENTS

Neither this project nor my now-grounded academic perseverance would stand as entrenched as it is today if it were not for the crucial intellectual nudges of my supervisor, Prof. Simon Wigley. He always guided me in the right direction, and as I walked along that path, I found more questions than answers, and these puzzles became an essential element of my academic disposition. From the day I started my academic journey onwards, I have been and will be grateful to him for every new question I encounter.

I would like to extend my warmest gratitude to Prof. Sandy Berkovski for his kind support during this pursuit. He was always generous with his time when I asked for the opportunity to discuss my arguments, and his support certainly fortified the foundations of this project.

I am wholeheartedly grateful for the kindest support of Prof. Lars Vinx, as his comments on the manuscript shaped how this project stands today. If I am confident in the position and the intuitions provided by my arguments today, his remarks and his meticulous scrutiny have a crucial and fundamental role in this.

I started working on this project in the wake of the recent passing of Prof. David Gauthier. Though I never had a chance to meet him, I feel as if I had; his research has been a constant companion like the Northern Star. There are many ways in which a researcher might be motivated towards a certain area of research, and Prof. Gauthier's philosophy has certainly influenced my current and prospective academic trajectory significantly, not only by his particular arguments but also by his interdisciplinary methodology which incorporates the theory of games and decisions into normative philosophical discussions. As a philosopher-in-training, I viewed my Master's as a pursuit of understanding these dynamics; I would be content if I could plausibly add a modest footnote to the ongoing debates about morality, rationality, and agreements.

Moving forward, I hope to further explore the path that Prof. Gauthier and many other inspiring scholars in this tradition have paved. May his soul rest in peace.

I would like to thank my partner, Selin Küçükoruç, for her continuous encouragements and for never doubting me. For what has been and what is to come, I have consistently felt her love and support, despite the miles and hours between us.

I would like to thank my family for their unconditional support and for believing in me. Their constant faith has been an anchor that allowed me to persevere during this pursuit.

I finally would like to thank Bilkent Philosophy faculty and cohort for thought-provoking discussions, my colleagues in Bilkent Philosophical Society for inviting me for a talk about this rather counter-intuitive realm of philosophy, and many more colleagues in academic philosophy who generously engaged with the earlier drafts of this project and unsparingly raised their worries and objections in various academic gatherings, which has been invaluable for bringing it to completion.

TABLE OF CONTENTS

ABSTRACT	vi
ÖZET.....	vii
ACKNOWLEDGEMENTS	viii
TABLE OF CONTENTS.....	x
TABLE OF FIGURES	xii
INTRODUCTION.....	1
CHAPTER 1.....	7
PUBLIC GOODS, FREE RIDING, AND THE PUZZLE OF OWNERSHIP	7
1.1. What is Free Riding?	7
1.2. Broad Definition of Free Riding.....	9
1.3. Narrow Definition of Free Riding and the Worries Behind it.....	10
1.4. What are Public Goods?	12
1.5. What are Club Goods?.....	14
1.6. The Ownership Puzzle.....	15
1.7. The Problem of Exclusion and the Lighthouse Model.....	17
CHAPTER 2.....	20
CONTRACTS AND THE NATURE OF MORALITY	20
2.1. Why Contractarianism?	20
2.2. Could There Be Proto-Contractual Norms?	22
2.3. The Principle of Minimax Relative Concession.....	24
2.4. Constrained Maximizer and Her Dispositions.....	25
CHAPTER 3.....	29

WHY WOULD ONE BUILD A LIGHTHOUSE? 29

 3.1. The State of Absence 29

 3.2. Acting First as a Rational Strategy 31

 3.3. Signals and Repetition 33

 3.4. The Scope of Cooperation 35

 3.5. Larger Scale Stability 39

CONCLUSIONARY REMARKS 47

REFERENCES..... 49

TABLE OF FIGURES

Figure 1: 2x2 Prisoners' Dilemma	1
Figure 2: 2x2 Chicken Game	8
Figure 3: 2x2 Assurance Game	9
Figure 4: Payoffs for one player in PD	26
Figure 5: Production game between Penny and Ben	32
Figure 6: Respective payoffs illustrating Penny's bargaining power	33
Figure 7: Ben's payoffs in repeated interactions.....	34
Figure 8: Coordination game between Ben and The Network.....	45

INTRODUCTION

According to Elinor Ostrom, who defines *collective action* as a shared goal in which more than one individual's contribution is necessary, the *collective action problem* occurs when rational individuals prefer short-term benefits over long-term coordinated payoffs (Ostrom, 2004). This assumption dates back to the emergence of the debate, when Mancur Olson famously identified the logic of individual inaction on a collective scale, and Russel Hardin followed up by arguing that the *problem* resembles a *Prisoner's Dilemma* model in game theory (Olson, 1965; R. Hardin, 1971).

The Prisoners' Dilemma (hereinafter PD) model is a non-cooperative strategy game in which traditionally two players choose simultaneously either to cooperate or defect (hereinafter C and D, respectively) over a prompt, without direct communication with the other party. It is observed from the structure of a two-player PD game that the dominant strategy for each player is to play a defecting strategy, as defecting when others cooperate (C,D; or D,C) creates the highest possible utility for the player (Figure 1).

Figure 1. 2x2 Prisoners' Dilemma

		Player 2	
		<i>C</i>	<i>D</i>
Player 1	<i>C</i>	3, 3	0, 5
	<i>D</i>	5, 0	1, 1

Both players are therefore tempted to defect as a result of this rational calculation, which transforms the outcome of mutual defection (D,D) into the *Nash equilibrium*, an equilibrium point in which no player can better their position by playing another strategy (Nash, 1950).

The curious case for the PD is that, although both players are rationally tempted to play a defecting strategy, the mutual cooperating outcome (C,C) yields greater payoffs than the mutual defecting outcome (D,D). The argument for the collective action problem, therefore, transforms into this game-theoretical description while assuming that although the mutual cooperating strategy over a shared goal, as Ostrom puts it, yields better payoffs for the individuals in a society; the same individuals have a rational temptation to defect from contributing to these shared goals (Ostrom, 2004).

In the crux of the matter lies the rational calculation that if an individual is the only one defecting (as in both C,D, and D,C outcomes), they would receive the highest possible payoffs (Notice 5,0 and 0,5 outcomes in Figure 1). The worrisome case for this rational agent is the fact that her fellow members of society are also tempted to think alike; hence, the Nash equilibrium, and therefore, the problem.

The emergence of the problem, therefore, necessitates a proposal to solve it; yet, it is curious that proposed solutions to the problem appear to be far more intuitive than the problem itself.

Consider Russel Hardin's follow-up on Olson's description, in which he argues that an account that reduces the problem to the traditional PD game has to assume that the plays resulting with the collective action problem (namely C,D, DC, and D,D) would remain Pareto-inferior, that is, states that are unfavorable to optimal or superior outcomes in which players may be better off, unless altered by external incentives for individuals' long-term motivation to cooperate (R. Hardin, 1971). This implies, conversely, that with such incentives, the Nash equilibrium would differ from a traditional PD setting, and a cooperating strategy would become a Condorcet winner, a strategy that is favorable to all players (R. Hardin, 1971).

Such external forces imply perhaps the most intuitive solution to the collective action problem: the State¹. Thomas Hobbes' theory of state revolves around the idea that self-interested individuals would yield nothing but harm to each other in the state of nature; and this condition of constant war must be abandoned to secure happiness and self-preservation of the individuals engaging in the social contract (Hobbes, 1985 [1651]: 190). Parallely, in *De Cive*, Hobbes argues for the nature of property rights in the state of nature that if everyone appears to enjoy every public good, no one can actually enjoy it (Hobbes, 1998 [1642]; Eggers, 2019: 60). Particularly, the latter conviction appears to be the root of the problem that this paper is interested in, *the tragedy of commons*, formulized and proposed by Garrett Hardin (G. Hardin, 1968).

For the Hobbesian account, the State is not only a solution to the problem; it is *the only* solution (cf. Kavka, 1986; Hampton, 1986). When transformed into an iterated PD game, the game-theoretical description of the Hobbesian problem of cooperation appears to justify the existence of a supra-agential entity concerned with how individuals interact socially and economically with one another (Taylor, 1987; cf. Gauthier, 1986).

It is still possible to argue within the same language that this form of governance, shadowed by the broader puzzle of the collective action problem, remains a Pareto-inferior state compared to potential scenarios that yield better outcomes for the engaging parties. Consider, for instance, an argument by the proponents of a different, perhaps decentralized, model of governance that highlights the *costs of enforcement* as a factor that needs not to be ignored when calculating the payoffs of mutual cooperating strategies in the Hobbesian state (Taylor, 1987:30; Ostrom, 1990). If Michael Taylor is right, then it may not be favorable for parties to outsource their nonexistent trust to the State, as that would mean receiving fewer payoffs than in a state where they are simply trusting each other to cooperate. Such claims, famously contended by Taylor and Ostrom, are a strong position that may potentially undermine the conclusions of the Hobbesian account.

¹ I will use 'the State' with 'capital S' to denote the centralized governing body; to mitigate potential confusions when I refer to 'states' as I am comparing the Pareto-optimality of the outcomes, or when referring to *conditions*.

The puzzle illustrated thus far has a direct influence on the nature of the free-riding problem. Ostrom argues that the problem of free riding is at the heart of any model of the collective action problem, as whenever one is not excluded from the benefits that others are providing, they are tempted not to contribute but to free-ride on the efforts of others (Ostrom, 1990:6).

Any account that I have highlighted regarding the nature of collective action problems (or the free-riding problem, on a narrower scale) focuses on ways and measures to implement plausible strategies that mitigate the unfavorable outcomes of defection being a rational strategy. The motivation behind this stems from the fact that the nature of this problem, if not addressed effectively, is defined as a tragedy (G. Hardin, 1968).

It is intuitive that a rational strategy may not be a moral one, yet the central conviction of this paper is that both concepts must be tied together to the greatest possible extent to mitigate collective action problems. I believe David Gauthier's theory of constrained maximization plausibly bridges the gap between a rational play and an agreeable moral action (Gauthier, 1986). This argument is relevant to the debates of the collective action problem, mainly because if it is possible to argue that any given action is morally unjustified by agreed-upon social norms, then the rational choice appears to be cooperating.

Gauthier's theory deftly answers the question of whether the moral play rationally equilibrates, particularly by comparing the states of plays between a straightforward (self-interest) maximizer and a constrained (self-interest) maximizer, which is a Bayesian player that expects rational compliance as bargaining, and adapts accordingly (Gauthier, 1986).

I am interested in taking Gauthier's theory one step further by proposing a temporal adjustment to situate his theory within the debates surrounding collective action problems. On its face value, the theory of constrained maximization proves that if two Bayesian players interact, they would aim to maximize their self-interest with moral constraints, and this would result in an agreeable equilibrium in which they are both better off than their starting position (Gauthier, 1986).

This proposal, however, only covers the simultaneous exchanges, and the outcome of mutual cooperation could be an agreeable equilibrium if and only if both players adopt CM as a rational strategy; which, as Gauthier puts it, is an adaptive strategy instead of a normative one (Gauthier, 1986). For interactions that are not temporally symmetrical, it is not certain how a Bayesian must act to secure the greatest possible utility for herself.

If she insists on interacting only with other Bayesians, then it would not be entirely rational for her to engage in any activity that yields an *ex ante* contractual obligation or invitation. In line with the debates surrounding collective action problems, I take such a conviction as conceding that it does not appear rational for an individual to create public benefits if reciprocity or compensation is uncertain. It seems intuitive, however, particularly if a society does not centrally allocate its resources, rational agents must undertake this role of creating public benefits while expecting reciprocity as Bayesians.

In this paper, I will present ex-ante constrained maximization as a rational strategy, particularly by highlighting the normative obligations of the beneficiary party to become a constrained maximizer *ex post*.

In the following chapter, I will investigate the nature of public goods that are central to this inquiry. I will argue that defining free riding exclusively in terms of public goods could be misleading. This would urge me to propose an ownership model for goods that are publicly available but privately owned, thereby making contractual engagement preferable.

In the second chapter, I will build on the claim that it is possible to define moral action as an agreeable equilibrium point accessible to Bayesian players. I will argue that to retain the rationality of this claim with extended temporally asymmetrical scenarios, further proto-contractual norms must also be introduced to serve as Bayesian signals.

In the third chapter, I will investigate the benefiting party's moral and rational plays and conclude that it is rational for them to become a constrained maximizer *ex post* if they are benefiting from a publicly available good. I will compare their actions to a traditional

free-rider, and present larger-scale outcomes that would incentivize a rational player to adopt an agreeable cooperative strategy.

I will conclude with the aim of convincing the reader that by extending the theory constrained maximization as a rational strategy to time-asymmetric scenarios, morality and rationality can reconcile the puzzling temptation to defect as a strategy in the context of collective action. As a result, particularly producing and benefiting from goods with a public nature would be a less worrisome case for mitigating the collective action problems.

CHAPTER 1

PUBLIC GOODS, FREE RIDING, AND THE PUZZLE OF OWNERSHIP

1.1. What is Free Riding?

Free-riding as a practice is, in fact, a rational temptation that is not entirely independent from the scale at which it is examined (R. Hardin, 1971). David Hume is one of the early theorists behind the concept, as he plausibly identifies it as a *political inconvenience* that necessitates a centralized body to allocate resources and skirt these ‘inconveniences’ (Hume, 1739:538 ; Barry, 2010: 378).

Mancur Olson and Russel Hardin would agree with the Humean conclusion that the practice of free-riding creates a *problem*; leaving also the Hobbesian solution aside, whether it is a problem that needs to be socially and politically mitigated is a puzzle that must be further investigated.

To this extent, John Rawls famously argued that the nature of free riding is comparable to the effect of which the action of free-riding is accounted towards (Rawls, 1971: 267). In other words, it may be possible to conclude that an action is inconsequential, as Olson presents a similar case for *latent groups* whose contribution may be negligible (Olson, 1965; Tuck, 2008).

Figure 2. 2x2 Chicken game

		Player 2	
		<i>C</i>	<i>D</i>
Player 1	<i>C</i>	5, 5	2, 10
	<i>D</i>	10, 2	0, 0

Richard Tuck, however, argued for the scaled contributions, that a contribution may be instrumentally valuable to meet certain ends for a deliberative threshold that makes the contributing action rational based on rule-utilitarian grounds (Tuck, 2008: 99). The problem of cooperation, then, if enough people are cooperating, resembles a *chicken game*, where the Nash-equilibria are both (C,D) and (D,C) outcomes². In other words, for this type of game, the cost of mutual defection (D,D) is so high (notice 0,0 in Figure 2) that a player might contemplate it is better to play a cooperative strategy rather than letting the system fail (Taylor, 1971).

Finally, Jean Hampton presents an intuitively similar but fundamentally different coordination model to explain the nature of the practice, essentially linking the nature of free-riding to the relationship between the production units of the goods and the costs of individual contributions (Hampton, 1987: 254)³. According to Hampton's model, the problem of cooperation is naturally a game of coordination, where (C, C) is a coordination equilibrium, a Pareto-efficient outcome fundamentally distinct from the Nash equilibrium in PD (Hampton, 1987: 253).

² Michael Taylor also presents a quite similar argument for the problem of cooperation (Taylor, 1987).

³ Hampton distinguishes her model of *battle-of-the-sexes* with Tuck's and Taylor's models of *chicken game* by arguing that the former has coordination-equilibria, which defined by reference to David Lewis as Pareto-efficient points of equilibrium that no player could become better off if any player acted differently (Lewis, 1969; Hampton, 1987: 253; Taylor, 1987; Tuck, 2008; also see Vanderschraaf, 2001: 41). However, without elaborate discussions on the *step goods* which Hampton crucially underlines, I prefer relying on the *assurance game* model to compare both. It is only important to note that while the *chicken game* asks 'Do we cooperate?', the *assurance game* asks, 'How do we cooperate?' (Lewis, 1969; Hampton, 1987: 253).

Figure 3. 2x2 Assurance game (also known as stag-hunt)

		Player 2	
		<i>C</i>	<i>D</i>
Player 1	<i>C</i>	5, 5	1, 2
	<i>D</i>	2, 1	3, 3

It is so far evident that the nature of goods is relevant to the discussion, particularly insofar as the investigation concerns the potential efforts to mitigate the problem on a population-level scale. It is nonetheless puzzling, though, as the scale of the free-riding phenomenon decreases, many theorists would be tempted to disregard the actions of a free-rider, since they are much more concerned with a population-level coordination than a bilateral contractual scale.

1.2. The Broad Definition of Free Riding

To elaborate on the controversial note at the end of the previous sub-chapter, I shall introduce the economic definition of a free-rider and present my worries accordingly. In the *Routledge Dictionary of Economics*, Donald Rutherford defines a free rider as an individual who does not pay for the goods or services she consumes (Rutherford, 2002). The peculiarity of this definition in relation to the worries raised by Rawls and Hampton is evident.

While investigating the emergence of the term in the relevant debates, Philippe Fontaine presents an interesting case about the development of the term in the historical context, particularly in the social sciences, in contrast to the common-sensical understanding of the concept and practice of free riding (Fontaine, 2014: 361).

I argue that we, philosophers, must not disregard the advantages of relying on a common-sensical understanding of this peculiar practice. So far, there are two viable paths from here, each with its own advantages.

First, as Rawls and Hampton do, we can scale the problem to a population level and argue for the possibility of cooperation, as long as it is plausible to expect so (Rawls, 1971; Hampton, 1987). This approach is commonsensical in its own sense, as it entails the observation that free riding is only relevant if the societal effects are commensurable.

Second, as I will argue below, we can disregard the puzzle of scaling and focus solely on the moral value of the action, without considering anything else. To further elaborate on this point, I will later introduce a contractarian model of moral normativity.

Both accounts are to some extent commonsensical, since the former emphasizes the effects while the latter underlines the moral permissibility of the action. I argue that we must abandon the former if we want to address a particularly relevant puzzle concerning not only heuristic maintenance, but also a deeper relationship between production and benefit.

1. 3. The Narrow Definition of Free Riding and Worries Behind It

The narrow definition of free riding includes *public goods* as its central element; in this sense, the famous case of *fare evasion* not only appears as an example, but also as a definition (Cullity, 1995). Both Olson's analysis of *latent groups* and Hardin's proposal of the *n-prisoners' dilemma* explain why it may be helpful to commonsensically disregard the question of free riding towards already established and somehow maintained public goods, such as public transportation (Olson, 1965; R. Hardin, 1971; Cullity, 1995).

This conclusion though, once again, is only heuristic. Practically, it is entirely plausible to scale the free rider problem only to instances where its effects are greater than its costs.

When someone deliberately omits buying a ticket for a ride in a system that only requires honor payments, they make a rational calculation of their expected value

compared to instances in which they could get inspected for this ride or not⁴. On the other hand, catching those who free-ride is beneficial to the transportation system, as long as the inspection rate is not too high to deter people from free riding altogether.

So, it seems that heuristically, the rival decisions between the operators of the train (let's call them the Administration) and the free-rider are, in fact, blended together. The Administration may expect, or even want, people to free-ride, simply because the revenue it needs to generate can be compensated by the difference between those who pay the price and those who attempt to free-ride, get caught, and face fines. It seems to me that there appears to be a sweet spot between regular inspections and the temptation to free-ride, and it seems arguably similar to a *tacit agreement* (in the Humean sense; Hume, 1739; Barry, 2010; Vanderschraaf & Valls, 2019) between the customers and the Administration⁵.

Then, I contend, the problem of free riding either reduces to a problem of coordination by explicit or tacit contractual engagement; or, the problem transforms into a mathematical inquiry of optimization. I prefer the former, since it is not yet evident how the nature of goods is central to this investigation, as I wish to examine in the following sub-chapter.

⁴ If the probability of inspection is p , then she must compare the price she pays (T) with the expected value of evasion, that in this case, only concerns the settled fine amount, if any (F), and p . If $T < p * F$; then she must calculate that it is rational to *pay the fee instead of free riding*. Suppose the ticket costs \$2.50 and there is 10% chance that there would be an inspection, that if she gets caught, will be fined \$60. Since $[2.50 < (0.1 * 60)]$, free riding would not be the rational strategy here.

⁵ Suppose that the Administration needs to generate \$2400 for the daily operation of the train (excluding the cost of enforcement), where T is again \$2.50 and 1000 people expected to ride daily. Administration then could also rationally calculate that having an inspection rate close to but not equal to 0, may incentivize some to free-ride. Say n (the people administration expects to free-ride) equals to 632. For an inspection rate lower as 4%, this could mean that 368 honest people buying the ticket would generate the less than half of the expected revenue since $[368 * \$2.50 = \$920]$, while 4% of the freeriders (in this case, about 25 people) do get caught, and pay $[25 * \$60 = \$1.500]$. In total, the Administration exceeds the expected revenue of \$2,400 by generating \$2,420 in revenues, excluding the costs of enforcement. The worrisome case for this illustration is that the amount of honest people is lower than the amount of free-riders, but nevertheless the result would work for the Administration.

1.4. What Are Public Goods?

Garrett Cullity provides an invaluable breakdown of the necessary and sufficient conditions for a good to be considered a public good, which I intend to follow. For Cullity, a public good is:

- *Jointly supplied*: If it's available to anyone, it is available to everyone (Schmidtz, 1991; Cullity, 1995)
- *Nonexclusive*: If everyone is enjoying the good, then no one can be prevented from enjoying it (Klosko, 1987; Schmidtz, 1991; Cullity, 1995)
- *Nonrival or jointly/collectively consumed*: One person's enjoyment of the good does not diminish the enjoyment of others (Samuelson, 1964; Taylor, 1987; Cullity, 1995)
- *Compulsory/open*: Enjoyment cannot be avoided without considerable inconveniences (Simmons, 1979; Cullity, 1995)
- *Equal/pure/indivisible*: Everyone consumes the same amount (Arneson, 1982; Cullity, 1995)

A good, therefore, is public if it is compulsorily, collectively, and similarly available to the enjoyment of everyone. Traditional examples are security, public health, and public infrastructure (Rutherford, 2002). Consider, however, the following challenges:

Arguably, it is unclear from this definition who owns the public goods. Two replies may follow. One may claim that everyone owns it, since they tacitly agree by participating in social interactions; they are bound by benefits and obligations by virtue of being a member⁶.

The member-ownership model is worrisome. It could potentially present a compelling response to a member living miles away from a road in a suburban area that she has and would never set foot on, by arguing that she is nonetheless a member in a holistic sense that entitles her to benefits and obligations jointly (Gilbert, 2006). Yet, this account would fail to respond to a challenge when someone from another country, or even, some

⁶ On the membership problem, see Gilbert, 2006, cf. Simmons, 1979.

alien from Mars, comes to use that road. The Martian is arguably not a member who entitles them to benefits contingent on membership; then, if public goods are collectively owned, members must be able to exclude them from their enjoyment, but alas, they cannot⁷.

One could also respond to the ownership challenge by arguing that no one owns public goods, or that the ownership puzzle is irrelevant. This account would rely on the fact that public goods are defined by how they are enjoyed, while private goods are defined by how they are owned or produced (cf. Samuelson, 1964; Ostrom, 2004). Then, it appears that if a good is publicly available, then the ownership problem must fundamentally dissolve.

I find this conclusion deeply unsatisfying, since most, if not all, goods and services that exemplify public goods are to some extent produced, and that production requires labor, which requires costs and so on. If the costs are distributed among members in the form of taxation, then members are either collectively or individually responsible for their share of the production. It must follow that unless members intentionally abandon their responsibility, someone (or some entity) must own public goods. Then, the irrelevance-response either circles back to the member-ownership model; or, the respondent may refer to the State as the sole owner of the public goods, which bears further concerns I will investigate below.

It appears that the definition of the public good is tailored to fit the intuitive goods and services that are commonsensically believed and practiced to be publicly available.

What is the difference between a public road and the house I built adjacent to it?

Following the definition of a public good, the demarcation point appears to be the scope or availability of enjoyment. Following my challenges, however, it appears that this demarcation is only intuitively plausible, and it is fundamentally worrisome unless further support is provided for the definition from a private ownership perspective.

⁷ At least practically. Following on the membership model, they must be able to collectively intend to exclude anyone from Mars to enjoy their commonly-owned member benefits, yet without a sci-fi solution that only allows Earthlings to ride on a certain route, it is practically impossible.

1.5. What are Club goods?

Before I raise further worries on what I call the ownership puzzle, consider club goods as an alternative model for goods that are somewhat publicly available. Buchanan defines club goods as an extension of ownership and consumption rights over a differing number of individuals (Buchanan, 1965). To this extent, club goods are not an invention out of thin air; instead, they are rational arrangements between individuals with similar demands to redistribute their individual costs.

For Buchanan, the scale of the club and the undertaken costs are simultaneous and interrelated (Buchanan, 1965). Simplistically following his example, suppose you want to build a pool in your neighborhood that would cost you c plus maintenance. It would be rational, if not advisable, for you to first ask your neighbors if they would be interested in participating in this project. Suppose you ask four of your neighbors, and they agree to participate. Then, the production and maintenance costs would be divided among five club members, as each would enjoy the pool equally once it is produced.

Notice two things: First, there is a clear-cut understanding of how the membership is formed based on the examination of who undertakes the costs and who will benefit from the good. Every remaining element of the pool, once built, fits the definition of public goods provided above.

Which brings us to the second caveat, that is, exclusiveness. Five neighbors deciding on the pool project would rationally expect an optimal cutting point towards which individuals could enjoy the benefits. This motivation concerns the relationship between cost distribution and enjoyment, that is, the very motivation an individual may be tempted to participate in a club-good-arrangement in the first place.

It is true that having more than five people on board for building a pool would lower the costs attributed to one single member; nevertheless, the amount enjoyed would reduce accordingly as more people become part of this arrangement (Buchanan, 1965). This is why the club arrangement is only effective to an extent of an optimal contribution.

The means to secure this optimality is again, exclusion, and the means for exclusion are property rights (Buchanan, 1965; Sandler & Tschirhart, 1997). Notice, however, that practically, goods and services arranged to be as club goods are not distinctively observed from the public goods, unless external measures are established.

If what makes a road a public good is the fact that there is no gate officer on every highway expecting payments of contribution from those who ride, while there may be fences around a pool that practically exclude an outsider; the distinction between public and club goods is not that of their nature, but how they are provided⁸.

1.6. The Ownership Puzzle

We have thus encountered a mandatory juncture for what has so far been argued on the nature of public goods. The puzzle I wish to examine essentially concerns the question of whether it is possible to attribute ownership to goods that are publicly available. Club goods presented a plausible argument for alternative modeling, with the caveat of exclusion. Hence, an objector, particularly one who relies on the irrelevance-argument described above, may claim that since club goods are exclusive, they are fundamentally different than public goods.

I contend that this objector is misled by the extent of definitions. It is true that the definition of a public good proposes nonexclusiveness as a central element (cf. Cullity, 1995). The ownership problem suggests that exclusiveness is contingent on ownership, and without a plausible account of ownership, the concept of public goods floats on vague grounds.

Consider the comparison between public roads and private/club pools once again. The only distinct nature that we have identified so far concerns the practical implications of the efforts of exclusion.

⁸ Ostrom's model of *common property resources* also provides a similar argument, particularly pertaining to the question of ownership (Ostrom, 1990). Since I am investigating the nature of public goods only instrumentally, it appears beyond the scope of this project.

Suppose you are observing a public road and an adjacent pool club. You observe that the same five people are driving the same road, parking on the ground, and jumping into the pool. There is no security whatsoever around the pool. Based on your observation, you conclude that you can also enjoy the pool. You drive the same road, park your car in the same place as other patrons, and so far, you face no problems. When you try to enter the pool, however, one of the patrons warns you that this is a private club, exclusive to its members.

You would be right to be befuddled when you encounter such an exclusion. The patrons, however, might show you relevant legal documents certifying their ownership of the pool grounds. The legal documents, such as the land title, would not practically exclude you from entering the pool, though; the practicality would only become relevant once the patrons invoke their rights of protection entitled by their ownership. You could practically enter the pool, but you would be removed if, for instance, law enforcement is summoned.

The fact that the same legal procedure was absent, as if you were riding the road that seems as available as the pool from the outside, is puzzling. The most intuitive way out of this puzzle is to attribute an ownership model for the road as well, for you to claim that you can safely enjoy the road because it is either collectively owned, or it is owned by the State, and you are allowed to enjoy it.

This conviction leads us to the crux of the puzzle, as I shall now argue that the ownership could only be attributed via hypothetical agreements or production. By production, I mean Locke's theory of labor (Locke, 1998 [1689]; Russel, 2004); while by agreements, I mean tacit compliances over already produced or appropriated resources, as per Locke and the legal doctrine of prior appropriation (Lewis, 1969; Tarlock, 2000; van der Vossen, 2020; also see Gauthier, 2022b).

The difference, then, that you are allowed to ride on the road but not allowed to enter the pool is a difference about who owns which goods. Notice that this is not simply a conclusion regarding the distinctiveness of private goods, since the pool already satisfies the necessary properties of public goods as described above.

Although it may sound controversial to an objector, an argument that follows may extend the scope of this conclusion by claiming that any good or service that is available to everyone, regardless of its intuitive nature, could be interpreted as a private good if and only if justly owned, even though it possesses the necessary properties of a public good.

1.7. The problem of exclusion and the lighthouse model

The crux of the ownership puzzle was to attribute a model of ownership for goods that resembles the nature of goods that are publicly available. Above, I attempted to investigate the necessity of the question of ownership by highlighting a peculiar concern. Now, I shall elaborate on the model I propose.

Goods and services that are publicly available must be somehow owned, be it collectively or privately, to be considered public goods or services. Otherwise, those who share a similar nature (such as the pool in the abovementioned case) would be ambiguous. This is so because the ownership problem is closely tied to the legal and practical means of exclusion.

If a collective entity owns a public infrastructure, then everyone can benefit from it. This entails that everyone can also abuse the good, by, say, free riding. It is curious, however, that the means of exclusion affect the scope of an individual's enjoyment of public goods that are arguably owned collectively. Suppose, for instance, you are one of the honest ticket-payers I have modeled above. You may be tempted to overlook the fact that free riding is inevitable, and the transportation system is utilitarianly optimized to nonetheless compensate for the costs necessary for maintenance.

It is optimized, however, to the extent of the society that you are also part of. The calculations the Administration did for optimizing free riding were based on social dispositions. You may therefore be tempted to tacitly agree with this convention that today you paid the fee, and tomorrow you could perhaps benefit from the extent of your fellow member who free-rode on that; by, say, using a road that she has never set foot on. I hold that without further elaboration on the rational strategies for larger-scale sustainability, such conventions are only rational for small-scale populations where tacit

agreement has direct influence on the costs you undertake and the benefits you receive. Similar to Ostrom's model of *common property resources*, minor deviations could plausibly be tolerated in small-scale resource management if the resources are collectively owned (Ostrom, 1990; Feeny et al., 1990).

If, again, a Martian free-rides on the train, then the optimization calculations would certainly fail, since the Administration calculated the probability of inspection based on observable social dispositions. You, therefore, would be justified in believing that the Martian should be excluded from train rides. This would either lead to stronger preventative measures that would eventually diminish the once-optimized model of resource management, or a model that is not entirely different from club goods.

Private ownership, whether through a club arrangement or not, resolves this puzzle by highlighting the means necessary for calculating the costs and benefits that must be considered to rationally produce and maintain goods that are publicly available.

To shed some light on this concern, let me introduce lighthouses as a curious case. The debate over the nature of lighthouses is certainly not novel, as it has concerned philosophers and economists for a while because of its peculiar nature (see Mill, 1994 [1848]; Sidgwick, 2011 [1901]; Pigou, 1932; Samuelson, 1964; Gauthier, 1986). The reader might be familiar with R.H. Coase's comprehensive account of the nature of lighthouses, which essentially optimizes private ownership with public duties (Coase, 1974: 376). Although highly persuasive, I am tempted not to take Coase's *Trinity House* account on its face value⁹.

This is mainly because a footnote presented in his analysis, quoting Kenneth Arrow. In this quote, Arrow provides a thought experiment that essentially answers the question of

⁹ I am mainly worried that Coase provides an empirical analysis, and not a normative one. I am not interested in examining empirical data for lighthouses around the world, to provide a descriptive account of their ownership model. Instead, I am arguing for necessary and sufficient conditions that could illustrate whether ownership and exclusion are possible. Nevertheless, contrarily to what some critics may contend (particularly David van Zandt and Elodie Bertrand), I believe Coase's account presents a plausible empirical ground for what I am about to argue on the possibility of private ownership for a publicly available infrastructure such as a lighthouse (van Zandt, 1993; Bertrand, 2006).

exclusion for lighthouses by suggesting that since the lighthouse-keeper can shut down the lights, exclusion is in fact possible¹⁰ (Coase, 1976: 375).

Coase initially disregards this thought experiment by claiming that it is surreal (Coase, 1976: 376). Nevertheless, neither he nor any critics of Coase's account presents a compelling account that explains why Arrow's exclusion thought-experiment does not normatively account for an exclusion within the debates of the nature of public goods.

Consider, for the sake of argument, that exclusion is possible. If so, again, the nature of a lighthouse fails to fit the described conditions of a public good (Cullity, 1995). Then, the reader might agree that this nature is peculiar, perhaps similar to club goods. This similarity, though, only extends to distinguish its nature from commonsensical public goods.

To clarify, I argue that a lighthouse, or any good or service that is publicly available, can be privately owned via Lockean labor (Coase, 1974: 360-362; Locke, 1980 [1690], §27), and this does not diminish the owner's entitlement to distribute her costs as expected payments. This conviction would necessitate me to present grounds for private ownership to not only the lighthouse as a building or establishment, but, quite controversially, to the light as well, by interpreting the light as a homestead-like contingency product of labor (see Rothbard, 1982: 52; Block, 1983: 12-15) and interpreting the provided service as the contingency of the ownership (see assurance contracts in Schmidtz, 1991: 55).

¹⁰ The quote is from a compendium of papers submitted to the Joint Economic Committee of the United States Congress in 1969.

CHAPTER 2

CONTRACTS AND THE NATURE OF MORALITY

2.1. Why Contractarianism?

An important step between the argument on the nature of public goods and a normative moral disposition to honor this nature requires further clarification of the grounds in morality. I shall, therefore, clarify my arguments for moral permissibility.

Recall that above, I illustrated that a transportation administration may expect some passengers to free-ride, and therefore optimize its expected revenue by inspecting an adequate proportion of free riders. Free riding is certainly impermissible in this example, since it is sanctioned. Nevertheless, notice that two-thirds of all the passengers were optimally expected to free-ride in my rather exaggerated example.

The abovementioned proportions matter to some extent, particularly when arguing against the possibility of objective moral standards. In *The Nature of Morality*, Gilbert Harman makes an interesting remark on observations and moral intuitions by investigating the phrase “When in Rome, do as the Romans do.” (Harman, 1977: 97). Harman’s account of observation follows a Humean notion of *tacit convention* when contesting moral objectivism (Harman, 1977; also see Vanderschraaf & Valls, 2019).

It is interesting to note that one could only challenge the Roman convention by referring to moral intuitions that somehow contrast with the observed dispositions. Following Harman's example, you would be disposed against becoming a cannibal in a cannibal society, likely because of your moral intuition against cannibalism (Harman, 1977:97-8).

The most crucial and relevant question concerns whether this mere intuition provides valid grounds to conclude that the intuition you feel when encountering such a society is objective or not. My answer would be no; in fact, quite contrarily, you would be disposed to such an intuition because of the conventions that you have been previously observed or have become disposed to¹¹.

To this extent, the reader may agree that we can leave the question of practical reason aside for the sake of argument, as we now have grounds to not only explain why certain societies are disposed to certain actions, but also why those actions may not be in accordance with our moral intuitions¹².

This conviction is essentially why I believe the account I am presenting is of a contractarian nature. Contractarianism prevents the potential pessimism towards questions against moral realism by grounding morally justified actions as rational constraints (Gauthier, 1991)¹³. It must follow, then, that you may be intuitively disposed against in Rome, doing what Romans do; nevertheless, you may be rationally inclined to follow the convention for self-interested reasons¹⁴.

¹¹ Michael Huemer cites Ruth Benedict to object a similar point (Huemer, 2016).

¹² The scope of this project compels me to ignore a crucial debate here; and I must certainly provide further grounds on the questions of moral motivation and practical reason in the future.

¹³ I wish to elaborate on a more implicit reference to Gauthier with this claim. In his 1991 chapter "Why contractarianism?" in Peter Vallentyne's edition *Contractarianism and Rational Choice*, Gauthier begins by proposing that 'morality faces a foundational crisis' (Gauthier, 1991: 15). This note is acknowledged by what I refer to as 'pessimism', as I also believe contractarianism is not only a moral theory on its own, but it is in fact a plausible response to cogent worries raised by moral nihilists.

¹⁴ For a cautious reader, this argument may sound as if an advocacy for becoming a cannibal in a cannibal society. Yet the reader must also note that a rational agent in such a scenario may be able to calculate the expected payoffs of various strategies, including, of course, conforming the norm, but it would unlikely be limited to it. Christina Biccheri presents compelling grounds of how such intuitively horrendous norms could rationally be conformed (Biccheri, 2014).

2.2. Could There be Proto-Contractual Norms?

The conclusion above, unfortunately, flouts the question of free riding that has thus far shadowed this investigation. It appears that my proposal for free-riding-optimized revenue generation of public transportation begs the question of whether free riding is morally permissible or not.

If we follow an orthodox account of contractarianism, say, by taking Hobbes at face value, we realize that the problem circles back to the PD (Vanderschraaf, 2001: 31)¹⁵. Notice that, however, the optimized model also allows for interpreting the problem as a coordination game, as Jean Hampton would agree (Hampton, 1987)¹⁶.

With a rational disposition to maximize self-interest, it is both a normative and practical challenge to establish grounds that would incentivize rational agents to interact with one another. It is indeed rational to calculate whether a contractual interaction would be beneficial for an agent or not; nevertheless, this agent must be able to expect that the other party would act justly. This concern is practically different than constrained maximization, since it does not urge the rational agent to calculate her payoffs compared to cooperative and noncooperative outcomes (Gauthier, 1986); instead, it prompts her to adopt, and expect the other party to adopt, certain rational dispositions during bargaining¹⁷. What I will now define as *proto-contractual norms* will serve as the normative grounds for my proposal of *ex ante* constrained maximization.

Suppose you enter a store. You like a t-shirt listed as \$50, which is above the price you are willing to pay. You ask for a discount, and the shopkeeper tells you, “Keep it, it is

¹⁵ Daniel Eggers argues for the opposite, that Hobbes’ account particularly in *De Cive* deviates from PD (see Eggers, 2011)

¹⁶ Initially, the expected cost of free riding was 40 (assumed the inspectors would catch 4% of the free-riders) divided by 632 (number of optimized free-riders), which gives us the probability of getting caught of 6.3%; times 60\$ (fine amount), which roughly calculates \$3.80. Since the ticket costs (\$2.50) less than the expected value of free riding (\$3.80), it was possible to calculate that about one-third of the daily riders would be honest-ticket payers. Notice that if the chance of getting inspected is 4%, then the expected value of free riding becomes even with the ticket cost with exactly 960 free-riders as $[(0.04)*\$60 \approx \$2.50]$. Then, the initial argument would also model that, everyone would be rational to free-ride; unless further scrutinized by the Administration, the expected value of free riding becomes less costly for even the honest ticket-payers.

¹⁷ The difference I wish to stress here is grounded by concerns raised by some scholars as the problem of ‘pure instrumental morality’ (Moore, 1994; Moehler, 2020).

free today”. You are confused as you walk out of the store with your new t-shirt. You realize, only later, that the person you thought was a shopkeeper was in fact just a random stranger.

It is intuitive to note that this encounter would entitle the real owner to coerce you into refunding your newly acquired item. Your bargain with the faux-shopkeeper was void, and I contend that a similar analogy could be made towards instrumentally rational moral bargainings.

It is plausible to expect, then, agreeable signals from moral others. These signals differ from the contractual norms governing the fairness of the interaction, such as the principle of minimax relative concession I will investigate below (Gauthier, 1986). Proto-contractual norms only signal that you are bargaining with the correct individual, bound by similar, if not the same, moral obligations limited only to the nature of the contract.

Various principles could exemplify such norms, and since I am not arguing for a universally agreed-upon principle, it would not be useful for me to attempt to create an exhaustive list. It is only important to note that signals precede the contract itself, and ensure that the bargain would lie on agreeable normative grounds.

Consider proto-contractual norms as salient signals that stand out in some conspicuous respects (Lewis, 1969: 35; Vanderschraaf, 2001: 53; see also Schelling, 1960). While rationality compels the agent to look and interpret those signals, morality compels her to adopt them if necessary.

Prior to bargaining, then, the rational agent must investigate whether she is engaging with moral and rational equivalents. She may look for conventions practiced based on legal principles such as *pacta sunt servanda* (see Sharp, 1941; Wahlberg, 1959), or, say, *lex mercatoria* (see Burdick, 1902). Alternatively, she may calculate whether the agent she is interacting with is Lockean, narrowly construed, particularly following Gauthier’s contractarian interpretation as a normative prohibition of bettering own situation while worsening others’ (Gauthier, 1986: 205; Arneson, 1987: 299).

I contend that proto-contractual norms are not necessary for the possibility of mutually-beneficial cooperative outcomes; rather, they are necessary for a rational agent to calculate that they are bargaining with a rational and moral equivalent.

This calculation need not be deliberative in the sense that compels agents to signal with unrealistic means, such as wearing a hat that reads ‘I am an honest person’. Rather, they can be adaptive, as Skyrms would suggest (Skyrms, 2010). This means that the rational agent must only be on the lookout for such signals to plausibly interact with others, and calculate only narrowly, whether the interaction she is about to engage in normatively satisfies the grounds necessary for her soon-to-be-optimized self-interests¹⁸.

2.3. The Principle of Minimax Relative Concession

During bargaining, fairness and impartiality serve as the basis of justice (Gauthier, 1986, p. 14). Gauthier introduces the principle of minimax relative concession (hereinafter MRC) as follows¹⁹:

“... the principle should state that given the range of outcomes, each of which requires concessions by [individuals] if it is to be selected, then an outcome [shall] be selected only if the greatest or maximum relative concession it requires, (...) or minimum, that is, no greater than the maximum relative concession required by every other outcome.” (Gauthier, 1986: 137).

The mathematical basis for interpreting the bargaining solutions as relative concessions is not novel to Gauthier (see Kalai & Smorodinsky, 1975; Vanderschraaf, 2023: 68-69). Gauthier’s contribution, I believe, is to define the initial bargaining position with a non-

¹⁸ An interesting example could be the perception of time, both as a signal and as a convention. Different cultures perceive time differently, and this perception signals a prospective contractor normative grounds she would rationally comply with. While a German may expect her to respect time boundaries much more strictly, a Turk may tolerate unsolicited delays. Hence a rational agent may be tempted to be more time-sensitive while meeting a German than meeting with a Turk, based on salient signals (Lewis, 1969). Notice, however, that she may be entirely wrong to assume that a Turk would be tolerant towards delays, because she only has prior knowledge based on her observations on conventions, but not on the proto-contractual norms of the certain Turkish rational agent she is about to interact with. Then, she would be rationally compelled to update her behavior (Skyrms, 2010).

¹⁹ Peter Vanderschraaf highlights that Gauthier’s most recent name for this concept is *maximin proportionate gain* (Vanderschraaf, 2023). To plausibly reflect on the 1986 version of *Morals by Agreement*, however, I will be using the name *minimax relative concession* (MRC).

cooperative outcome (Gauthier, 1986: 132). In other words, he plausibly distinguishes between instances where the games are played as if the players are capable of coordinating with each other and those where they are not. Consequently, he proposes that MRC is a principle of justice that prompts the players to seek relative utilities not based on what they have (initial bargaining position) but based on what they could achieve (concessions).

Thus, rational agents are motivated to engage in moral bargains without the fear of not achieving optimal and beneficial outcomes (Gauthier, 1986: 132-133). This conviction is particularly crucial to questions of morality, since cooperation is not a good-to-have, but rather, a must-have to secure at least minimally tenable grounds of moral permissibility²⁰.

2.4. Constrained Maximizer and Her Dispositions

In PD, the expected payoffs of one player are denoted by the following inequality: $T > R > P > S$ (Axelrod & Hamilton, 1981). Notice that T denotes the greatest payoffs, whereas $R > \frac{S+T}{2}$ provides that, in fact, cooperation yields greater payoffs than both noncooperative payoffs. Since T pays off the most, both players would want that, and consequently receive P , even though $R > P$. Notice in Figure 4 that denotes payoffs for one player in PD that the temptation (T) is greater than the reward of cooperation (R) is greater than the penalty of defection (P) is greater than the payoffs of a sucker (S) as $5 > 3 > 1 > 0$; which also holds for $2R > (S+T)$.

²⁰ Gauthier presents the former equilibrium as the Zeuthen-Nash-Harsanyi bargaining solution (Gauthier, 1986: 146). It is important to note that the fairness element in MRC is presented as an alternative to this account that proposes that utility-differences within the cooperative surplus should be maximized instead of relative concessions (Gauthier, 1986: 148).

Figure 4. Payoffs for one player in PD.

		Player 2	
		<i>C</i>	<i>D</i>
Player 1	<i>C</i>	$R = 3$	$S = 0$
	<i>D</i>	$T = 5$	$P = 1$

The reason why the player seeks T is that she is a self-interest maximizer. She plays for T not only because she seeks the greatest individual payoff, but also, she avoids being the *sucker* by playing a cooperative strategy unilaterally (Axelrod & Hamilton, 1981)²¹. Gauthier acknowledges this nature of the PD game by correctly distinguishing between a straightforward maximizer (hereinafter SM) and a constrained maximizer (hereinafter CM) (Gauthier, 1986: 167).

In its essence, an SM is one who calculates $T > R$, whereas a CM is one who calculates $R > P$. The problem for an SM is, however, that she knows she could never achieve T if surrounded by SMs like her, but only P . This is why Gauthier compares an SM to Hobbes' *Foole* (Gauthier, 1986: 167; Hobbes, 1985 [1651]; cf. Vanderschraaf, 2010).

A CM, therefore, is conditionally disposed to cooperate in ways that which if followed by others, generate the optimal payoffs (Gauthier, 1986: 177). In other words, she is only rationally disposed to cooperate with fellow CMs and avoid SMs²².

Let us denote the mathematical plausibility of this conclusion. Let r be the percentage of moral agents (CMs) in a given society; let p be the probability of CMs recognizing each

²¹ Notice the conspicuous tension between the *sucker* in Axelrod and Hamilton's paper and the *fool* in Gauthier's, by reference to Hobbes. While the sucker plays cooperative strategy (with C,D outcome), the *Foole* plays a defecting strategy (with D,C outcome), and neither are plausible.

²² Here, Gauthier provides an ambiguous comparison between a *translucent* CM vs. a *transparent* CM (Gauthier, 1986: 174). These definitions are vague, and are not entirely compatible with my account of signaling and proto-contractual norms.

other, and q be the probability that a CM fails to recognize an SM and gets exploited. Gauthier argues that an agent must be a CM if and only if p divided by q is greater than the gain from cheating $(1 - T)$ divided by the gain from cooperating $(R - T)$; summed with the ratio of SMs in the society $(1 - r)$ times T divided by CMs in the society (r) times the gain from cooperating $(R - T)$ (Gauthier, 1986: 175-177). This inequality suggests that CM is an adaptive strategy rather than a normative disposition, and the strategy to become a CM is directly correlated to r being the percentage of moral equals in a given society. The inequality, is, then: $\frac{p}{q} > \frac{1-T}{R-T} + \frac{(1-r)T}{r(R-T)}$ ²³.

My prior suggestion of proto-contractual norms becomes relevant here, as it provides some degree of moral homogeneity by normatively prompting r to be not lower than an implausible threshold that would prompt a morally disposed rational agent to act immorally. Notice that we expect the added value on the inequality, that is, the population risk, to be lower when r is as close to 1 as possible.

Then, the reader might agree that the problem circles back to collective action problems, since the proposal of constrained maximization works effectively only if it is plausibly signaled with similar dispositions. Without prior knowledge of r , constrained maximization does not function as a normative framework that prompts the rational agent to always be disposed towards cooperation; instead, this would be irrational.

Signals, once again, appear to be the missing link. Updating strategies based on prior and posterior knowledge can be incorporated within the rational framework of constrained maximization only if proto-contractual moral norms are adequately established and signaled accordingly. This, then, must be an account of acting like Romans in Rome (Harman, 1977: 97).

Only that we are neither Romans nor in Rome, and, hypothetically, if we are only bound by our own moral and rational capacity to flourish our society, then we must agree on normative dispositions and treat defection with proportionate moral sanctions (see

²³ The inequality denotes that if the ratio of p divided by q is greater than the population risk, than the rational agent would be better off if CM (Gauthier, 1986: 177). Note that T and R are for denotation purposes only.

Bicchieri, 2014; Gauthier, 2022b). Then, the question becomes: “Who starts first?” This will be the starting point for my argument for *ex ante* constrained maximization.

CHAPTER 3

WHY WOULD ONE BUILD A LIGHTHOUSE?

3.1. The State of Absence

In a decentralized social order, where resources are not centrally allocated to mitigate rational temptation to defection, a rational agent would be signaled that the resources are produced from the bottom up (cf. Skyrms, 1996). This potential signalling is the reason why I am interested in a state where the centralized resource allocation schemes are absent.

I am tempted to refer to this bottom-up state as a ‘state of absence of centralization’. This is not a state of nature in a traditional social contractalist interpretation, which would define it as a proto-State status quo (cf. Hobbes, 1998 [1642]; Kavka, 1983; Hampton, 1986; Narveson, 1988; Arneson, 1991; Gauthier, 2022a; Vanderschraaf, 2023²⁴). Although similar, this is also not a political theory of decentralization, as I am only clarifying the necessary grounds for a philosophical argument, rather than conducting an empirical investigation into the social and economic nature of decentralization. The ‘state of absence’ would, by definition, be natural (see Nozick,

²⁴ In his book *Bargaining Theory*, Vanderschraaf makes a remark on agents using bargaining theory as a tool may rationally decide among alternative social contracts, characterized by how parties fare at a state of nature (Vanderschraaf, 2023: 76). My proposal, in this sense is to treat the *state of nature* as an alternative contract may which eventually yield greater payoffs for engaged parties. This is one the reasons why I do not use the *state of absence* interchangeably with the *state of nature*; instead, in my account, the *state of absence* would be closer to a social contract than the state of nature.

1974; Simmons, 1992); hence, it may be possible to use the ‘state of nature’ only as a reference, but not as a philosophical terminology.

For the Hobbesian account, every agent in the state of nature is a self-interest maximizer, which would eventually prompt them to agree on the establishment of the State, providing them with safety and security that are absent in the state of nature (Hobbes, 1985 [1651]: 192). They are not, however, according to Gauthier’s account, SMs that require the cooperation of other moral agents to fulfill their aim of achieving the payoffs of temptation. The difference, as stated above, is that SMs need others to be CMs, while the same condition applies to CMs, that they also need others to be CMs.

I propose to take this final strategy as the starting point of my argument for the nature of the state of absence, rather than the traditional Hobbesian account. If the problem is that rational agents cannot agree on morals, suppose that there are valid normative grounds that they are rationally better off if they agree, and worse off if they do not.

To argue on this controversial statement’s plausibility, I must present two grounds. First ground is mainly covered so far that morality is in fact normative grounds that are agreed upon and complied with; insofar as the sanctions of defection are adequately disincentivizing (see Biccheri, 2014: 210). If plausible, this argument would follow in the state of absence.

The second ground extends and locates the argument for the ownership in this state. I argue that the strategy of constrained maximization needs to be reevaluated to address the issue of starting first, proposing that this would occur if agents were socially and economically isolated from centralized decision-making for resource allocation and appropriation.

Notice that our pursuit so far is to signal the rational agent that r ratio (the ratio of the number of moral players) is sufficiently high to engage in contracts. I argue that in the state of absence, the threshold is more likely to be taken as a normative ground. The r ratio requires moral homogeneity, and this homogeneity may be imposed by showing

that without an adequately high r , no rational agent would act *ex ante* to provide certain resources for the benefit of others.

Consider a society that has not produced public goods whatsoever, i.e., no roads, no centralized healthcare system, no centralized security, no lighthouses, and so forth. This society is in need of some crucial outputs: Goods with a public nature that are enjoyable to any interested party and a rational agent to produce and maintain those resources. The latter directly follows the former, and I argue that it must also be followed by moral and rational compliance with the expected costs of the enjoyed goods and services.

3.2. Acting First as a Rational Strategy

Suppose that Penny walks in an open valley. It is getting dark, and she needs something to light her way up. She finds a stick on the ground and remembers that she had a lighter in her backpack. She lights the stick on fire, forming a torch. Now it is completely dark out there, and the only source of light is her handmade torch. As soon as she starts walking with the help of her newly formed torch, she realizes that someone is pacing up behind her. “Thank God,” shouts Ben, a complete stranger, “you have a torch!”

Our previous discussions on the nature of public goods are pertinent to this puzzle I am about to investigate. This interaction between Penny and Ben is puzzling because Ben appears to be waiting for someone to produce a public good that he can benefit from, while Penny only acts to maximize her self-interest as she mitigates the cost of walking in the dark by producing a public resource. This interaction is reminiscent of a PD-like game between Penny and Ben, which is denoted in Figure 5 where V denotes the value of the light while c is the respective cost.

The setup for this interaction resembles our concerns for the state of absence in which centralized resource allocation schemes are absent, meaning that agents are isolated in their private rational calculations for resource management. This does not, by default, entail that Penny, the producer of the good, is a CM in Gauthier’s framework, since she is not yet capable of calculating whether she is about to interact with moral others who are rationally tempted and morally motivated to interact with her (Gauthier, 1986; on the

distinction between axiomatic and strategic bargaining, see Vanderschraaf, 2023: 36-37).

Figure 5. The production game between Penny and Ben

		Ben	
		<i>Act</i>	<i>Wait</i>
Penny	<i>Act</i>	$V - \frac{C}{2}, V - \frac{C}{2}$	$V - C, V$
	<i>Wait</i>	$V, V - C$	$0, 0$

This conviction illustrates the inadequacy of Gauthier’s account of constrained maximization, as its scope does not extend to the moral calculations prior to any bargaining (Gauthier, 1986: 173). The possibility of *ex ante* constrained maximization would allow us to conclude several outcomes regarding the interaction of Penny and Ben, as well as analogous larger-scale interactions that will follow.

First, the possibility of *ex ante* constrained maximization would imply that any action creating external benefits would grant the owner legitimacy to morally use and abuse her production, as an extension of her Lockean self-ownership (Locke, 1980 [1690] §27; for *abuse* cf. Nozick, 1974: 171). This nature would rationally incentivize agents to act as an *ex ante* CM in the state of absence, since it mitigates the costs of production with agreeable benefits.

Second upshot is, then, the benefits. In the state of absence, a producer of a public good would be an *ex ante* CM insofar as she allows other CMs to enjoy this resource to the extent of their contribution. If one is an *ex ante* CM, the *ex post* contribution she expects should be fair, in accordance with the principle of MRC (Gauthier, 1986: 135).

Third, she must be able to normatively expect cooperation. One is only rational to act as an *ex ante* CM in a state where she can rationally expect and foresee cooperation. It follows that although she does not have the temporal ability to bargain, she has several strategies to strengthen her hand. It is only important to note that while this inability to

bargain highlights the parting of this account with Gauthier's, a rational agent interacting with an *ex ante* CM must normatively calculate that she still has the means to leave the bargaining table, say, by putting off the light (Figure 6).

Figure 6. Respective payoffs illustrating Penny's bargaining power

		Ben	
		<i>Cooperate</i>	<i>Free Ride</i>
Penny	<i>Maintain</i>	<i>R, R</i>	<i>S, T</i>
	<i>Punish</i>	<i>P, P</i>	<i>P, P</i>

The most important element for the possibility of *ex ante* CM is therefore its setup, i.e., the state of absence. *Ex post* cooperation, then, is not only a rational strategy, but also a normative obligation that would sustain the resource management in such a state as a reciprocity.

3.3. Signals and Repetition

As an *ex ante* CM, Penny has several tools in her inventory that would ground her rational strategy of acting first. One important tool is the *grim trigger*, in accordance with the Folk theorem (Friedman, 1971; Fudenberg & Tirole, 1991). Suppose that the moment Penny and Ben start walking together, the game is played repetitively until they no longer need the light. In this repetition, Penny can update her strategy based on the signals she receives from Ben (for inductive learning, see Vanderschraaf, 2001: 111).

Penny could play unconditionally, meaning that, since she had already produced a public good, regardless of whether she received any cooperative plays or not, she could keep the resource as it is. In that case, a free-riding Ben would receive the greatest payoff of T (the temptation of free riding) now, and the value of light (*V*) as the interaction repeats (Figure 8).

Figure 7. Ben’s payoffs in repeated interactions²⁵

		Ben	
		<i>Cooperate</i>	<i>Free Ride</i>
Penny	<i>Grim Trigger</i>	$R + \beta V$	$T + \beta P$
	<i>Unconditional</i>	$R + \beta V$	$T + \beta V$

Unconditional plays, however, are not entirely compatible with constrained maximization because it is not entirely clear what moral motivation would prompt her to act unconditionally. Since we are investigating a scenario in which players are not acquaintances further beyond being moral and rational equals, an objector claiming that Penny could provide a resource unconditionally must present plausible moral grounds that would be compatible with her moral and rational motivation. Since I present an account in which the sole reason for Penny to act first is the rational expectation of compliance, the objector supporting unconditionality must prove that she could maximize different values that would essentially rationalize her accepting the lowest cooperative payoffs²⁶.

If Penny chooses to play with grim-trigger, meaning that the signal of Ben’s defection would prompt her to switch to a non-cooperative strategy, a free-riding Ben would receive fewer payoffs compared to cooperative outcomes, since a grim-triggering Penny would extinguish the light (Figure 7).

Also, notice that a cooperative Ben would not realize any difference between his payoffs whether Penny plays with a grim trigger strategy or not. This is the most crucial signal

²⁵ Notice that the regular game-theoretic notation for the discount factor is δ ; however, since I use it to denote the cost of contribution, I used β here to avoid confusion. This figure illustrates that if Ben signals free-riding, a grim-triggering Penny would switch to punishment forever. Ben’s payoffs would then become $T(\text{now})$ plus P (as the interaction repeats).

²⁶ The objector must also provide that this value would not only prompt Penny to accept the lowest payoffs but also motivate her to act *ex ante*. While the former could be argued by framing the maximized value as charitable or benevolent motivations, it would not extend to the latter.

that would essentially prompt us to consider the necessity of normative obligations to become CM *ex post* whenever interacting with CMs in the state of absence.

Notice that how r , the ratio of the moral agents in a given society, becomes no longer puzzling when deciding whether to interact with an *ex ante* CM (cf. Gauthier, 1986: 175). This peculiarity transforms the rational calculation of becoming a CM in the state of absence by aligning it closer to normative moral obligations than a descriptive and adaptive strategy.

In other words, the payoffs I have denoted not only illustrate why *ex ante* CM is rational, but also how it is normatively grounded on larger-scale contemplations concerning resource management. Since both parties play with the conviction that a resource would not otherwise be provided ‘if not acted upon’, I will build on this conviction by incorporating further elements of ‘if not complied with’.

3.4. The Scope of Cooperation

Regardless of Penny’s strategy in a repeated game, Ben receives the greatest payoffs if he cooperates with her. The nature of cooperation, though, is not yet clear in this account. The second tool I must introduce, then, is Penny’s ability to deliberate a fair and agreeable cost of compliance, as an extension of her self-ownership, which would essentially frame the scope of Ben’s expected cost of cooperation.

Penny can calculate an agreeable share of contribution based on the principle of MRC. MRC provides that agents may bargain with relative concessions regarding what they rationally strategize to achieve as a result of this interaction (Gauthier, 1986). It is only important to note that Penny has a higher bargaining position compared to Ben, set by her capacity to determine a threat point (cf. Nash, 1953), but this does not entitle Penny to exploit Ben if her proposal means Ben could settle on much advantageous outcomes with alternative actions. Nor, it entitle Ben to exploit Penny by proposing that a bargain is not necessary since the good is already produced.

We have so far only established that Penny has produced a public good, and she signals cooperative strategies as an *ex ante* CM. Consider the following scenarios:

Scenario 1: *Penny allows Ben to walk with her and expects him to say ‘Thank you’ after this interaction.*

Scenario 2: *Penny allows Ben to walk with her and expects him to pay \$5 when they are safe outside of the woods.*

Scenario 3: *Penny extinguishes the light the moment she sees Ben, and expects him to contribute to the production of the good, say, by finding a better stick.*

Scenario 4: *Penny has no rational strategy whatsoever pertinent to this interaction, and Ben walks a few steps behind her until the end of the woods.*

We can easily eliminate Scenario 4 since it is evident that Penny, in this case, would not be a CM. Even a charity-like value maximization would allow us to conclude that she was a CM, but this does not seem to be the case, nor is it possible to argue that such a strategy falls within the scope of this project.

In Scenario 3, Penny’s action of extinguishing is not fair since it is done preemptively without adequate signaling. Such an action, with bad faith, would clearly be morally impermissible since Penny doubles her costs only to bargain Ben to concede further.

In Scenario 2, Penny acts as an *ex ante* CM, but exploits her higher bargaining position by prompting Ben to make unfair concessions, as long as it is possible to argue that the expected contribution of \$5 would exceed the value of light (V) and Penny’s cost of production (c). In perfect market conditions, Penny is capable of evaluating the cost of her production, but this evaluation should be made in good faith insofar as it prompts Ben to engage in future interactions with Penny (Gauthier, 1986: 85). In this case, the surplus ($2V - c$) could be rationally divided according to MRC if and only if the amount of the expected optimized transfer payment (t^*) is equal to $\frac{c}{2}$. Only if this equality holds, then $U_{Penny} = U_{Ben} = V - \frac{c}{2}$ would hold under the perfect market conditions.

Notice that it is rational for Penny to expect a \$5 payment, but as an *ex ante* CM, this would be morally impermissible. If her expected contribution is not in accordance with

the principle that prompts rational agents to engage in bargains based on their relative concessions, she would instead be an *ex ante* SM.

It is disparaging for Penny to adopt such a strategy if she foresees future interactions with Ben and their fellows, who, if CMs, would treat Penny as an exploitive person that is not fit to bargain with. In aversion to such an outcome, an *ex ante* CM would be disposed to expect contributions compared to otherwise beneficial arrangements; in other words, if $t^* = \frac{c}{2}$ does not hold for a \$5 payment for the light, she must be able to calculate that Ben would be better off not cooperating with her, which would decrease the expected payoffs of both parties.

If Penny signals Ben that she would expect an adverse contribution for benefiting from her production, Ben would be rational to seek different alternatives, such as investigating means to produce his own stick, but also, he would not be immoral if he decides to free ride on her production, as it would be similar to adopting an SM strategy against an SM.

The outcomes of Scenario 2 and alike are certainly unfavorable, as they encourage straightforward maximization over social cooperation. The crux of the possibility of *ex ante* constrained maximization is the moral motivation to mitigate the rational strategy of acting first in the state of absence. Hence, although capable of rational calculations over the external benefits, an *ex ante* CM must be normatively disposed to proposing only agreeable contributions that would not only be fair but also beneficial to all the engaged parties. The rational calculation of future interactions must also not be disregarded.

It is therefore evident why only the Scenario 1 works as an auspicious outcome for an *ex ante* CM in such an interaction. In perfect market conditions, Penny can certainly evaluate her product with a value much greater than a simple ‘thank you’; nevertheless, as an *ex ante* CM, she must preemptively concede from this pure evaluation towards a cooperative outcome that would be fair to ask, but also would prompt Ben to act accordingly, compared to him adopting an SM strategy. This could also be denoted as

Penny incorporating a value of her moral recognition (ϵ) to her expected payoffs, and comparing being a CM to her surplus as a perfect market SM²⁷.

Consequently, Penny's expected utilities in Scenario 1 could be denoted as $U_{Penny} = (V - c) + \epsilon$. Comparatively, this expected utility would be greater than her expectations in Scenario 2 that we denoted as $V - \frac{c}{2}$, if ϵ value of moral recognition is significant, which I assume that it is. This explains the rationality of Penny being an *ex ante* CM.

Notice that for Ben, every strategy other than free riding would prompt him to subtract a compliance cost (δ) from his expected value²⁸. This differs from the one I have argued above as t^* which denotes a monetary fee agreed upon under perfect market conditions. While t^* denotes only a monetary fee, the δ cost can either be monetary or moral, as it is in the case of having to say 'thank you' as a fair and agreeable cost of cooperation.

If CM, then Ben must not calculate that the value of free riding (V) is greater than the value of cooperating ($V - \delta$), which would essentially appear as if Ben is better off if he walks away without bargaining. Instead, he must understand that any action of his signalling a noncooperative play would rationally prompt and morally allow Penny to extinguish the light. Therefore, Ben, if *ex post* CM, must compare the value of cooperation ($V - \delta$) with the value of darkness (0).

²⁷ It is important to distinguish Penny's expected moral value of ϵ with our discussions on the unconditional plays above. An objector may claim that Penny could expect an ϵ value where she plays unconditionally, say, by allowing Ben to benefit from the light without any expectation. In that case, however, I contend that Penny does not play a strategy that conforms to any moral conventions (see Lewis, 1969; Biccheri, 2014), insofar as it is possible to establish that moral play in the instances I am interested in, particularly in the state of absence, must be prompted by rational expectation of compliance. This is the reason why I put an emphasis on *proto-contractual norms* in relation with the ratio of moral agents in a social order. I contend that, Penny could only expect a value of ϵ if and only if she is a rational self-interest maximizer in a condition in which she sits at the hypothetical bargaining table prior to any actual arrangement, with the rational expectation of compliance; and, moral plays must be conformed accordingly, particularly in order to sustain the resource-management in such conditions (see Chapter 3.5).

²⁸ Also note that Ben's δ does not need to be equal to Penny's moral recognition of ϵ . Unlike an expected monetary transfer fee of t^* in perfect market conditions where one agent's loss (-\$5) is equal to other's gain (+\$5), I hold that normative signaling would allow for asymmetries. Even, the δ cost might be negligible compared to arguably significant moral recognition for Penny, or the other way around. This asymmetry, and the possibility of a positive sum, further illustrates the plausibility of Scenario 1 against Scenario 2.

In other words, since $V > V - \delta$, the value of free riding is greater than the value of cooperating for Ben. If Ben's strategy is to free ride, however, Ben's payoffs would equal to 0 since Penny is rational to punish him with darkness.

Then, Ben's strategy would transform into the following contemplation that suggests that he must cooperate if and only if $U_{cooperate} > U_{darkness}$; in other words, if $V - \delta > 0$.

3.5. Larger Scale Stability

Presupposing that Penny has ownership over her production, which allows her to expect reciprocity as others enjoy the benefits, does neither entail that Penny can exploit this capability, nor that Ben can morally free ride on the benefits he has enjoyed. In the state of absence, agents are isolated to produce and maintain goods with a public nature. In such a state, it is crucial for agents to be the ones who sit at the bargaining table with their hands full, while expecting rational compliance.

Notice that according to Gauthier's description of a CM, she must be able to calculate whether to adopt a constrained maximizing strategy if the ratio of CMs in a given society (r) is adequately high (Gauthier, 1986: 175). In my account, such a calculation is impractical since, in the state of absence, it is difficult, if not impossible, for an agent to abandon a CM strategy altogether while acting *ex ante*. As if Penny extinguishing her light, this would prompt an agent that say, who builds roads, to demolish the road infrastructure if the expected value of compliance is not met by the beneficiaries²⁹.

The state of absence is similar to the dark valley that Penny walks with Ben. Our pursuit is to find normative moral grounds that shed some light on it. I have so far illustrated why Penny is rational to be an *ex ante* CM, and Ben must rationally become an *ex post* CM the moment he enjoys the benefits created by her. My final goal is to show that in the state of absence, we can descriptively presuppose that the r ratio is sufficiently high,

²⁹ An objector may claim that this analogy does not hold, since demolishing the road requires further costs that would not be a proportionate response against non-compliance. If so, suppose that the road-builder closes the gates allowing others to ride on the road.

so that it allows agents to sustain social, economic, and moral interactions without any centralized body of governance.

To illustrate, let W be the lifetime value of participating in such interactions, i.e., being a cooperative and a moral member of a social network in the state of absence. I assume this value would be scaled in accordance with the social network's population (N) according to Metcalfe's law, as it expands in proportion to the value of cooperation (Metcalfe, 2013). The agent, particularly the party that enjoys the produced benefits, must rationally compare the value W with the expected value of being alienated from social and economic interactions (A).

If Penny is an *ex ante* CM, she expects a fair cost of contribution from Ben in accordance with the MRC principle (Gauthier, 1986). I have illustrated above that if this expected cost of contribution is unfairly high, this would mean that Penny is not an *ex ante* CM; therefore, Ben cannot be rationally expected to become one. Suppose that as an *ex ante* CM, Penny expects a fair contribution, such as having Ben thank her as a result of this interaction. Ben is then better off complying with her since his otherwise signals would prompt Penny to update her strategy of CM.

On larger-scale social interactions, however, this analogy would not hold, as it cannot always be realistically expected that all CMs in a society will update their strategy when encountering a defector. This could quite contrarily mean that a defector is signaled that he can defect and face no consequences, since he has the higher bargaining position as the producer of a public good cannot rationally demolish the production because someone is free riding.

Then, Ben's cost of contribution (δ) once again becomes relevant. Notice that despite normative obligations, Ben can rationally calculate that he has the higher bargaining position if and only if $V_{evade} > V_{cooperate}$ assuming $V_{cooperate} = V - \delta$ and $V_{evade} = V$. This inequality will always result in favor of Ben not complying in one-instance interactions unless strengthened by larger-scale factors, since even the simplest expected compliance of having to say 'thank you' would skew the inequality in favor of defecting.

For larger-scale interactions, therefore, it would be unfavorable to rely on this single-instance expected value calculation, since the defector might be signaled that he has the higher bargaining position, as the producer cannot demolish the resources she has produced.

Since it would also be objectionable to coerce Ben into becoming an *ex post* CM, we must present him with an alternative of being excluded from future social and economic interactions. In other words, on the larger social scale, Ben would no longer be rational to calculate a single instance comparison between V_{evade} and $V_{cooperate}$, since the strategy he plays in this single instance interaction would be directly related to his expected value of social inclusion (W) and, otherwise, the value of his being alienated (A).

We can therefore conclude that the value of being included in social interactions (W) would have an advantageous surplus over the value of alienation (A), assuming that the agent does not prefer *autarky* or a similar condition, which could essentially prompt him that this value is greater than the simple cost of expected contribution as $W - A > \delta$. If this inequality holds, which I assume is likely to hold than not, then any incidental cost of compliance would yield favorable payoffs in the long run.

This translates into the following contemplation for Ben: “Penny has already produced a light, I’d better give her what she is asking because in the long run, I am better off having good social and economic relationships with her than not.”

On the other hand, Ben could as well contemplate as follows: “Penny produced a light, and I am expected to somehow contribute. Who cares? I will never see her again, nor ever set foot in these woods!”³⁰

To illustrate, consider an elaborate case of public infrastructure. Suppose that Penny lives near a seashore and realizes that many ships move across the shore every night

³⁰ This SM strategy for Ben is also rational in my account, as he calculated the expected value of being included in larger-scale social interactions to be low. Notice that this strategy holds if and only if Ben gives up any future payoffs he could receive by not only interacting with Penny, but also with her fellows.

without adequate lighting. She decides to build a lighthouse on the shore to help the ships navigate at night. She believes that by building a lighthouse, she would receive an adequate amount of payment to cover both the lighthouse's maintenance and her daily needs.

Penny is once again an *ex ante* CM since she sits at the bargaining table prior to any agreement with her hands full. Her constraints, i.e., producing a public resource with expected cooperation, are fair and rational insofar as she expects a fair payment that is compatible with the principle of MRC.

Further, suppose that she provides two other services contingent on the lighthouse property: a tavern and a dock. Arguing that docking itself would require a fee would fall under the Coasean analysis of exclusion that would not only be compatible with my arguments on ownership so far, but also with the economic theory of clubs (Coase, 1974; Buchanan, 1965). In an attempt to more elaborately demonstrate the crucialness of *ex ante* CM in the state of absence, I will skirt this conclusion for now by ignoring her exclusion for the contingencies, and investigate a distinct form of benefit, which is the light itself.

Suppose that she expects three distinct forms of compliance: First, if you dock near the marina, you pay, or you will be excluded. Second, if you buy a beer in the tavern, you must pay; otherwise, you will be excluded. Third, if you benefit from the light I have provided, I expect you to donate a fair share so that I can maintain the lighthouse for the future³¹.

Notice why it is plausible to ignore the first two, since exclusion is not only tied to the nature of the goods and services, but also compatible with our arguments for *ex ante* CM so far. The third expectation, however, is rather controversial.

Why, Ben, now a ship captain, would pay a fair share of the donation so that Penny can maintain the lighthouse? This decision is entirely based on the calculations I provided

³¹ My argument for the ownership of light as a public good contingent on a private property is hinted as a summary above, see Chapter 1.7. Here, I would also like to underline David Schmitz argument for preserving goods for the future generations (Schmitz, 1991: 24).

above regarding the value of social and economic inclusion (W) and the value of being excluded from social interactions (A).

The fair donation, which is not even evaluated by Penny, is based on MRC, as she expects agents to pay according to their relative concessions. Ben must calculate whether the cost δ exceeds the surplus value of social inclusion ($W - A$). In other words, he must contemplate whether the cost δ is relatively smaller than the expected surplus of being included in future interactions with Penny and her fellows.

It is evident that Ben the Captain can rationally evade the honor payment (δ) if he believes he will never return to this coast and interact with anyone involved in this social and economic network, since $V_{evade} > V_{cooperate}$. This decision would be rational if the value of social inclusion (W) is smaller than the value of evasion/free riding/alienation (A), in which Ben the Captain has rational grounds to conclude that he will not be interacting with anyone within this social order in the future. In that case, Ben could rationally calculate that $W \approx 0$.

Nevertheless, consider the following: Ben the Captain docks one day at Penny's marina by paying the necessary fee. He walks into Penny's tavern, asks for a beer, and again, pays his tab³². As he tries to find a table, he glances at the box dedicated to honor payments for the maintenance of the lighthouse. Befuddled, "Who pays for the light anyway?" he asks one of his fellow captains.

The answer to this inquiry is directly related to the ratio of moral agents (r) in a given society that we have so far been concerned with. Whether his fellow captain would agree on this intuitively ridiculous idea of paying an honor fee for benefiting from a public light signals Ben the moral play.

Suppose that the fellow captain, let us call her Claire, is wary of the value of social inclusion (W). As a CM, she is aware that: (1) There are proto-contractual norms that

³² In both instances, the expected value comparison of $V_{evade} > V_{cooperate}$ does not hold since his evasion means he would be excluded from enjoying the benefits. If he asks for a beer and does not pay its price, he will be removed from the tavern, with potential legal and/or social sanctions.

agents are better off complying with than defecting; (2) She is a voluntary member of a social order that satisfies the necessary conditions of the state of absence; (3) Penny has provided for this community that otherwise no one has; (3.1) If Penny’s expectation were to be unfair, members have grounds to bargain with her; (3.2) If Penny’s expectations were to be unfair, then someone else can also operate a lighthouse next to hers, ask a much fair contribution, and she will be free to engage with them instead of Penny; (4) She is better off within this community than being alienated; (5) If there are other communities that yields her more payoffs, no one is coercing her into obligations within this community, so she can leave.

Having understood Ben’s astonishment, Claire then tells him, “We all give; I, for instance, give 10% of my daily purchases; it is *nothing* compared to the *benefits* I receive”. Notice that Claire is aware that she is not paying for the light *per se*; instead, she is rationally complying with the moral necessities of the social order she is voluntarily being a part of.

If Ben shrugs over Claire’s explanation and insists on his morally unjustified conviction that it is irrational to pay a light, Claire, and any other member of this social and economic network, would get signaled that Ben would not make a plausible business partner, nor a party to make agreements with. Then, Ben would only be better off if he boards his ship and never returns to this community.

If *ex ante* CM is possible and rational, Ben’s decision appears to be transformed into a game in which he plays against the social network of CMs. Suppose that there are N members in this society, and I assume the ratio of moral agents in the society (r) is close to 1 to illustrate that the aggregated utility of this social order scales with the number of members, as there will be more signals towards cooperation if *ex ante* CM is rational.

The following summation denotes the of the value of social inclusion (W) for Ben³³:

$$W_{Ben} = \sum_{i=1}^{N-1} (V_i - \delta) \approx (N - 1)(V - \delta).$$

³³ I assume that this value would scale with the population, in accordance with the Metcalfe’s law (Metcalfe, 2013).

This means that the value he receives as he cooperates in one-instance interactions scales accordingly with the population increase per Metcalfe’s law, which would eventually allow us to conclude for Ben, if *ex post* CM, the following inequality would hold true: $W \gg A$.

The contemplation of his paying an honor payment for Penny transforms into an assurance/coordination game in which norm confirmation is at stake for the social network (the established norm could also act as a device for equilibrium selection, cf. Harsanyi & Selten, 1988; Vanderschraaf, 2001: 43). In this game, Ben plays for W while the social network plays for conforming to its norms, which means that in the state of absence, the production-benefit relationship with *ex ante* CM must be sustained by *ex post* CM (Figure 8).

Figure 8. Coordination game between Ben and the network³⁴

		The Network	
		<i>Include</i>	<i>Alienate</i>
Ben	<i>Comply (ex post CM)</i>	W , Norm conformed	$-\delta$, Norm erodes
	<i>Defect (SM)</i>	V_{evade} , Norm erodes	A , Norm conformed

Notice that Ben’s strategies are $W \gg V_{evade} > A > -\delta$; whereas for the social network, conforming to the norm pays off more than eroding the norm, since they are in the state of absence, in which goods would otherwise not be produced.

An objector might claim that Ben’s decision to comply is still shadowed by uncertainty, since in this game, Ben’s action of compliance might yield him the lowest payoff (or, sanction) of $-\delta$. The safe response would be to claim that the social network would

³⁴ Notice two Nash equilibria of (comply, include) and (defect, alienate).

never play a norm-eroding strategy since, first, they are better off as N increases³⁵; and, second, they reward the signal of cooperation in the state of absence.

If this response is still unsatisfying, let us compare the expected value of compliance with defection for Ben. Let p be the probability that the society will conform to the norm (plays include). Then, Ben only plays compliance (ex post CM) if $U_{Comply} > U_{Defect}$. Let us transform this into the following: $p(W) + (1-p)(-\delta) > p(V) + (1-p)(A)$. Let us once again assume A is equal to or closer to 0³⁶.

Then, solving for p , the inequality transforms into $p > \frac{\delta}{W-V+\delta}$. Since we already assumed $W \gg \delta$ and V , even with a low trust of $p \approx 0$, Ben would be rational to cooperate, since the payoffs of social inclusion (W) are assumed to be substantial.

Consequently, the rational strategy of *ex post* CM transforms into an assurance game that the reader might be familiar with, as it is argued for the problem of cooperation (Hampton, 1987; Taylor, 1987). Notice, only that, in this case, the assurance game model is not descriptive; instead, it is normative, which signals the agent to become an *ex post* CM to be able to engage in social and economic interactions that would yield substantial payoffs compared to a noncooperative strategy.

³⁵ Assuming per the Metcalfe's Law (Metcalfe, 2013). This would not hold if the increase of N mean a decrease of r , so I argue this would not be the case since Ben's inclusion is only rationally permitted if he is a CM. So, an increase of N would scale with r accordingly.

³⁶ I am aware that this is a far-fetched assumption, and an objector would claim that A could be substantially valuable compared to W . Despite being a crucial point, arguing against this would be beyond the scope of this project.

CONCLUSIONARY REMARKS

My pursuit started with a quixotic acknowledgement pertinent to the decades-long debate surrounding the collective action problem, that the intuitive solutions remain worrisome to this day. I believe this conviction and motivation are quixotic not only for my account but for any account that challenges the intuitive solution of the problem, that is, the State.

I wanted to present an alternative account based on moral and rational decision-making. In this sense, relying on Gauthier's invaluable proposal of constrained maximization would seem to be working, until, unfortunately, it is not.

The problem with constrained maximization was not fundamental, but rather incidental. Since Gauthier was mainly concerned with an ethical theory, he understandably was not adequately engaged with the puzzles related to public goods. It was my challenge, if I wanted to incorporate constrained maximization into this debate, to bridge these two concepts.

First, I needed to present grounds for voluntary engagement between parties, and I did so by arguing for proto-contractual norms. By proto-contractual norms, agents are able to signal their readiness to others for social, economic, and moral bargains.

Temporality then appeared as a major obstacle, since Gauthier's account of translucency only appears to be working on a simultaneous level (Gauthier, 1986: 174). This, however, was not adequate for me to argue for the rationality of acting first.

Acting first is a problem only for the state of absence, a state that I define as similar to the state of nature, but not as a proto-State condition. In the state of absence, agents are isolated within the scope of their engagement network, providing public goods based on Lockean appropriation and production. This condition is the alternative I presented for addressing the collective action problem.

Then, the question of *ex ante* constrained maximization became the question of "Who starts first?" I presented rational grounds for such an initiative and attempted to tie it up with a rational decision to engage in such interactions by becoming an *ex post* constrained maximizer.

Finally, moral norms emerged as an assurance game between the beneficiary and the network of rational agents in the state of absence. Interpreting the problem of cooperation as an assurance game is not novel to my account; however, I presented grounds for its normative nature, instead of being a descriptive illustration.

Consequently, I argued that within the described conditions, the ratio of moral agents in a state devoid of a centralized resource allocation scheme could be rationally interpreted as close to 1 as a convergence, since agents need to rely on themselves; by providing and fairly compensating each other, they can sustain larger-scale social, moral, and economic interactions.

As a final remark, I must say that I argued on the *possibility* of this rational calculation. The *plausibility* of this account, particularly in light of the Hobbesian theory of the state, is a crucial and fundamental question that I must leave for future studies.

REFERENCES

- Arneson, R. J. (1982). "The Principle of Fairness and Free-Rider Problems" *Ethics* 92(4), 616–633. <http://www.jstor.org/stable/2380394>
- Arneson, R. J. (1987). "Locke versus Hobbes in Gauthier's ethics" *Inquiry: An Interdisciplinary Journal of Philosophy* 30 (3):295–316.
<https://doi.org/10.1080/00201748708602125>
- Arneson, R. J. (1991). "Lockean Self-Ownership: Towards a Demolition" *Political Studies* 39(1), 36-54. <https://doi.org/10.1111/j.1467-9248.1991.tb00580.x>
<http://dx.doi.org/10.1111/j.1467-9248.1991.tb00580.x>
- Axelrod, R., & Hamilton, W. D. (1981). "The Evolution of Cooperation" *Science*, 211(4489), 1390–1396. <https://doi.org/10.1126/science.7466396>
- Barry, B. (2010) "David Hume as a Social Theorist" *Utilitas* 22 (4), 369–392.
<https://doi.org/10.1017/S0953820810000300>
- Bertrand, B. (2006). "The Coasean Analysis of Lighthouse Financing: Myths and Realities" *Cambridge Journal of Economics* 30(3), 389–402. <https://doi.org/10.1093/cje/bei068>
- Bicchieri, C. (2014). "Norms, Conventions, and the Power of Expectations" in Nancy Cartwright & Eleonora Montuschi (eds.) *Philosophy of Social Science: A New Introduction*. Oxford University Press.
- Block, W. (1983). "Public Goods and Externalities: The Case of Roads". *The Journal of Libertarian Studies*, 7(1), 1–34.

- Bogart, J. H. (1985). "Lockean Provisos and State of Nature Theories" *Ethics* 95 (4) 828–836. <https://www.jstor.org/stable/2381259>
- Burdick, F. M. (1902). "What is the Law Merchant?" *Columbia Law Review* 2 (7), 470–485. <https://www.jstor.org/stable/1109926>
- Buchanan, J. M. (1965). "An Economic Theory of Clubs" *Economica*, 32(125), 1–14. <https://doi.org/10.2307/2552442>
- Coase, R.H. (1974) "Lighthouse in Economics" *Journal of Law and Economics* 17 (2) 357–376
- Cullity, G. (1995). "Moral Free Riding" *Philosophy and Public Affairs* 24(1), 3–34. <https://doi.org/10.1111/j.1088-4963.1995.tb00020.x>
- Cullity, G. (2025). "The Free Rider Problem" *The Stanford Encyclopedia of Philosophy* (Fall 2025 edition), Edward N. Zalta & Uri Nodelman (eds.). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2025/entries/free-rider/>
- Fudenberg, D. & Tirole, J. (1991). *Game Theory*. MIT Press Books.
- Eggers, D. (2011) "Hobbes And Game Theory Revisited: Zero-Sum Games In The State Of Nature" *The Southern Journal of Philosophy* 49 (3) 193–226 <https://doi.org/10.1111/j.2041-6962.2011.00071.x>
- Eggers, D. (2019). "Hobbes, Kant, and the Universal 'right to all things', or Why We Have to Leave the State of Nature" *Hobbes Studies* 32(1), 46–70. <https://doi.org/10.1163/18750257-03201004>
- Feeny, D., Berkes, F., McCay, B. J., & Acheson, J. M. (1990). "The Tragedy of the Commons: Twenty-Two Years Later" *Human Ecology* 19(1), 1–19. <https://doi.org/10.1007/BF00889070>
- Fontaine, P. (2014). "Free Riding" *Journal of the History of Economic Thought*, 36(3), 359–376. doi:10.1017/S1053837214000376
- Friedman, J. W (1971) "A Non-cooperative Equilibrium for Supergames" *The Review of Economic Studies* 38 (1), 1–12, <https://doi.org/10.2307/2296617>

- Gauthier, D. (1986). *Morals by agreement*. Oxford University Press.
- Gauthier, D. (1991). “Why Contractarianism?” In Peter Vallentyne (ed.), *Contractarianism and Rational Choice: Essays on Gauthier’s Morals By Agreement*. Cambridge University Press.
- Gauthier, D. (2022a) “Taming Leviathan” in Susan Dimock, Claire Finkelstein, and Christopher W. Morris (eds.), *Hobbes and Political Contractarianism: Selected Writings*
- Gauthier, D. (2022b) 'Political Contractarianism', in Susan Dimock, Claire Finkelstein, and Christopher W. Morris (eds), *Hobbes and Political Contractarianism: Selected Writings* <https://doi.org/10.1093/oso/9780192843005.003.0009>
- Gilbert, M. (2006). *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford University Press.
- Hampton, J. (1986). *Hobbes and the Social Contract Tradition*. Cambridge University Press.
- Hampton, J. (1987). “Free-Rider Problems in the Production of Collective Goods” *Economics and Philosophy* 3 (2), 245–273. doi:10.1017/S0266267100002911
- Hardin, G. (1968). “The Tragedy of the Commons: The population problem has no technical solution; it requires a fundamental extension in morality” *Science* 162(3859), 1243– 1248. <https://doi.org/10.1126/science.162.3859.1243>
- Hardin, R. (1971). “Collective Action as an Agreeable n-Prisoners’ Dilemma” *Behavioral Science* 16(5), 472–481. <https://doi.org/10.1002/bs.3830160507>
- Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics*. Oxford University Press.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., (...) Oliphant, T. E. (2020). “Array programming with NumPy” *Nature*, 585 (7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>

- Harsanyi, J. & Selten, R. (1988) *A General Theory of Equilibrium Selection in Games*. MIT Press.
- Hobbes, T. (1985 [1651]). *Leviathan* (C. B. Macpherson, ed.). Penguin Books.
- Hobbes, T. (1998 [1642]). *On the Citizen* (R. Tuck & M. Silverthorne, Eds.). Cambridge University Press
- Huemer, M. (2016). “A Liberal Realist Answer to Debunking Skeptics: The Empirical Case for Realism” *Philosophical Studies* 173 (7), 1983 – 2010.
<https://www.jstor.org/stable/43895543>
- Hume, D. (1985 [1739]). *A Treatise of Human Nature* (E.C. Mossner, ed.). Penguin Books.
- Kalai, E., & Smorodinsky, M. (1975). “Other Solutions to Nash’s Bargaining Problem” *Econometrica* 43(3), 513–518. <https://doi.org/10.2307/1914280>
- Kavka, G. S. (1983). “Hobbes’s War of All Against All” *Ethics*, 93(2), 291–310.
<http://www.jstor.org/stable/2380421>
- Kavka, G. S. (1986). *Hobbesian Moral and Political Theory*. Princeton University Press.
- Klosko, G. (1987). “The Principle of Fairness and Political Obligation.” *Ethics* 97 (2) pp. 353–362. <https://doi.org/10.1086/292843>
- Kritikos, A.S., & Meran, G. (1998) “Social Norms, Moral Commitment, and Cooperation” *Homo Oeconomicus* 15, 71-92.
<https://ideas.repec.org/a/hom/homoec/v15y1998p71-92.html>
- Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press.
- Locke, J. (1980 [1690]). *Second treatise of government* (C. B. Macpherson, Ed.). Hackett Publishing Company.
- Metcalf, B. (2013). “Metcalf’s Law after 40 Years of Ethernet” *Computer* 46 (12), 26 – 31. <https://doi.org/10.1109/MC.2013.374>
- Mill, J. S. (1994 [1848]). *Principles of Political Economy* (J. Riley, Ed.). Oxford University Press.

- Moehler, M. (2020) “Minimal Morality, Bargaining Power and Moral Constraints: Replies to D’Agostino, Thrasher, Morris, and Vanderschraaf” *Analytic Philosophy* 16 (1) 87 – 100. <https://doi.org/10.1111/phib.12182>
- Moore, M. (1994). *The Social Contract Tradition from Hobbes to Rawls*. Routledge.
- Narveson, J. (1998). *The Libertarian Idea*. Temple University Press
- Nash, J. F. (1950). "The Bargaining Problem." *Econometrica*, 18(2), 155-162
<https://www.jstor.org/stable/1907266>
- Nash, J. F. (1953). “Two-Person Cooperative Games" *Econometrica*, 21(1), 128–140
<https://www.jstor.org/stable/1906951>
- Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books.
- NumPy Developers (2024). *NumPy* (Version 2.3.0) [computer software].
<https://numpy.org>
- Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.
- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.
- Ostrom, E. & Ostrom, V. (2004) “The Quest for Meaning in Public Choice” *The American Journal of Economics and Sociology* 63 (1) 105–147.
<https://doi.org/10.1111/j.1536-7150.2004.00277.x>
- Pigou, A. C. (1932). *The Economics of Welfare*. Macmillan and Co.
- Rawls, J. (1971). *A Theory of Justice*. The Belknap Press of Harvard University Press.
- Rothbard, M. N. (1982). *The Ethics of Liberty*. Humanities Press
- Rutherford, D. (2002). *Routledge Dictionary of Economics*. London: Routledge.
- Samuelson, P. A. (1964) *Economics: An Introductory Analysis* (6th ed.)
- Sandler, T., Tschirhart, J. (1997) “Club theory: Thirty years later” *Public Choice* 93, 335–355 <https://doi.org/10.1023/A:1017952723093>
- Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press.

- Schmidtz, D. (1991). *The Limits of Government: An Essay on the Public Goods Argument*. Westview Press.
- Sharp, M. P. (1941). “Pacta Sunt Servanda” *Columbia Law Review*, 41(5), 783–798.
<https://doi.org/10.2307/1117839>
- Sidgwick, H. (2011 [1901]). *The Principles of Political Economy*. Cambridge University Press.
- Simmons, A. J. (1979). *Moral Principles and Political Obligations*. Princeton University Press. <https://doi.org/10.2307/j.ctvzxx97q>
- Simmons, A. J. (1992). *The Lockean Theory of Rights*. Princeton University Press.
<https://doi.org/10.2307/j.ctv173f2n3>
- Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge University Press.
- Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press.
- Tarlock, A. D. (2000). “Prior appropriation: rule, principle, or rhetoric” *North Dakota Law Review*, 76(4), 881-910. <https://commons.und.edu/ndlr/vol76/iss4/5>
- Taylor, M. (1987). *The possibility of cooperation*. Cambridge University Press.
- The Axelrod Project Developers (2023). *Axelrod-Python/Axelrod: v.4.12.0* (Version 4.13.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.7861907>
- Tuck, R. (2008). *Free Riding*. Harvard University Press
- Vanderschraaf, P. (2001). *Learning and coordination: Inductive deliberation, equilibrium and convention*. Routledge.
- Vanderschraaf, P. (2010) “The Invisible Foole” *Philosophical Studies* 147 (1), 37–58.
- Vanderschraaf, P. & Valls, A. (2019) “Hume on Decisions, Conventions and Justice” in Angela Coventry & Andrew Valls (eds.) *David Hume on Morals, Politics and Society*, Yale University Press, 317–339.
- Vanderschraaf, P. (2023). *Bargaining Theory*. Cambridge University Press.

- van Zandt, D.E. (1993) “The Lessons of Lighthouse: ‘Government’ or ‘Private’ Provision of Goods” *The Journal of Legal Studies* 22 (1), 47 – 72.
<http://dx.doi.org/10.1086/468157>
- van der Vossen, B. (2021). “Property, the environment, and the Lockean Proviso” *Economics and Philosophy* 37 (3):395 - 412.
<https://doi.org/10.1017/S0266267120000401>
- Wahlberg, H. (1959). “Pacta Sunt Servanda” *The American Journal of International Law*, 53(4), pp. 775–786. <https://www.jstor.org/stable/2195750>